Direct Methods in High Resolution Electron Microscopy

D. Van Dyck
University of Antwerp, Belgium

M. Op de Beeck
University of Antwerp, Belgium

Utah State University
MERRILL-CAZIER LIBRARY

DIRECT METHODS IN HIGH RESOLUTION ELECTRON MICROSCOPY

D. Van Dyck[*] and M. Op de Beeck

University of Antwerp (RUCA), Groenenborgerlaan 171
B-2020 Antwerpen, Belgium

Abstract

New approaches are proposed to retrieve the wavefunction at the object and from this, to retrieve the projected structure of the object. The wavefunction is retrieved by capturing images at a series of closely spaced focus values and to process the whole 3D data. The structure of the object is retrieved using a formalism based on electron channelling.

Key Words: direct methods, wavefunction reconstruction, image reconstruction, phase retrieval, high resolution electron microscopy

* Address for correspondence:
D. Van Dyck
University of Antwerp (RUCA)
Groenenborgerlaan 171
B-2020 Antwerpen, Belgium
                Phone No.: +32-2-2180258

Introduction

We are living in a very exciting period for structural research using HREM. Indeed, the possibility to "see" the individual atoms of which matter is constituted seems within reach. Recent technological improvements permit a resolution of about 0.1 nm to be obtained. However, the potential power of the technique is still severely limited by the problem of quantitative interpretation of the images. For instance, the use of computer simulation images requires much a priori knowledge which makes HREM very dependent on other techniques. The situation can be compared with the early days of X-ray diffraction. Recent developments make it possible to retrieve the object structure directly from the electron micrographs.

As is well known, the coherent transfer of the wave function in the spatial frequency domain is given by a multiplication with the phase transfer function (PTF) of the electron microscope (Figure 1), which can be considered as a complex band filter. At optimum focus, the phase aberration caused by the spherical aberration can be complemented by a slight underfocus of the objective lens, causing the phase transfer function to be nearly constant over a wide range of spatial frequencies. In this regime, the electron microscope acts as a phase contrast microscope. If the object is thin, the image then directly reveals the phase of the object, which is proportional to the projected potential. The resolution of the microscope is then called interpretable resolution ρ_s (structural resolution or point resolution) and is given by the first zero of the PTF. Another type of resolution is the information limit ρ_I given by the point where the information disappears in the noise. This limit is caused by the damping of the PTF due to spatial and temporal incoherence. No information beyond ρ_I can be retrieved from the image.

A promising way of increasing the resolution is by restoring the information that is present between ρ_s and ρ_I and that is still present in the image, albeit with the wrong phase. For this purpose, image processing will be indispensable. In that case, the resolution will be determined by ρ_I. ρ_I can be improved drastically by using a field emission gun (FEG) which reduces the spatial as well as the temporal incoherence. With the present technology, an information limit $\rho_I = 0.1$ nm is within reach.

The ultimate resolution however is determined by the object itself. The intensity of the scattered information

(beams) decreases with increasing diffraction angle. This is a consequence of the requirement for energy conservation (Ewald sphere) and Heisenberg's uncertainty relation [1]. Combining these results with the voltage limit for displacement damage, it is found that the ultimate resolution will always be of the order of the Bohr diameter (0.1 nm). For a thorough discussion on the ultimate resolution we refer to [2].

Image Interpretation

The most difficult problem in high resolution electron microscopy (HREM) is the problem of the interpretation of the images. At present, the only way for a reliable interpretation of the electron micrographs is by comparing them with computer simulations calculated for plausible trial-structures. However this technique is very tedious, requires a number of usually unknown parameters, and can only be applied with some success if the number of possible structure models is very limited. This makes HREM very much dependent on the availability of prior information obtained from other techniques. HREM would be much more powerful if a direct method exists to extract the structural information directly from the electron micrographs.

Direct Methods

A direct method should consist of three stages. First the wavefunction in the image plane has to be reconstructed (phase problem). Then the wavefunction at the exit face of the object has to be calculated. Then finally from this the structure of the object has to be retrieved.

Phase retrieval [3]

The phase problem can be solved mainly in two ways, by using holography or by using the focus as an external controllable parameter. In electron holography, the beam is split by an electrostatic biprism into a reference beam and a beam that traverses the object. Interference of both beams in the image plane then yields fringes, the positions of which yield the phase information [4] [5]. In order to assess this information one needs a very high resolution camera (CCD), a powerful image processor, and a field emission gun to provide the necessary spatial coherence. We will present another method, in which the focus is used as an extra parameter. Images are captured at very close focus values so as to collect all information in the three-dimensional image space. Each image contains linear information and nonlinear information. By Fourier transforming all 3D image space, the linear information of all images is superimposed onto a sphere in reciprocal space, which can be considered as an Ewald sphere (Figure 2). By separating this linear information the phase can be retrieved [3]. This can be proven as follows.

Consider an image plane at a particular focus value in which we want to retrieve the phase. For convenience we choose the origin of focus in that plane. Writing the wavefunction as a Fourier integral, we have

$$\psi(\boldsymbol{R},0) = C + \int_{g\neq 0} \phi(\boldsymbol{g}) \exp(2\pi i\boldsymbol{g}.\boldsymbol{R}) \ d\boldsymbol{g} \qquad (1)$$
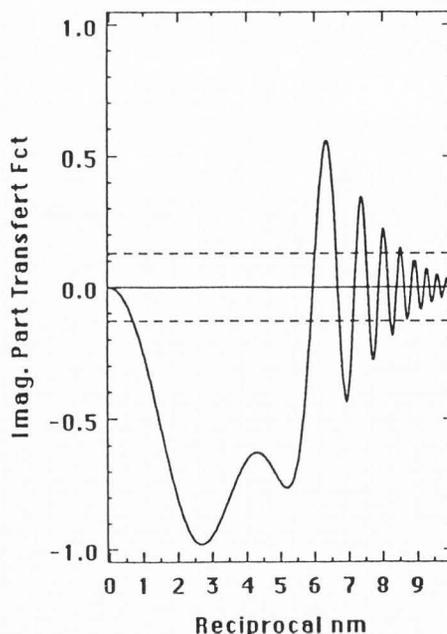
Figure 1. Phase transfer function for a 300 keV instrument ($Cs = 0.7$ nm, $C_c = 1.3$ nm).

$\phi(\boldsymbol{g})$ are the Fourier components. We have separated the zeroth order component (constant term).

The wavefunction at a particular focus value z can be obtained from (1) by propagation, i.e.

$$\psi(\boldsymbol{R},z) = C + \int_{g\neq 0} \phi(\boldsymbol{g}) \exp(2\pi i\boldsymbol{g}.\boldsymbol{R}) \ \exp(-i\pi\lambda g^2 z) \ d\boldsymbol{g} \qquad (2)$$

3D Fourier transforming the image intensity $|\psi(\boldsymbol{R},z)|^2$ now yields, using (2),

$$I(\boldsymbol{g},\xi) = |C|^2 \delta(\boldsymbol{g}) + C^*\phi(\boldsymbol{g}) \delta\left(\xi - \frac{\lambda g^2}{2}\right)$$

$$+ C\phi^*(-\boldsymbol{g}) \delta\left(\xi + \frac{\lambda g^2}{2}\right)$$

$$\int_{\substack{g'\neq 0 \\ g+g'\neq 0}} d'\boldsymbol{g} \ \phi^*(\boldsymbol{g}') \ \phi(\boldsymbol{g}+\boldsymbol{g}') \ \delta\left\{\xi - \frac{\lambda}{2}\left[(\boldsymbol{g}+\boldsymbol{g}')^2 - g'^2\right]\right\} \qquad (3)$$

where δ are Dirac functionals and $\boldsymbol{g}$ and ξ are the conjugates of $\boldsymbol{R}$ and z. The first term on the r.h.s. of (3) only yields a contribution in the origin. The second and third linear terms give a sharply peaked contribution which is located on a paraboloid in reciprocal space which can be considered as the Ewald sphere in vacuum (Figure 2). The last term gives a contribution which is more continuously spread through reciprocal space. It is immediately clear that by selecting the information concentrated on the paraboloid one directly obtains $\phi(\boldsymbol{g})$ and $\phi^*(\boldsymbol{g})$ so that from (1) the total wavefunction
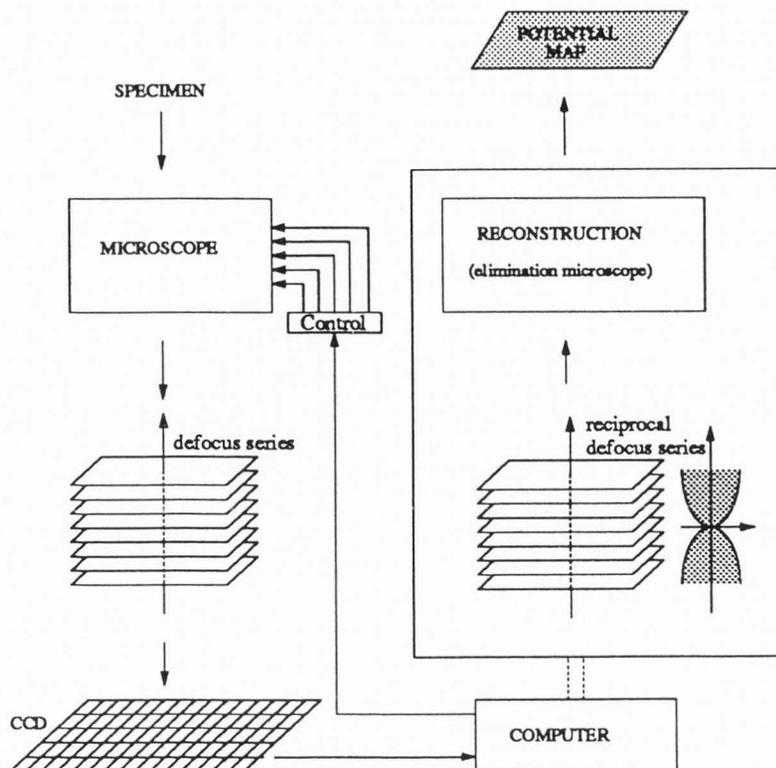
Figure 2. Schematical representation of the phase retrieval procedure. The paraboloid which contains the linear information in reciprocal space is also shown.

at focus 0 is retrieved. In principle, this can by done by taking a nearly continuous series of images at very small focus intervals, 3D Fourier transforming and selecting the sphere. However, this procedure is rather impractical. Hence we proceed as follows. We take a series of images at focus values $z_1, z_2, z_3 \dots$. The focus interval is of the order of 3 nm. Each of the images $I(R, z_n)$ is then transformed into $I(g, z_n)$. Finally we calculate the series $\sum_{n} I(g, z_n) \exp(-i\pi\lambda g^2 z_n)$.

In this way, the sphere for $\xi > 0$ is in a sense projected in the plane, apart from a known weighting factor. We can do the same for $\xi < 0$. From these data it is easy to calculate $\phi(g)$. In a sense, all the images are back propagated to zero focus, where the linear part of each image superimposes and increases with respect to the nonlinear part. However, as seen from (3), the integral also gives a contribution to the sphere which may influence the results. This contribution can be taken corrected by using (3) in an iterative way. Another advantage of this method is that it is relatively easy to compensate for the effect of chromatic aberration. It is well known that chromatic aberration results from a spread in the focus due to instabilities in voltage and lens current. Hence the image intensity is convoluted with a focal spread function $I(R, z) = I_0(R, z) * f(z)$. In reciprocal space, the convolution product is a direct product with the Fourier transform of $f(z)$ $I(g, \xi) = I_0(g, \xi) . f(\xi)$. If the spread function f is known, it is easy to compensate for this effect by

dividing by $f(\xi)$. Since this has to be done only at the sphere, blowing up effects are largely reduced. Another advantage of the method is that, since the contribution of the noise is more homogeneously distributed in space, the selection of the sphere automatically increases the signal to noise ratio.

Structure retrieval [6]

The final step consists in retrieving the projected structure of the object from the wavefunction at the exit face. If the object is thin enough to act as a phase object, the phase is proportional to the electrostatic potential of the structure, projected along the beam direction so that the retrieval is straightforward. If the object is thicker, the problem is much more complicated. However, if the object is a crystal viewed along a zone axis, the incident beam is parallel to the atom columns. It can be shown that in such a case, the electrons are trapped in the positive electrostatic potential of the atom columns, which then act as pipes. This effect is known as electron channelling and can be explained as follows.

If the crystal object is perfectly oriented along a zone axis, the electrons are trapped in the positive potential of the columns. The columns then, in a sense, act as channels for the electrons. If the distance between the columns is not too small, a one-to-one correspondence between the wavefunction at the exit face and the column structure of the crystal is established. Within the columns, the electrons oscillate as a function of depth without however leaving the column (Figure 3). Hence the classical picture of electrons traversing the crystal as plane-like waves in the direction of the Bragg beams, which historically stems from X-ray diffraction, is in fact equivalent but misleading. It is important to note that channelling is not a property of a crystal, but occurs even in
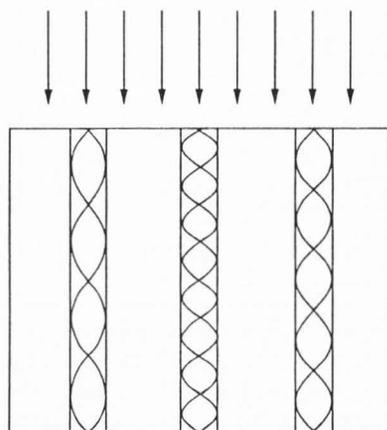
Figure 3. Schematical representation of the electron channelling.

an isolated column and is not much affected by the neighbouring columns, provided the distance is not too close.

The channelling can best be understood as follows [6] [7] [8]. Assuming normal incidence and taking the z axis perpendicular to the specimen foil, the high energy equation describing the dynamical electron scattering in real space is equivalent to the time-dependent Schrödinger equation

$$- \frac{\hbar}{i} \frac{\partial \phi}{\partial t} (R,t) = H \phi(R,t) \qquad (4)$$

in which the time is replaced by the depth z using $t = mz/hk$ and in which the Hamiltonian is given by

$$H = - \frac{\hbar^2}{2m} \Delta - e \ U(R,t) \qquad (5)$$

with $U(R,t)$ the electrostatic crystal potential, m and k the relativistic electron mass and wavenumber. This can be understood by assuming that in the direction of propagation (z axis) the high energy electron behaves as a classical particle with a constant velocity equal to hk/m. In this way the z axis plays the role of a time axis. We will further on use t instead of z.

It is easy to verify that the solution of (4) which obeys the boundary condition $\phi(R,0)$ is now given by

$$\phi(R,t) = 1 + \sum_n C_n \ \phi_n(R) \left[\exp\left(-\frac{i}{\hbar} E_n t \right) - 1 \right] \qquad (6)$$

with $\phi_n(R)$ the bounded eigenstates of the Hamiltonian and E_n its energy ($E_n < 0$)

$$H \ \phi_n(R) = E_n \ \phi_n(R) \qquad (7)$$

In case of a rotationally symmetric situation, which occurs when the incident beam is exactly parallel to the single column, only symmetrical states can be excited (to be compared with s states). If the atoms are not too heavy and the accelerating potential is not too high only one bound state

appears so that

$$\phi(R,t) = 1 + C \ \phi(R) \ \exp\left(\frac{-iEt}{\hbar} - 1 \right) \qquad (8)$$

From this it is clear that the electron wavefunction varies perfectly periodically with depth, the periodicity being determined by E, which is related to the mass of the column. From (8) it is clear that $\phi(R)$ represents a kind of impulse response function for that particular column. Its Fourier transform can then be considered as the maximum scattering factor for that column. The scattering factor varies periodically between zero and this maximum. This effect is known as "dynamical extinction". In a sense, the resolution limited by the object then also varies periodically with depth. The best resolution is obtained for those values for which (8) becomes maximal. However, the variation is different for different types of columns.

In case of an assembly of columns, located at positions R_i, the total wavefunction is now from (8)

$$\phi(R,t) = 1 + \sum_i C_i \ \phi_i(R-R_i) \ \exp\left(\frac{-iE_i t}{\hbar} - 1 \right) \qquad (9)$$

(8) now in principle allows to retrieve the object structure, i.e. the type and position of each column, once the wavefunction at the exit face is known. Indeed, from (8)

$$\frac{\text{Re} \left(\phi(R,t) - 1 \right)}{\text{Im} \left(\phi(R,t) - 1 \right)} = \text{tg} \ \frac{Et}{2\hbar} \qquad (10)$$

should be constant over the column area.

From this, the energy E can be calculated, which yields a measure for the "weight" of the column. Substitution into (8) then yields $C \ \phi(R)$ from which the form of the potential, and hence also the exact column position can be obtained.

Results

Figure 4 shows the results of the retrieval procedure for a crystal of $Ba_2NaNb_5O_{15}$ and Figure 5 shows the results for a model of amorphous Si. In both cases, images are simulated, including noise, which are then used again to retrieve the structure (left: image at optimum focus; centre: retrieved structure; right: original structure). By comparing the retrieved with the original structure, the merits of the method can be evaluated. It is clear from these results that the position of the projected atoms can be retrieved with an accuracy of about 1 Å. This is particularly important for amorphous structures because thus far, no technique exists to obtain reliable structural information of this kind.

Another interesting point to make is that, within a direct method, the concept of resolution becomes completely different to the original Rayleigh concept. In fact, if the types of the atoms are known, only their positions have to be determined and resolution is then reduced to the accuracy with which these positions can be obtained.
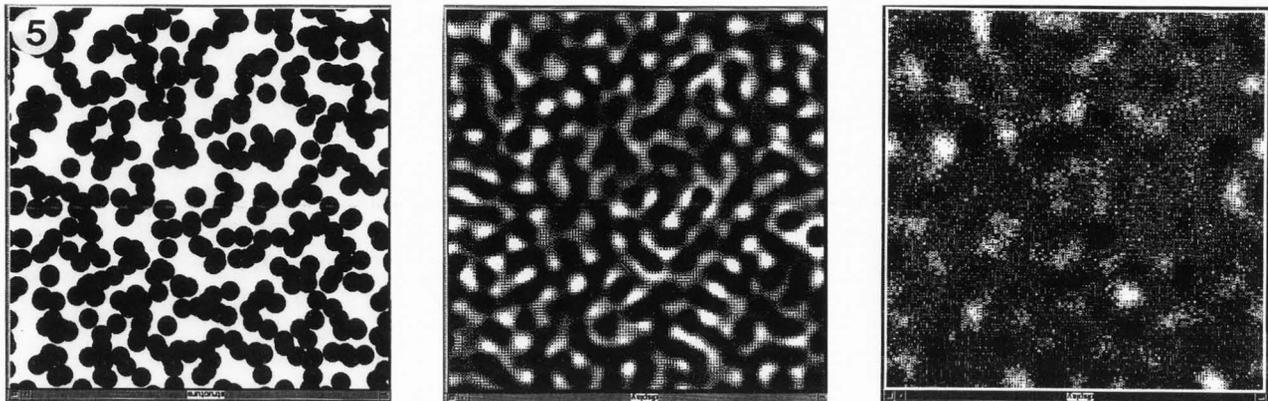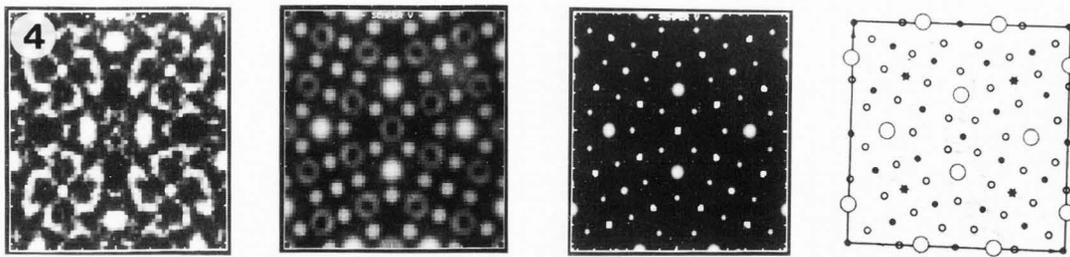
Figure 4. Structure of $Ba_2NaNb_5O_{15}$ as retrieved from simulated images. The simulations are performed using the transfer function of Fig. 1 adding 10% noise. From left to right: image at optimum focus; retrieved phase at object; retrieved structure; structure model.

Figure 5. Structure of amorphous Si as retrieved from simulated images (object thickness 2 nm). The simulations are performed using the transfer function of Fig. 1 adding 10% noise. From left to right: image at optimum focus; retrieved phase at object; structure model.

For instance, in a crystal, the number of projected positions to be determined cannot be larger than the number of Fourier components (beams) which constitute the image. This leads to a critical density of atoms per unit area beyond which the atom positions cannot be discriminated. In this view, resolution is reduced to a critical distance below which atoms cannot be discriminated and above which the positions can be determined with an accuracy, which is only limited by the accuracy of the recording.

Instrumentation

In order to put this method into practice one needs a medium voltage high resolution electron microscope, equipped with a field emission gun (FEG), high resolution CCD camera with a high DQE value, directly coupled to a fast image processing system. The microscope should be aligned in an automatical way. Recently an European Brite-Euram project has been set up, which is funded by the European community and in which the ultimate goal is to obtain direct 1 Å structural information using holography and focus variation. Furthermore, the FEG allows the use of all illumination angles whereas the CCD collects all electrons either in image space or in diffraction space. In the future it would be desirable to equip such an instrument with an energy filter above and below the specimen. In this way nearly all information that can be obtained with electrons can be assessed.

References

1. D. Van Dyck (1988). On the ultimate resolution attainable with HREM. In: Proceedings 9th European Congress on Electron Microscopy (York), Inst. Phys. Conf. Ser. nr. 93, Vol. **2**, p. 349-350.

2. D. Van Dyck, A.F. De Jong (1992). Ultimate resolution and information in electron microscopy. Ultramicroscopy, in the press.

3. D. Van Dyck (1990). New direct methods for phase and structure retrieval in HREM. In: Proceedings XIIth International Congress for Electron Microscopy (Seattle), San Francisco Press Inc., p. 26-27.

4. H. Lichte (1986). Electron holography approaching atomic resolution. Ultramicroscopy **20**, 293-304.

5. H. Lichte (1991). Electron image plane off-axis holography of atomic structures. In: Advances in Optical and Electron Microscopy, Academic Press, Vol. **12**, 25-91.

6. D. Van Dyck (1990). A simple real-space channelling theory for electron diffraction and HREM. In: Proceedings XIIth International Congress for Electron Microscopy (Seattle), San Francisco Press Inc., p. 64.

7. K. Kambe, G. Lempfuhl, F. Fujimoto (1974). Interpretation of electron channelling by the dynamical theory of electron diffraction. Z. Natur. **29A**, 1034-1044.

8. B.F. Buxton, P.T. Tremewan (1980). The atomic string approximation in cross-grating high-energy electron diffraction. Acta Cryst. **A36**, 304-315.

Discussion with Reviewers

P. Rez: What is the range of defocus values that can be used in focal series restoration and is it affected by temporal coherence?

- Isn't the suggestion for dividing by the focal spread function numerically unstable?

Authors: The range of defocus values that can be used in focal series restoration is mainly determined by the spatial coherence. However when using a FEG, this range can be extended to several 100 nm.

- Dividing by the focal spread function is not a deconvolution in the strict sense, since it is only carried out at the paraboloid. Recent experiments have shown that it is not unstable.

P. Rez: One serious problem with focal series restoration is image registration (see E.J. Kirkland (1984). Improved high resolution image processing of bright field electron micrographs. Ultramicroscopy **15**, 151-172). What methods do the authors propose to eliminate this problem?

Authors: With the large and fast memories of modern image processors, image registration is not a serious problem. In the near future, recording can probably be done directly on optical disks.

P. Rez: Are there problems with the method in large unit cell materials, such as complex oxides, when the Fourier Period is large and only a small part of the "paraboloid" is sampled?

Authors: The method can in principle even be used (and has already been) for aperiodical objects. However, since the information in the oscillating part of the transfer function is highly delocalized, a large number of recording pixels (e.g. 1000 or more) is required.

P. Rez: Is it still conceivable that nonlinear contributions will lie on the paraboloid. Can the authors give an estimate of the thickness in either Au or Si (110) or (111) projections when such nonlinear terms will cause the method to fail?

Authors: The method seems to work for object thicknesses of the order of 10 nm. We do not have an estimate for the thickness in Au or Si (110).

P. Rez: If the authors are correct the problem of inverting dynamical diffraction is trivial once the complex wave at the exit surface has been recovered, as the positions of the atomic columns can be directly determined. Is this really true when atomic columns are very close to each other as in semiconductors in the (110) projection? How do the authors reconcile their views with the displacements of up to 0.3 Å found by Saxton and Smith (W.O. Saxton, D.J. Smith (1985). The determination of atomic positions in high-resolution electron micrographs. Ultramicroscopy **18**, 39-48) in studies using image simulations?

Authors: It is true that the positions of close overlapping columns may apparently be displaced. It is our belief that the channelling approach may be improved to handle this situation, for instance by using perturbation theory.

J. Rodenburg: Can you please elaborate on the question of ultimate resolution being specimen dependent. I can understand this in the case of a thin crystalline material where the Ewald sphere may unfortunately pass through a perfect minimum in the reciprocal space scattering function of the specimen, but surely this is not an absolute limitation? At worst one may need to perform several experiments at different angles of specimen tilt.

Authors: The ultimate resolution will indeed be limited by the object. The derivation in this paper however is oversimplified. In practice also inelastic scattering will influence the resolution. For a thorough discussion we refer to Ref. [2] which we for this purpose have added in proof.

W. Coene: The advantages of ultra-high resolution microscopy are in practice concerned with structure information in complex structures (like oxides) or high index zone-axis orientations, in which case the projected distance between atom columns becomes small, so that neighbouring atom columns will "feel each other" while diffracting the incident electrons. How will this affect a possible breakdown of the channelling concept? How severely is channelling affected in the case of more than one important bound state in relation (8)? What do the authors think about a fine-tuning structure reconstruction step by means of a "maximum-likelihood" iteration scheme for the highly non-linear diffraction problem (like in non-linear image reconstruction)?

Authors: If more than one bound state is present, channelling will become much more complicated. However, not only the energy E but the product Et matters, so that for thin objects in an exact zone orientation, and a not too high accelerating voltage, only one bound state will appear for most types of columns. We believe that the maximum likelihood procedure will improve the accuracy of the results but we are not sure whether the profit will balance the computation efforts.

W. Coene: The high spatial coherence of the field emission gun (FEG) is very essential in reconstruction by electron holography. Is the improvement in spatial coherence as necessary for reconstruction by focus variation, or can one maybe benefit from an intentionally reduced spatial coherence?

Authors: In focus variation the spatial coherence does not affect the applicability of the method itself but it affects the information limit and hence the ultimate resolution.

W. Coene: The authors argue in the Introduction that the ultimate resolution in HREM is determined by the object itself, and that it is limited to about 0.1 nm, which is roughly the same value as the one that can now be reached technologically using an FEG. Do the authors think that this is the final physical limit, or, in other words, is the additional possible benefit of energy filters in HREM (as suggested in the Results section) severely limited by effects of electron diffraction in the specimen?

Authors: We do believe that energy filters can only be used to improve the resolution by reducing the noise, if operated on the zero loss mode.