Aberration correction: zooming out to overview

BY A. HOWIE*

Cavendish Laboratory, University of Cambridge, J.J. Thomson Avenue, Cambridge CB3 0HE, UK

In the structural characterization of thin specimens by projection (atomic column) imaging, the revolutionary development of aberration-corrected electron microscopy has already brought significant improvements not only in spatial resolution but also in improved image contrast. Some highlights from the symposium are summarized. Despite the purchasing and operating costs as well as the demands they place on operator skills, a staggering number of these new microscopes has already been installed worldwide. Serious challenges, therefore, arise including the need to attract customers from a wide range of disciplines where complex structure problems may require the development of new imaging modes. The ability to image at large scattering angles may be useful in mitigating some of the many as-yet uncorrected delocalization mechanisms that then arise and are systematically identified here. Larger specimen volumes made possible by chromatic aberration correction will benefit the development of more relevant in situ observations, particularly in materials science and catalysis. With additional incorporation of phase shifting electrodes or other devices, these chromatic aberration correctors could also be important for efficient phase contrast imaging in easily damaged biological structures. For many of these formidable problems, earlier experience of the optical microscopy community may teach some lessons.

Keywords: image delocalization mechanisms; hollow cone imaging; confocal imaging; depth sectioning

1. Introduction

A clear message from the symposium is that aberration-corrected electron microscopy (ACEM), in a triumph of sophisticated focusing hardware and computer control, has wrought a step function change in challenges as well as opportunities. In the special but widely used case of projection imaging and energy loss analysis of atomic columns at both planar and well-aligned linear defects in crystals, this exciting development is already demonstrating the expected improvements in spatial resolution. Somewhat less expected, but possibly even more significant, has been the new life brought to many other aspects of electron microscopy. As the first of several instances, improved contrast makes much easier the projection imaging of columns of light atoms such as

*ah30@cam.ac.uk

One contribution of 14 to a Discussion Meeting Issue 'New possibilities with aberration-corrected electron microscopy'.
oxygen in conventional transmission electron microscopy (CTEM). One factor in this advance, demonstrated at the symposium by Urban, has been the discovery that phase contrast is improved by working with a very small negative value of the spherical aberration constant $C_3$ and in overfocus conditions so that the atomic columns appear bright. In high-angle dark field scanning transmission electron microscopy (HAADF-STEM), which is the other widely used method for atomic column imaging and spectroscopy, similar improvements in contrast as well as resolution have followed the introduction of ACEM. However, an additional and possibly even more significant benefit here (made clear by the contributions of Pennycook and Muller at the symposium) has been the consequent great increase in probe current, leading to much shorter acquisition times and/or less noisy data.

Discussion of contrast in conventional high-resolution imaging already takes us beyond electron optics and into issues of specimen scattering, some aspects of which remain inadequately understood but where investigation has been reinvigorated by the availability of ACEM. Although only a few of these were touched on at the symposium, a brief account of recent developments in quantitative simulation and on the status of projection imaging and channelling is given in §2. Beyond the atomic column elastic imaging procedure, a rich variety of as-yet uncorrected ‘delocalization’ effects come into play and these generally dominate over electron optics in determining the effective spatial resolution. If these problems are to be systematically addressed, a classification of the kind attempted in §3 will be essential. As discussed in §4, the increased range of imaging angles made available in ACEM offers opportunities for mitigating some of the lateral delocalization effects and also for improving the longitudinal delocalization through reduced depth of focus.

Other potentially important spin-off consequences of the development of spherical aberration correctors can be discerned, including a systematic appraisal and enhancement of the whole microscope system as described in Krivanek’s contribution at the symposium. High-resolution performance at operating energies of 60 keV or even lower is now becoming available and will greatly facilitate the study of light atom samples such as carbon nanotubes susceptible to knock-on beam damage. Improved spectroscopy, especially at energy losses down to 1 eV and below, is facilitated by monochromators, although with a considerable loss of beam intensity. At the symposium, we also learnt from Haider that chromatic aberration correctors are now in the final stages of development. These should be particularly useful for high resolution, in situ, materials science and chemistry observations involving large specimen volumes and long working distances when the increase in chromatic aberration is potentially serious. Chromatic aberration correction in the STEM mode operates in the pre-specimen region and thus simply compensates the probe-forming system for the incident beam energy spread. In the CTEM mode, the chromatic aberration corrector is placed after the specimen and can thus in principle also correct for the much larger energy spread generated by inelastic scattering there. This is most likely to be particularly useful in biology where the images from different energy slices can be sufficiently similar owing to very delocalized inelastic scattering with image contrast at all energy losses arising only from elastic scattering events. At least equally exciting for biologists would be the development of improved phase shifting devices for efficient imaging of nearly pure phase objects sensitive to ionization damage. Renewed investigations with phase shifting plates of the
Zernike type used in optical microscopy are underway. However, these could be eclipsed by a successful outcome of the scheme described at the symposium by Rose for achieving the same result with additional charged electrodes in the aberration corrector. It will be fascinating to discover whether the combination of chromatic aberration correction and phase shifting will improve the spatial resolution generally attainable in biological imaging to the point where spherical aberration becomes necessary. The resolution limit of chromatic aberration-corrected phase imaging will be determined not only by the phase shifting device performance but also in thicker samples by the angular spread of valence losses.

It may be noted that the developments just described are less directed towards sub-angstrom resolution than to widening the scope of electron microscopy applications. Progress on this front was reported at the meeting by Colliex, Pennycook, Kirkland and Gai and reflects a response to one of the several challenges discussed in §5 that ACEM poses for the microscopy community. The history of aberration correction in optical microscopy exhibits a number of interesting parallels with the current position in electron microscopy and may even offer clues to future events.

2. Understanding and analysing high-resolution microscopy images

(a) Quantitative simulation

Computer simulation is frequently an essential step in identifying a structural model or models that will give rise to the observed high-resolution microscopy (HREM) image. Fully quantitative agreement with image intensities or contrast is rarely achieved, however, and indeed image intensity levels relative to the incident beam are not often measured. Overcoming this shortcoming is not simply a question of confirming the validity of Schrödinger’s equation, but would also add a crucial additional test factor providing scope for warning signals when quantitatively successful matching is not achieved. There are welcome signs that the availability of ACEM has brought renewed enthusiasm for addressing this difficulty.

In CTEM high-resolution imaging, there is a frequently observed discrepancy between computed and observed image contrast referred to as the Stobbs factor with a value typically about 3 (Hytch & Stobbs 1994). The modular transfer function of the detector system, hitherto ignored in discussion of this problem, may resolve much of the difficulty (Thust 2008). There is still, however, the issue of the image contribution made by the thermal diffuse scattering within the objective aperture. Some direct information about this, obtained by comparing holographic centre-band with side-band images (Boothroyd & Dunin-Borkowski 2004), suggests that its effect is significant, tending to reduce contrast in the usual underfocus imaging conditions and to increase it in the more recently adopted overfocus condition in ACEM. Apart from these basic problems, the discussion at the meeting about background intensity levels, which followed Urban’s presentation, indicated that further attention to experimental procedures could be needed for further progress towards full quantification of conventional images.

In HAADF-STEM imaging, the observed improvements in resolution and contrast as well as the significant increase in probe current consequent on the larger illumination aperture can all be quite well accounted for in computer
simulations of the total (mainly thermal diffuse scattered) intensity reaching
the annular detector (Dwyer & Etheridge 2003). Normalization of the image
intensities by deflecting the incident beam into the annular detector is
experimentally difficult because of the enormous intensity ratio involved but has
been achieved recently using a 24 bit detection system (LeBeau & Stemmer 2008).
Analysis of the results obtained shows good agreement with theory (LeBeau et al.
2008), although still with some recourse to an adjustable source size. The door
may therefore be open for fully quantitative STEM work.

(b) Understanding the power and limitations of atomic column imaging

Off-axis channelling, with wave intensities concentrated in the regions between
atomic columns, was established many years ago as a key component in
diffraction contrast imaging of thick crystals. Axial channelling, with the
wave intensities concentrated on the atomic columns and therefore strongly
attenuated by diffuse scattering, has become the basis of virtually all STEM
or CTEM high-resolution imaging where much thinner crystals are used.
Adequate electron optical resolution is then usually a sufficient as well as a
necessary condition for the atomic columns to be clearly resolved at planar
interfaces or linear defects such as dislocations lying along a prominent
crystal axis.

Even before the advent of ACEM, there were grounds for suspicion about
the complete reliability of projection imaging, particularly for dislocation cores.
Significant structural variations along the dislocation line can arise from a
number of effects, but may not be strongly visible in the usual images. Surface
relaxations such as the Eshelby twist at a screw dislocation, although easily
noticed in diffraction contrast imaging (Tunstall et al. 1964), are much less
perceptible in HREM images (Sigle 1999; Kisielowski et al. 2006). Minimum
energy configurations for partial dislocations do not in general lie parallel to
one another or to the thin film normal (Hazzledine et al. 1975). Dislocation
core reconstruction effects, although clearly visible in sophisticated plan view
structure imaging (Kolar et al. 1996), cannot be detected in projection imaging or
diffraction (Koch et al. 2000). The increased numerical apertures now available for
imaging often go beyond the angular range for channelling along atomic columns
and so may be more sensitive to failures of the method as well as providing
opportunities for extending it.

Although there have been many simulations of atomic column images at
defects, both for CTEM and for HAADF-STEM, these have mostly employed
the slice method that, although reliable and efficient, is less useful than the
Bloch wave method in revealing the basic scattering processes (Peng et al. 2004).
The Bloch wave current flow is normal to the dispersion surface and so, for a
perfectly aligned crystal with CTEM plane wave illumination, is always parallel
to the axis with no spreading within the crystal in the absence of defect or
thermal diffuse scattering. With the increased imaging angles of ACEM, the
most highly excited, strongly channelling Bloch wave state (with a rather flat
dispersion surface) is now collected to higher diffraction orders and hence with
sharper spatial resolution. High-order Bloch waves with large transverse energy
and hence poor channelling properties can also now be imaged more strongly, but
their excitation is very weak.
3. Lateral delocalization mechanisms

Electron optical or what may perhaps be termed Scherzer delocalization has so far tended to dominate discussions of spatial resolution, particularly in the context of elastic imaging of atomic columns. However, as the light microscopists discovered more than a century ago, the optics can rarely if ever be separated from the scattering processes in the specimen. Because of the great strength of electron interactions, we have to worry about inelastic as well as multiple elastic scattering in all but the thinnest specimens. These interactions constitute many other sources of delocalization, which it may be helpful to list if a systematic attempt is to be made to minimize them in the new era of ACEM.

Bragg delocalization might be a suitable term to describe the somewhat complex effect of elastic scattering crudely described by the lateral beam spreading of order 2t\textit{n}\theta_B for the \textit{n}th order Bragg reflection in a crystal of thickness \textit{t}. Fortunately, in crystals, this potentially very serious effect is substantially mitigated by the dynamical scattering and channelling processes. As mentioned earlier, the dispersion surfaces are quite flat in the vicinity of Bragg conditions with a corresponding current flow parallel to the Bragg planes. This forms the basis for the column approximation in diffraction contrast imaging. Close to a major crystal axis, two-dimensional channelling along atomic columns underpins virtually all recent work in HREM whether in conventional or scanning mode. What ‘close’ means depends on the strength of the attractive potential in the column of atoms but, particularly in low atomic number elements, the increased range of angles, and hence transverse momentum, associated with ACEM means that the channelling condition may be broken leading to beam spreading, cross-talk between atomic columns, etc.

We now have to explain why these attractive channelling properties are apparently so robust in the presence of an abrupt scattering centre, such as a planar interface that can certainly provide to the fast electron large changes \Delta q_x in transverse momentum, i.e. in the direction normal to the interface. Unless, however, the interface introduces some new periodicity in the \textit{z}-direction parallel to the electron beam, the available momentum change \Delta q_z in this direction will be very small (limited in magnitude by the uncertainty principle to 1/\textit{t}, where \textit{t} is the crystal thickness). Such small values of \Delta q_z would seem to rule out most interband Bloch wave transitions and even most intraband transitions apart from those on very flat branches of the dispersion surface where there will be little effect on the current flow direction. The apparent absence of Bragg delocalization effects in ACEM images of sharp interfaces may then perhaps be explained. Careful examination of interface images in crystals of different thicknesses would certainly be interesting however.

Debye delocalization is the image blurring effect owing to thermal vibrations of atoms and is described in a crystal by the Debye–Waller factor, reducing the amplitude of higher order Bragg reflections. In some nanostructured specimens, particularly when there is a low frequency (and therefore highly excited) vibration mode, there may be a quite directly visible effect in the image. For instance, by measuring the image blurring at various distances from the free end of a single-walled carbon nanotube, Treacy \textit{et al.} (1996) determined the cantilever mode frequency. With aberration correction, it may be possible to extend this technique to smaller amplitude vibrations, particularly, for example, the torsional...
Figure 1. Use of a hollow cone illumination aperture in STEM to provide high spatial resolution valence excitation spectroscopy by selecting only high-momentum transfer events.

vibrations of a nanoparticle lying on a support. The increased beam current associated with aberration-corrected operation may of course be an added source of vibration.

Coulomb delocalization arises because electron scattering is governed not by a point force but by a long-range interaction. For elastic scattering, this effect is already included in the atomic scattering amplitude, but for inelastic scattering of an electron of velocity $v$ with energy loss $\Delta E = \hbar \omega$, it can be described somewhat crudely by the classical Bohr impact parameter $\gamma v / \omega$, where $\gamma = 1 / \sqrt{1 - v^2 / c^2}$. An uncertainty principle argument can be used to relate this to the minimum momentum transfer in the collision. Although this impact parameter has a general significance in inelastic scattering, the simple classical trajectory excitation model is strictly only valid when all of the inelastic scattering is collected (Ritchie & Howie 1988). In valence excitation where Coulomb delocalization is most serious, this requirement is mostly met without too much difficulty particularly in CTEM imaging, and many results have been interpreted with simple local dielectric response functions $\varepsilon(\omega)$. In STEM energy loss imaging, Muller’s contribution at the symposium emphasized the need to use a spectrometer-acceptance aperture significantly larger than the illumination aperture for the simple classical impact parameter model to be valid. In core-level excitation, the failure of the classical model is sometimes apparent, and a more complete quantum mechanical theory using the mixed dynamic form factor (Kohl & Rose 1985) is needed.

The feasibility of reducing Coulomb delocalization by collecting only the higher momentum transfer components of the collision has already been demonstrated for core excitations by Tafto & Krivanek (1982) and for valence excitations by Muller & Silcox (1995), but the greater angular acceptance range of ACEM inelastic imaging clearly creates further opportunities. The hollow cone illumination mode depicted in figure 1 illustrates one possible
scheme. At the typical accessible scattering angle $\theta$ of 30 mrad corresponding to a momentum transfer $\hbar q = mv\theta$, the valence electron excitations would typically lie in the Compton scattering regime with $\Delta E = (\hbar q)^2/2m = 50$ eV. Plasmon loss contributions would also be collected because of large angle quasi-elastic scattering, but could be eliminated by setting a suitable spectrometer energy-acceptance window. Detailed interpretation of the results would again probably depend on the quantum mechanical theory rather than on the classical excitation theory.

The Coulomb delocalization effect governs not only the direct interaction of the specimen with the incident electron beam, but also the interactions within the specimen such as the coupling of plasmon and other valence excitation modes at different points of a nanostructure. For instance, the modes of wavevector $q$ at the two interfaces of a sandwich structure A–B–A with an intermediate layer B of thickness $d < q^{-1}$ will be strongly coupled, leading to symmetric and antisymmetric combinations with energy splitting $\Delta E$. For the splitting of the surface plasmon loss of energy $\hbar \omega_s$ at a thin slab of material of thickness $d$, we find $\Delta E = \hbar \omega_s \exp[-\omega_s d/\gamma v]$. Paradoxically, therefore, an improvement in energy loss resolution could produce greater sensitivity to more distant couplings. When detectable, these Coulomb couplings between excitations at different places also clearly introduce an extra $q$-dependence in the excitation spectra.

Landau delocalization arises because the probability of inelastic excitation at any point in the specimen can be influenced by the scattering of these excitations on the surrounding structure within a distance governed by their mean free path. This phenomenon is exemplified by the extended X-ray absorption fine structure effect and its analogue in the above-edge structure at core losses in electron energy loss spectra. A particularly extreme form of this Landau delocalization effect occurs in the case of Cherenkov radiation, in which the mean free path is very long. Cherenkov delocalization has been observed in complex media (Zabala et al. 2003), but, in principle, can easily be eliminated by using either the lower beam energies or the larger momentum transfers that are now accessible for high-resolution work owing to ACEM.

Except for the Cherenkov modes, Landau delocalization effects are completely missed in the widely used valence electron excitation theory based on a momentum-independent dielectric response function $\epsilon(\omega)$. Although the excitations have a lifetime, they have no velocity and thus the mean free path is zero. The use of a momentum-dependent response function $\epsilon(\omega, q)$ removes these shortcomings, although the consequent changes in the energy loss function are usually small. This modification is adequate only for a homogeneous specimen, however, and assumes that the effect on the excitation generated at a point $r$ of the surrounding structure at $r'$ depends only on $|r - r'|$. More generally, the response depends on $r$ and $r'$ separately. The appropriate response function in a sufficiently large crystalline region near $r$ would then be $\epsilon(\omega, q, q' + g)$, but in the more generally inhomogeneous specimens relevant to microscopy, a real space function $\epsilon(\omega, r, r')$ will usually be needed.

Apart from the case of Cherenkov excitation already mentioned, Landau delocalization effects have not so far been very apparent in valence excitation. They could perhaps be more significant in ALOOF beam generation of excitations in nanostructures if their scattering on surface steps and other features could be detected. They should also be noticeable in the Compton excitation mode.

Phil. Trans. R. Soc. A (2009)
previously depicted in figure 1 since the excited electrons then have energies very similar to those responsible for near edge structure in core excitation spectroscopy.

4. Longitudinal delocalization: depth sectioning

Electron microscopists may look with some envy on the tremendous success in three-dimensional reconstruction of complex structures achieved with scanning confocal optical microscopy. Particularly striking is recent static and dynamic imaging of defects such as dislocations in colloid crystals and of ‘amorphous’ colloid structures (Schall et al. 2007a,b). A key factor in the success of the reconstruction scheme has been to attain sufficiently weak (i.e. kinematical) scattering conditions by using in the interstices between the colloid spheres a fluid differing from them only slightly in refractive index. Sample thicknesses are typically equivalent to about only 10 colloid layers and thus far less likely to generate the image overlap problems familiar in electron microscopy. However, another special factor in successful applications of confocal scanning optical microscopy to thicker biological samples has often been the use of multi-photon excitation techniques.

The large scattering angles associated with narrow depth of focus in kinematical scattering arise most easily in HAADF-STEM imaging, although also equivalently in TEM imaging with hollow cone illumination. Early work (Gibson & Howie 1979) indicated the possible advantages of hollow cone illumination in reducing overlap problems by defining a more restricted coherence volume in the dark field imaging of amorphous material structure. With the later development of fluctuation microscopy (Treacy & Gibson 1996), it became apparent that the immediate usefulness of this approach for amorphous structure determination lay in imaging at moderate rather than at high spatial resolution (for a recent review, see Treacy et al. 2005).

The 40 mrad scattering angles now accessible for imaging in ACEM imply a kinematical depth of field as little as 3 nm for a small scattering centre and have provoked fresh investigations of the opportunity for simple means of high-resolution depth discrimination. It has indeed proved possible to identify individual dopant atoms at different depths in semiconductors by HAADF-STEM imaging (Voyles et al. 2002; Varela et al. 2004; Van Benthem et al. 2006). In Bleloch’s contribution at the symposium, the depth-resolved imaging of individual gold atoms in a silicon nanowire was demonstrated. Unfortunately, multiple scattering effects, particularly channelling, impose their own depth signature and can greatly complicate any simple depth-sectioning procedure by focal scanning. One option may be simply to choose orientations that avoid these channelling conditions (Xin et al. 2008). Another possibility might be to use large-angle hollow cone illumination centred on a crystal axis (Cosgriff & Nellist 2007). More ambitiously still, Nellist’s contribution at the symposium described the use of confocal electron microscopy using a doubly corrected CTEM/STEM instrument.

An even more radical confocal procedure for high-resolution depth sectioning in crystals, including possibly thicker crystals, would be deliberately to exploit either on-atomic column or between-atomic column channelling on two different crystal axes as indicated in figure 2. To maintain channelling conditions, particularly for
between-column channelling, the illumination and imaging apertures should not be too large. However, each aperture either has to operate at some angle to the microscope axis or a magnetic deflection system has to be employed as in figure 2. For these experiments, positional sensitivity arises from the intersection of the two channelling beams, and depth sectioning would be provided by a specimen stage with three-directional piezo control. It may be seen as yet another consequence of the great success of sophisticated electron optical instrumentation that such a proposal may no longer be regarded as entirely fanciful.

For many samples, Z-contrast tomography must be the most obvious and already fully operational competitor to these various depth-sectioning schemes and can yield a three-dimensional resolution of approximately 1 nm³ (Arslan et al. 2005). Whether some of them prove to be superior to tomography in special cases, perhaps because of higher spatial resolution or because a smaller dose can be used in beam-sensitive samples, is possible (Borisevich et al. 2006), but still not entirely clear.

5. Current challenges and historical parallels

Staggering numbers of these costly and sophisticated new ACEM instruments have now been installed worldwide sometimes through the generosity of private donors or regional agencies as well as through the more usual national research councils. This massive initial funding frequently fails to cover on-going costs,
raising serious questions about continuing maintenance and user training. The high running costs have prompted an urgent search for customers particularly in the chemistry, materials and electronics industries. A substantial move into various forms of in situ microscopy, going far beyond the traditional DIY mode, is already apparent. A beneficial stimulus has thus been provided to the manufacturers of special stages, high-pressure cells and facilities for combined electron microscopy and scanned probe microscopy or for specimen manipulation. As already mentioned, the successful development of chromatic aberration correctors should allow high-resolution imaging at longer working distances in larger specimen volumes. With the addition of efficient phase shifting devices, this combination could be very attractive for biological imaging of easily damaged samples.

Expanded training programmes at all levels are urgently required. The levels of skill required for specimen preparation, microscope adjustment and operation, image analysis and interpretation have all, so far at least and contrary to some assertions, been significantly raised by ACEM. There is also an educational challenge at the more basic level of the occasional user. This poses an immediate problem for every ACEM facility, but also needs the involvement of the whole microscopy community. The active involvement of the manufacturers in assisting these training programmes and in designing systems with improved user friendliness will clearly be crucial.

In addressing many of these issues, useful lessons may be available from the history of optical microscopy over the past 200 years (e.g. Bradbury 1967). The nineteenth century development of lenses for optical microscopes culminated in the celebrated apochromat—an objective lens colour corrected at three different wavelengths and with operating distances chosen to minimize spherical aberration at two intervening wavelengths. Such a lens could cost as much as £50—a formidable sum for the individual microscopist of the time to find. For optimum performance, the lens was used under oil immersion conditions with the overall design and operation taking account of effects beyond lens optics such as the refractive properties of the sample and cover glass. In the hands of expert microscopists, but they had to be expert, many highly significant observations were made particularly in bacteriology. Errors in the interpretation of earlier observations were identified such as the so-called globular theory to explain image features that were mainly diffraction and defocus artefacts. For many microscopists, however, these impressive microscopes were simply too expensive, sophisticated and difficult to operate. With hindsight, this development in high-resolution optical lens design is seen at least as a temporary high water mark in the discipline, which became increasingly clear with the advent of electron microscopy.

We may recognize some parallels between optical history and the current situation in ACEM. With present day image computation facilities, electron microscopists can cautiously hope that they have been saved from the more egregious errors of misinterpretation of the kind that were caused two centuries earlier by uncorrected and poorly understood aberrations in glass lenses. The more recent remarkable progress in optical instrumentation can, nevertheless, be a source of inspiration to us as has already been mentioned for phase contrast imaging and depth sectioning confocal microscopy. Even on the issue of cost and sophistication, it is interesting to note that the most elaborate optical lenses now
used in semiconductor device lithography and operating in the UV region can each cost as much as the whole of the TEAM project described at the symposium by Dahlem.

I am grateful to a number of colleagues for useful discussions, particularly Peter Evennett, Peter Nellist, Stephen Pennycook and John Spence.

References


