

Towards Quantitative Structure Determination Through Electron Holographic Methods

D. Van Dyck,* E. Bettens,* J. Sijbers,* M. Op de Beeck,[†] A. van den Bos,[‡] A. J. den Dekker,* J. Jansen,[§] and H. Zandbergen[§]

*University of Antwerp (RUCA), Groenenborgerlaan 171, B-2020 Antwerpen, Belgium; †Philips Research Laboratories, Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands; ‡Department of Applied Physics, Delft University of Technology, Lorentzweg 1, NL-2628 CJ Delft, The Netherlands; and [§]National Centre for HREM, Laboratory of Materials Science, Delft University of Technology, Rotterdamseweg 137, NL-2628 Delft, The Netherlands

The problem of quantitative interpretation of high-resolution electron microscopy is studied in the framework of parameter estimation. Ideally, quantitative interpretation means that unknown structural parameters of an object such as atom types and coordinates are determined from fitting with the experimental dataset. However, in the imaging process, the influence of these parameters is completely scrambled over a large area of the image. As a consequence, the fitting becomes a search process in the higher dimensional space of all coupled parameters. The real importance of holographic methods such as off-axis holography and focus variation is that they restore (deblur) to some extent the imaging process so as to unscramble the influence of the different model parameters. In this way the dimension of the search space becomes manageable. In this framework the concept of resolution in the sense of Rayleigh is not valid anymore, but it has to be replaced by the notice of parameter precision. In case two atoms are very close, the parameter space may become degenerate so that the atoms cannot be discriminated. The probability of this degeneracy is a function of the distance between the atoms and the dose of the imaging particles. © Elsevier Science Inc., 1999. All rights reserved.

INTRODUCTION

The ultimate goal of high resolution electron microscopy is to determine the atomic structure of an object. In this respect, the electron microscope can be considered as an information channel that carries this information from the object to the observer.

The transfer of information proceeds in three successive steps as sketched in Fig. 1. First, the electron interacts with the atoms in the object, through multiple scattering. Second, the exit wave of the object is transferred through the microscope to the image plane. This process is described by a convolution product with the impulse response function (point spread function) of the electron microscope. Because the imaging process is coherent, the exit wave as well as the impulse response function are two-dimensional complex functions with an amplitude and a phase. In the last step, the image is recorded either on photograph film or by a (CCD) camera. In this step, only the intensity of the image wave is recorded and the phase is lost. Incoherent effects are changes in the imaging conditions causing changes in the image intensity that are integrated during the time of recording.

A major problem is the interpretation of the image. Indeed, the structural information (atomic types and positions) of the object is usually hidden in the images and cannot easily be assessed. Therefore, a quantitative approach is required in which all steps in the imaging process are taken into account. Two main approaches have been followed so far in the literature: (a) the indirect approach, in which the images are simulated for various plausible trial structures of the object and compared with the experimental images; and (b) a direct approach, in which the lost phase information is retrieved using holographic techniques so as to "deblur" the effect of the microscope and to reveal directly the atomic structure of the object.

In this paper we will discuss the problem of quantitative interpretation of the highresolution images in the framework of parameter estimation. It is shown that both direct and indirect methods fit within this framework. But only direct methods can make quantitative structure determination possible for completely unknown objects. We will also discuss the problem of resolution in the same context.

QUANTITATIVE IMAGE INTERPRETATION

In principle, one is not interested in highresolution images as such, but in the structure of the object under study. High-resolution images are to be considered as data planes from which the structural information has to be extracted in a quantitative way. Ideally, this should be done as follows: one has a model for the object and for the imaging process (Fig. 1), including electron object interaction, microscope transfer, and image detection. The model contains parameters that have to be determined by the experiment. The parameters can be estimated from the fit between the theoretical images and the experimental images. The goodness of the fit is evaluated using a criterium of goodness of fit such as likelihood, mean square difference or R-factor (cfr. X-ray crystallography). For each set of parameters of the model, one can calculate this goodness of fit, so as to yield a fitness function in parameter space. The parameters for which the

goodness of fit is maximal, then yields the best estimates, which can be derived from the experiment. In a sense, one is searching for a maximum (or minimum depending on the criterion) of the criterion of goodness of



FIG. 1. Due to lens imperfections, the recorded images are a blurred representation of the crystal structure. Here, an analogon is shown, using real impulse response functions. (a) Crystal structure; (b) crystal potential; (c) schematic representation of the impulse response function; and (d) blurred image of the crystal potential due to the lens imperfections.

fit in the parameter space, the dimension of which is equal to the number of parameters.

The object model that describes the interaction with the electrons consists of the assembly of the electrostatic potentials of the constituting atoms. Because for each atom type the electrostatic potential is known, the model parameters then reduce to atom numbers and coordinates, Debye Waller factors, object thickness, and orientation (if inelastic scattering is neglected). The imaging process is characterized by a small number of unknown (or not exactly known) parameters such as defocus, spherical aberration, etc.

A major problem is now that the structural information of the object can be strongly delocalized by the image transfer in the electron microscope (Fig. 1) so that the effect of the structural parameters is completely scrambled in the high-resolution images. For instance, if the position of one atom in the object is changed, this affects the image over a large area. Due to this coupling, one has to refine all parameters simultaneously, which poses a combinatorial problem. Indeed, the dimension of the parameter space becomes so high that one cannot use advanced optimization techniques such as genetic algorithms, simulated annealing, tabu search, etc., without the risk of ending in local maxima. Furthermore, each evaluation of the criterion of the goodness of fit requires a full image calculation so that the procedure is very cumbersome. The problem is only manageable if the object is a crystal with a very small unit cell and a small number of object parameters [1–3], or if sufficient prior knowledge is available to reduce the number of unknown parameters to a few. In X-ray crystallography, this problem can be solved by using direct methods or maximum entropy methods, which provide a pathway toward the global maximum in parameter space. In high-resolution electron microscopy, this problem can be solved by deblurring the information, so as to unscramble the influence of the different object parameters in the image. In this way, the structural parameters can be uncoupled and the dimension of the parameter space reduced. This can be achieved in different ways: high-voltage microscopy, correction of the microscopic aberrations, or holographic methods.

Holographic methods have the particular advantage that they first retrieve the whole wave function in the image plane, that is, amplitude and phase. In this way, they use all possible information. In the other two methods, one starts from the image intensity only and inevitably misses the information that is predominantly present in the phase. Ideally, one should combine highvoltage microscopy or aberration correction with holography so as to combine the advantage of holography with a broader field of view, that is, a larger reconstructible field. However, this has not yet been done in practice.

A full holographic reconstruction method consists of three stages. First, one has to reconstruct the wave function in the image plane (phase retrieval). Then one has to reconstruct the exit wave of the object. Finally, one has to "invert" the scattering in the object so as to retrieve the object structure. Ideally, one should be able to disentangle the object parameters to the level where the positions of all atom columns (viewed along the incident beam) can be



FIG. 2. Schematic representation of the focus variation wavefunction reconstruction procedure.



FIG. 3. An experimentally recorded focal series for $Ba_2NaNb_5O_{15}$. The defocus values are indicated.

fitted independently. This then leads to an approximate structure model.

This structure model then provides a starting point for a final refinement by fitting with the original images (i.e., in the high dimensional parameter space) that is sufficiently close to the global maximum so as to guarantee convergence.

In the case of perfect crystals, one can combine the information in the high-resolution images with that of electron diffraction patterns. Because the diffraction pattern usually yields information up to higher spatial frequencies than the images, one can, in this way, extend the resolution to beyond 0.1nm.

HOLOGRAPHIC RECONSTRUCTION

Two main holographic approaches have been developed to solve the phase problem: off axis holography, and focus variation. Here, we will only mention the results of the focus variation method. For the principles and the details we refer to [5–11].



FIG 4. Reconstructed amplitude and phase of a series of 20 images of $Ba_2NaNb_5O_{15}$. The amplitude (a) mainly shows the heavy atoms, while the phase (b) represents the light atoms. The result of the structure reconstruction step is shown in (c), with the real structure as inset.



FIG. 5. HREM image (a) and phase of the experimentally reconstructed exit wave (b) of an MgSi precipitate in an Al matrix.

First, a series of about 20 images is taken under computer control, at regular focus intervals at both sides of a reference focus. Each image contains essentially the same information about amplitude and phase of the reference image, but scrambled in a different way. By suitable image processing of the whole image series it is possible to retrieve this amplitude and phase separately. Figure 2 shows a schematical setup.

Once the wave function in the image plane is retrieved, one can easily reconstruct the exit wave of the object, provided the instrumental parameters are known with sufficient accuracy. If the instrumental parameters are only approximately known, the reconstructed exit wave still contains residual aberrations, which can be eliminated in the final fitting procedure. At this stage the resolution is only limited by the information limit of the electron microscope and can, in case a field emission source is used, reach the order of 0.1nm (at 300keV). It should be noted that the reconstruction can be done off line.

Figure 3 shows a part of through focus series of the material Ba₂NaNb₅O₁₅, and Fig. 4 (top) shows the reconstructed exit



FIG. 6. Structure model obtained with MSLS from the fitting procedure described in the text.



FIG. 7. One Gaussian function a = 300; $\sigma = 70$.

wave. In the exit wave all atom columns (in projection) can be discriminated. The resolution in the exit wave clearly exceeds the point resolution of the microscope, which, for this experiment (200keV) is only of the order of 0.25nm. However, in this case, the heavy columns are only revealed in the amplitude of the exit wave (left) and the light columns only in the phase (right).

To interpret the amplitude and phase images in terms of mass and position of the projected columns, one has to "invert" the dynamical scattering of the elections in the object. For this purpose, a simple and invertible albeit approximate channeling theory has been proposed in which each atom column acts as a channel for the electrons so as to keep a one-to-one correspondence between projected object structure and exit wave. The details are given in [12]. Figure 4 (c) shows a projection of the structure obtained from the channeling theory. The structure model obtained in this way yields accurate values for the positions of the columns and approximate values for the weights (the model is shown in the inset).

It should be noted that it is an intrinsic limitation of HREM that fast electrons parallel to a column direction are insensitive to variations along the beam direction but sensitive to perpendicular variations. For instance, it is impossible to discriminate between a column consisting of, say, atoms of mass 50 every 0.5nm and atoms of mass 100 every 1.0nm.

In a final step, the approximate retrieved structure model can be used as a good starting point for a fitting procedure with the original dataset (the whole focal series). However, if the exit wave reconstruction is carried out successfully, the focal series recalculated from this exit wave does not differ from the original images, within the noise level. This means that the exit wave contains all information that was present in the original images, and can be used for the final structure model fitting. However, because the very high and very low spatial



FIG. 8. Simulated experiment with the Gaussian function of Fig. 6 as the probability function; total number of counts = 10,000.



FIG. 9. log(*L*) in function of the position parameter *a* for the simulated experiment of Fig. 8.



FIG. 10. Sum of two gaussian functions with parameters $a_1 = 300$, $a_2 = 700$, $\sigma = 70$.

frequencies are cut off by the imaging and the reconstruction process, the reconstructed exit wave is only unique within a certain spatial frequency band (typically between 0.1 and 0.5nm).

Recently a hybrid method has been proposed, the Multislice least-squares refinement (MSLS), to determine the structure of unknown microcrystals by first obtaining an approximate model using focus variation exit wave reconstruction, which afterwards, could be refined using several selected area diffraction patterns.

An application of MSLS refinement is shown in Figs. 5 and 6. Figure 5 (left) shows an HREM image of a Mg/Si precipitate in an Al matrix. Figure 5 (right) shows the phase of the exit wave that is reconstructed experimentally using the focus variation method. From this, an approximate structure model can be deduced. From different precipitates and different zones, electron diffraction patterns could be obtained that were used simultaneously for a final fitting with MSLS.

For each diffraction pattern the crystal thickness as well as the local orientation were also treated as fittable parameters. The obtained R-factors are of the order of 5%, which is well below the R-factor using kinematical refinement, that do not account for the dynamical electron scattering. Figure 6 shows the structure obtained after refinement. Details of this study have been published by Jansen et al. [4].

PARAMETER ESTIMATION

As shown above, the ultimate structure model is refined by fitting theoretical with experimental data. We will now discuss the fitting procedure in detail. Consider an experiment with possible outcomes x_i . These can be the pixels in an image plane that are hit by the imaging electrons. In this respect,



FIG. 11. Simulated experiment with the sum of the two gaussian functions of Fig. 10 as the probability function; total number of counts = 10,000.

the whole focal series described above can be regarded as one experiment. Let us call $\{a_n\}$ the parameters of the model, including object, interaction, imaging and recording.

By means of the model one should be able to predict the probabilities $p(x_i/\{a_n\})$ that the outcome of the experiment is x_i , that is, that the electron hits the pixel x_i given the information that the model parameters are $\{a_n\}$.

The whole experiment now consists in collecting *N* events [*N* electrons reaching the image(s)]. Let us call n_i the frequency of the outcome x_i with Eq. (1)

$$\sum_{i} n_i = N \tag{1}$$

The problem then consists in estimating the model parameters $\{a_n\}$ from the outcome $\{n_i\}$ of the experiment. All the prior knowledge should be in the model, the only unknowns being the parameters. The model may contain parameters that are of

interest, such as atom coordinates and atom types and parameters that are not of interest such as microscope settings or the structure of an amorphous layer.

In some experiments one has degrees of freedom that can be chosen so as to optimize the experiment in function of the desired parameters. This is called experimental design. In [13], it has, for instance, been shown that the optimal focus sequence for the focus variation method is close to equidistant.

If the probability density function of the observations is known, it may be used to construct a precise estimator as follows. First, the available observations are substituted in their probability density function. This produces a function of the parameters only, called the likelihood function of the parameters. The maximum likelihood estimator of the parameters is defined as the parameter values that maximize the likelihood function. Maximum likelihood esti-



FIG. 12. log(L) in function of the position parameters a_1 and a_2 for the simulated experiment of Fig. 11.



FIG. 13. Sum of two gaussian functions with parameters $a_1 = 440$, $a_2 = 560$, $\sigma = 70$.



FIG. 14. Simulated experiment with the sum of the two Gaussian functions of Fig. 13 as the probability function; total number of counts = 10,000.

mators have a number of favorable properties. It is known (e.g., [14]) that if there exists an unbiased estimator that attains the Minimum Variance Bound (or Cramer-Rao Bound), this estimator is given by the Maximum Likelihood method. If the outcome of the experiment is the set $\{n_i\}$, the likelihood function (*L*) is found with the aid of multinominal statistics:

$$L = N! \prod_{i} \frac{p^{n_i}(x_i / \{a_n\})}{n_i!}$$
(2)

where $p(x_i/\{a_n\})$ is the probability that the measurement yields the value x_i , given that the model parameters are $\{a_n\}$. This probability is given by the model. For instance, in the case of HREM, $p(x_i/\{a_n\})$ represents the probability that the electron hits the pixel x_i , in the image if all parameters of the model (object structure and imaging parameters) are given, n_i then represents the measured intensity (in number of electrons) of the pixel x_i . In practice, it is more convenient to use the logarithm of the likelihood function, defined as log(L). Because log is a monotonic function, log(L) yields the same maxima. Using Eq. (2), we find for log(L) [Eq. (3)],

$$\log(L) = \sum_{i} n_{i} \log p(x_{i}/\{a_{n}\}) + \text{constant}$$
(3)

The base of the logarithm is not important because it only changes the constant. Each possible set of parameters can be represented by a point in parameter space. The dimension of this space is equal to the number of parameters in the model. The function log(L) can be calculated for each point in this space. The best estimate for the model parameters is then given by the point for which log(L) is maximal. Log(L)can then be considered as a criterion of goodness of fit. In principle, the search for the best parameter set is then reduced to the search for optimal fitness in parameter space. Different optimization methods exist (e.g., hill climbing, genetic algorithms,



FIG. 15. Log(L) in function of the position parameters a_1 and a_2 for the simulated experiment of Fig. 14.

Tabu search, simulated annealing, etc.), but they all fail if the dimension of the search space is too high. search space can be reduced drastically by using reconstruction schemes that undo the imaging process so as to uncouple the model parameters. In the case of HREM,

As shown in §3, the dimension of the



FIG. 16. Log(*L*) in function of the distance parameter $d = a_1 - a_2$. In this case, the two peaks can be resolved.

the reconstruction would be ideal if all individual atom columns could be isolated.

RESOLUTION

To study the aspects of resolution in the framework of parameter estimation, we will use a very simple example. Consider an experiment that consists in localizing a one-dimensional object. The object has a shape f(x) and is located at position a. The probability that the outcome of a measurement is x_i is then given by:

$$p(x_i/a) = f(x_i - a) \tag{4}$$

Figure 7 shows an example with a Gaussian object with position 300 and standard deviation 70 (in pixel units). Figure 8 now shows a simulated experiment using Eq. (4) with N = 10,000 samples. From this experiment one now has to estimate the position of the Gaussian object. The function log(L) for this experiment is shown in Fig. 9. The best estimate for *a* corresponds with the maximum of *L*. The value may differ from the theoretical value. If the experiment would be repeated, different values will be found for *a*. It can be proven that the average over all possible experiments yields the theoretical value for *a* and that the standard deviation on *a* is given by Eq. (5)

$$\sigma = \frac{\sigma_0}{\sqrt{N}} \tag{5}$$

where σ_0 is the width of the Gaussian object, and *N* is the number of samples in the experiment. The resolution of an imaging system can now be described as follows: suppose the object would be an ideal point object, the "image" of which is spread by the imaging system into a Gaussian peak. Then σ_0 would be a measure of the resolution of the system in the original sense of Rayleigh. However, as shown above, this concept of resolution does not hold in the framework of parameter estimation. Because the form of the object is known, the figure of merit is now the standard deviation on the estimated position of the object



FIG. 17. Log(*L*) in function of the distance parameter $d = a_1 - a_2$. In this case, the two peaks cannot be resolved.



and it is equal to the Rayleigh resolution divided by the square root of the number of samples (counts).

Resolution in the sense of resolving power can be studied on the hand of the following simple example. The experiment now consists of locating two identical one-dimensional objects with the same shape function f(x) and located at positions a_1 resp. a_2 .

The probability for an experimental outcome of x_i is now Eq. (6)

$$p(x_i/a_1a_2) = \frac{1}{2}[f(x_i-a_1) + f(x_i-a_2)] \quad (6)$$

Figure 10 shows an example of two Gaussian objects with standard deviation 70 at respective positions 300 and 700 (in pixel units). Figure 11 shows a simulated experiment with N = 10,000 samples (counts). The parameter space is now two dimensional. The function log(L) of this experiment is shown in Fig. 12 (a contour map is also shown in projection). L has two maxima, which are symmetrical along the symmetry plane $a_1 = a_2$ as a consequence of the symmetry of the problem. If the two peaks are very close, as is shown in Fig. 13 with the corresponding experiment in Fig. 14, the two maxima of log(L) may merge, as shown in Fig. 15. The standard deviation on the individual peaks along the line interconnecting the maxima will now increase as [15] [Eq. (7)]:

$$\sigma = \frac{2\sqrt{2\sigma_0^2}}{\sqrt{N}|d|} \tag{7}$$

with $d = a_1 - a_2$ the distance between the peaks. If $\sigma < |d|$, the peaks can still be resolved, which is only the case if the number of counts exceeds [Eq. (8)]

$$N > 8 \left(\frac{\sigma_0}{d}\right)^4 \tag{8}$$

If this is not the case, the two maxima will approach each other within the distance of the standard deviation and may merge into one maximum so as to make the peaks unresolvable. This is a degeneracy of parameter space.

Resolving two objects (peaks) has now become a yes or no problem. Figure 16

shows $\log(L)$ as a function of d, that is, along the line connecting the two maxima in Fig. 15. Only d > 0 is shown. The function is symmetrical for d < 0. log(*L*) shows a maximum at a non zero value for d, indicating that the peaks are resolved. Figure 17 shows the same log(L) in case of degeneracy. Here the maximum occurs at d = 0, that is, the objects are not resolved. In [16] this degeneracy has been described using the catastrophe theory. The problem has been studied in more detail in [17-19]. A critical parameter that judges the ability to resolve the objects is the curvature of log(L)in the point d = 0. If this curvature is negative, the objects are not resolved, if it is positive, they are resolved. Figure 18 shows the statistical distribution of this curvature *B* for the ensemble of all possible experiments. The important point to note here is that one can define a probability of resolution, that is, the probability given by the unshaded area in Fig. 18. The probability that the objects will not be resolved is given by the shaded area in Fig. 18. These probabilities are functions of d, σ_0 , and N and can be calculated explicitly.

CONCLUSION

Indirect and direct methods for interpreting HREM images can be seen as different ways of matching model parameters from an experiment. In the case of unknown structures, the problem becomes unmanageable unless the model parameters can be uncoupled by holographic reconstruction methods. The concept of resolution is reconsidered in the framework of parameter estimation. It is shown that the nonresolution of objects is due to a degeneracy in parameter space. This leads to the definition of probability of resolution, which can be calculated explicitly.

References

1. A. Thust, M. Lentzen, and K. Urban: Nonlinear reconstruction of the exit plane wave function for periodic high resolution electron microscopic images. *Ultramicroscopy* 53:101–120 (1994).

- R. Bierwolf and M. Hohenstein: Remise free reconstruction of the exit-surface wave function in HREM. *Ultramicroscopy* 56:32 (1994).
- 3. M. Lentzen and K. Urban: Reconstruction of the projected crystal potential from a periodic high-resolution electron microscopy exit plane wave function. *Ultramicroscopy* 62:89–102 (1996).
- J. Jansen, D. Tang, H. W. Zandbergen, and H. Schenk: MSLS, a least squares procedure for accurate crystal structure refinement from dynamical electron diffraction patterns. *Acta Crystallogr.* A54: 91–101 (1998).
- P. Schiske: Focal series reconstruction, In: *Image* Processing and Computer Aided Design in Electron Optics, Hawkes, P. W., ed., Academic Press, New York, pp. 82–90 (1973).
- E. J. Kirkland, B. M. Siegel, N. Uyeda, and F. Fujiyoshi: Digital reconstruction of bright field phase contrast images from high resolution electron micrographs. *Ultramicroscopy* 5:479–503 (1980).
- E. J. Kirkland: Improved high resolution image processing of bright field electron micrographs. *Ultramicroscopy* 15:151–172 (1984).
- W. O. Saxton: *Focal Series Restoration in HREM.* Proceedings of the XIth International Congress on Electron Microscopy, Kyoto, pp. 1–4 (1986).
- 9. D. Van Dyck and M. Op de Beeck: Direct methods in high resolution electron microscopy. *Scanning Microsc. Suppl.* 6:115 (1992).
- W. Coene, G. Janssen, M. Op de Beeck, and D. Van Dyck: Phase retrieval through focus variation for ultraresolution in field-emission transmission

electron microscopy. *Phys. Rev. Lett.* 69:3743–3746 (1992).

- M. Op de Beeck and D. Van Dyck: Direct structure reconstruction in HRTEM. *Ultramicroscopy* 64:153– 165 (1996).
- D. Van Dyck and M. Op de Beeck: A simple intuitive theory for electron diffraction. *Ultramicroscopy* 64:99–107 (1996).
- A. Buist, A. van den Bos, and M.A.O. Miedema: Optimal experimental design for exit wave reconstruction from focal series in TEM. *Ultramicroscopy* 64:137–152 (1996).
- A. van den Bos: Parameter estimation. In: *Handbook* of *Measurement Science*, Vol. 1, P. H. Sydenham, ed., John Wiley, New York, pp. 331–377 (1982).
- E. Bettens, A. J. den Dekker, D. Van Dyck, and J. Sijbers: Ultimate Resolution in the Framework of Parameter Estimation. Proceedings of the International Conference Signal and Image Processing (SIP'98), Las Vegas, NV (1998).
- A. van den Bos: Optical resolution: An analysis based on catastrophe theory. J. Optical Soc. Am. 4: 1402–1406 (1987).
- A. van den Bos: Ultimate resolution: A mathematical framework. *Ultramicroscopy* 47:298–306 (1992).
- A. van den Bos and A. J. den Dekker: Ultimate resolution in the presence of coherence. *Ultramicroscopy* 60, 345–348 (1995).
- A. J. den Dekker: Model Based Optical Two-Point Resolution. Proceedings of the ICASSP 96, Atlanta, GA. IEEE (1996), pp. 2395–2398 (ISBN 0-7803-3192-3) (1996).

Received February 1998; accepted October 1998.