FROM HIGH RESOLUTION IMAGE TO ATOMIC STRUCTURE: HOW FAR ARE WE?

D. Van Dyck^{1*}, E. Bettens¹, J. Sijbers¹, M. Op de Beeck², A. van den Bos³ and A.J. den Dekker¹

¹ University of Antwerp (RUCA), Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

² Philips Research Laboratories, Prof. Holstlaan 4, NL-5156 AA Eindhoven, The Netherlands

³ Delft University of Technology, Lorentzweg 1, NL-2628 CJ Delft, The Netherlands

Abstract

The problem of quantitative interpretation of highresolution electron micrographs 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 the so-called direct methods such as holographic exit wave reconstruction and direct structure retrieval 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 to occur is a function of the distance between the atoms and the dose of the imaging particles.

Key Words: High resolution electron microscopy, resolution, parameter estimation, structure retrieval.

*Address for correspondence:

D. Van Dyck

Department of Physics, University of Antwerp (RUCA)

Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

Telephone Number +32-3-2180258

FAX Number: +32-3-2180318

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 Figure 1. First, the electron interacts with the atoms in the object, through multiple scattering. Secondly, 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. Since the imaging process is coherent, both the exit wave and 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 photographic film or by a chargecoupled device (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.
- (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 high-resolution 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

E-mail: dvd@ruca.ua.ac.be

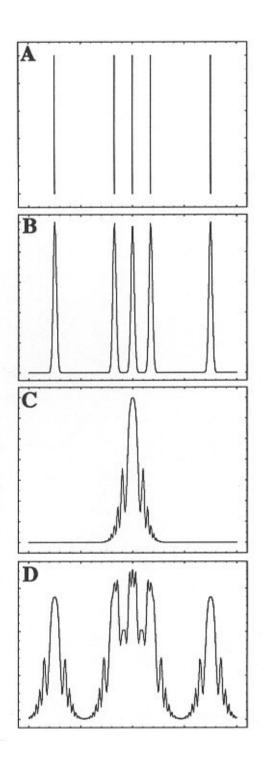


Figure 1. Due to lens imperfections, the recorded images are a blurred representation of the crystal structure. Here, an analogon is shown, using real impulse response functions: (a) Crystal structure, (b) Crystal potential, (c) Schematic representation of the impulse response function, (d) Blurred image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

of resolution in the same context.

Quantitative Image Interpretation

In principle, one is not interested in high-resolution 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 (Figures 1 and 2), 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 criterion of goodness of fit such as likelihood, mean square difference or R-factor (cf. X-ray crystallography). For each set of parameters of the model, one can calculate the value of this criterion of goodness of fit, so as to yield a goodness of fit function in parameter space. The parameters for which the goodness of fit is maximal then yields the best estimates that 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 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. Since 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 characterised 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 delocalised by the image transfer in the electron microscope (Figure 1) so that the effect of the structural parameters is completely scrambled in the highresolution 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 optimisation 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 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 (Thust et al. 1994; Bierwolf and Hohenstein, 1994; Lentzen and Urban, 1996) or if sufficient prior knowledge is available

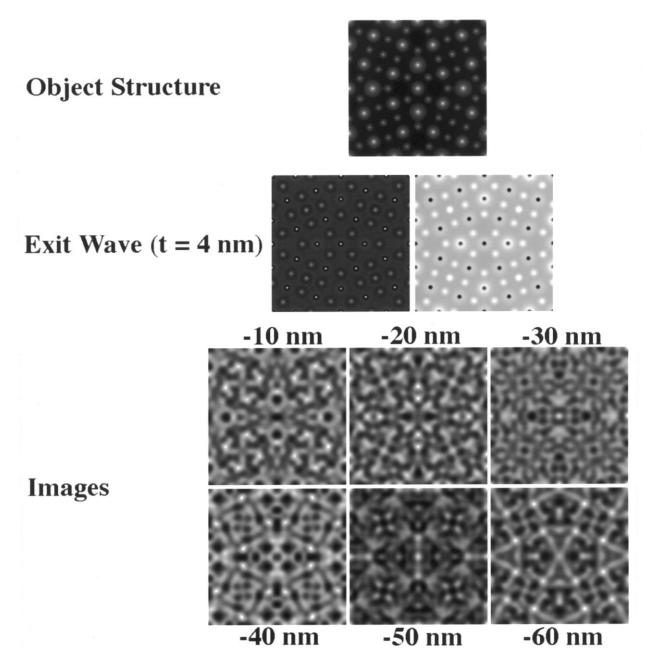


Figure 2. Due to dynamical diffraction, scattering with the object structure will yield a complex exit wavefunction, which further will be blurred by the microscope's impulse response function.

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, i.e. amplitude and phase. In this way, they use all possible information. In the other two methods, one starts from the

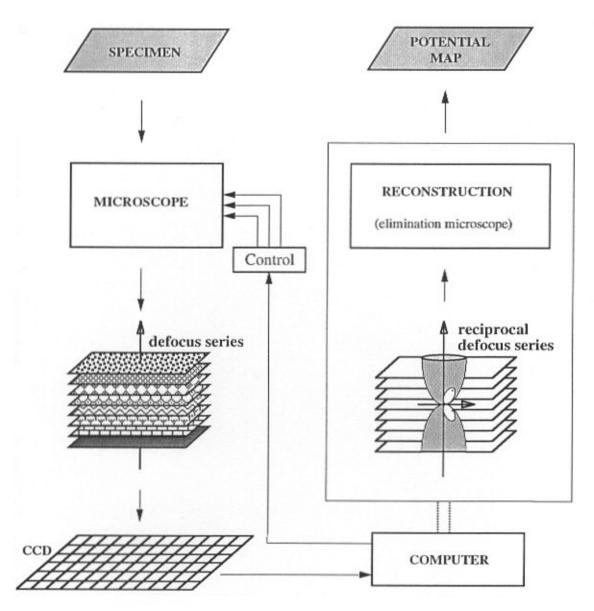


Figure 3. Schematic representation of the focus variation wavefunction reconstruction procedure.

image intensity only and inevitably misses the information that is predominantly present in the phase. Ideally, one should combine high voltage microscopy or aberration correction with holography so as to combine the advantage of holography with a broader field of view, i.e. a larger reconstructable 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 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.

It has to be noticed that in case of perfect crystals, one can combine the information in the high-resolution images with that of an electron diffraction pattern. Since the diffraction pattern usually yields information up to higher spatial frequencies than the images, one can in this way extend the resolution to beyond $0.1\,\mathrm{nm}$.

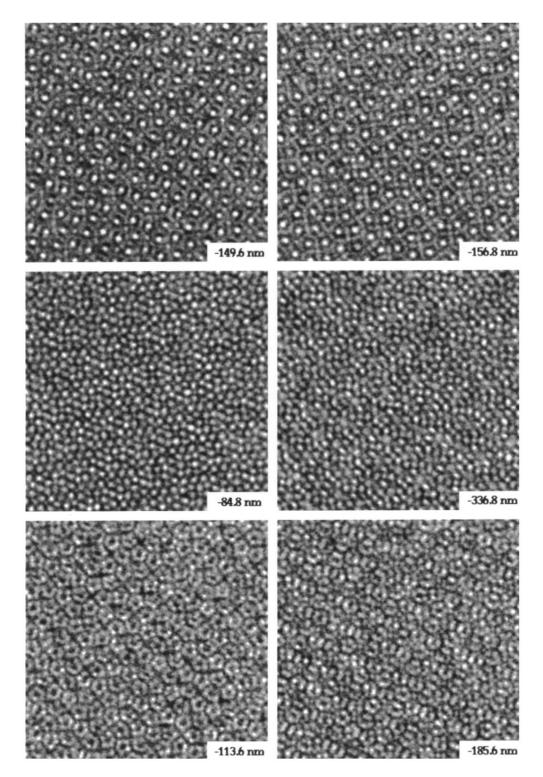


Figure 4. An experimentally recorded focal series for $Ba_2NaNb_5O_{15}$. The defocus values are indicated.

Recently it has been shown how very accurate structure determination can be achieved by starting from an

exit wave reconstruction, and further refinement using dynamical electron diffraction data, yielding R factors well

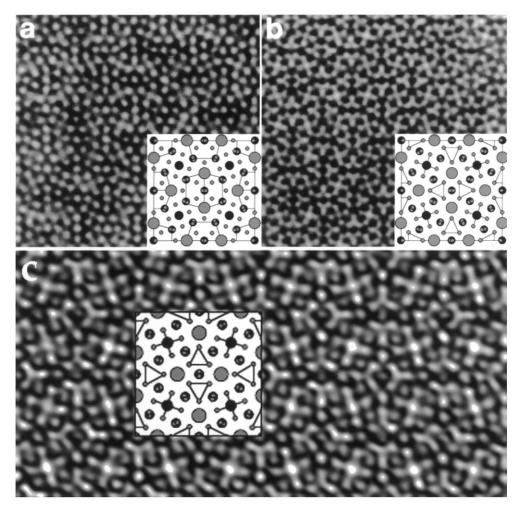


Figure 5. Reconstructed amplitude and phase of a series of 20 images of Ba₂NaNb₅O₁₅. The amplitude (a) mainly shows the heavy atoms, while the phase (b) represents the light atoms. The result of the structure reconstruction step is shown in (c) with the real structure as inset.

below 5% (Jansen et al., 1997).

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 Schiske (1973), Kirkland *et al.* (1980), Kirkland (1984), Saxton (1986), Van Dyck and Op de Beeck (1992), Coene *et al.* (1992), and Op de Beeck and Van Dyck (1996). 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 3 shows a schematical set-up.

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 0.1 nm (at 300 keV). It should be noted that the reconstruction can be done off-line.

Figure 4 shows a part of through focus series of the

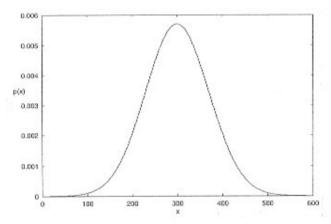


Figure 6. One Gaussian function a = 300; $\sigma = 70$.

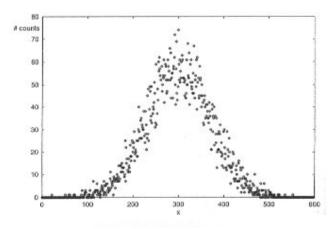


Figure 7. Simulated experiment with the N function of Figure 6 as the probability function, total number of counts N = 10000.

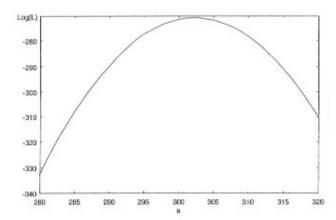


Figure 8. Log likelihood function log(L) as a function of the position parameter a for the simulated experiment of Figure 7.

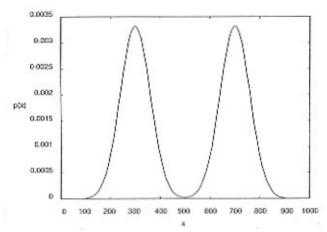


Figure 9. Sum of two Gaussian functions with parameters $a_1 = 300$, $a_2 = 700$, $\sigma = 70$.

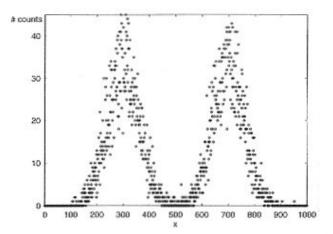


Figure 10. Simulated experiment with the sum of the two N functions of Figure 9 as the probability function, total number of counts N = 10000.

material $\mathrm{Ba_2NaNb_5O_{15}}$ and Figure 5 (top) shows the reconstructed exit wave (to compare with the simulations in Figure 2). 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 (200 keV) is only of the order of 0.25 nm. 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).

In order 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 channelling theory has been proposed in which

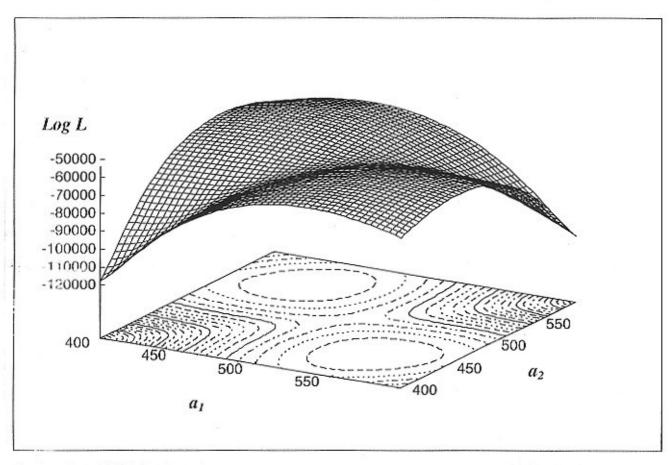


Figure 11. Log likelihood function log(L) as a function of the position parameters a_1 and a_2 for the simulated experiment of Figure 10.

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 Van Dyck and Op de Beeck (1996). Figure 5 (bottom) shows a projection of the structure obtained from the channelling 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 high-resolution electron microscopy (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 0.5 nm and atoms of mass 100 every 1.0 nm.

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).

Parameter Estimation

As shown above, the ultimate structure model is refined by fitting theoretical to 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, 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 $p(x/\{a_n\})$ that the outcome of the experiment is x_i , i.e. that the electron hits the pixel x_i given the information that the model parameters are $\{a_i\}$.

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

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

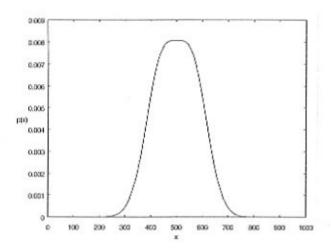


Figure 12. Sum of two Gaussian functions with parameters $a_1 = 440$, $a_2 = 560$, $\sigma = 70$.

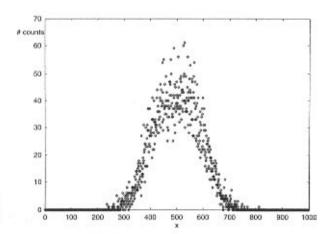


Figure 13. Simulated experiment with the sum of the two N functions of Figure 12 as the probability function, total number of counts N = 10000.

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 optimise the experiment in function of the desired parameters. This is called experimental design. In Buist and Van den Bos (1996), 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 joint 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 estimators have a number of favorable properties. It is known (e.g. Van den Bos, 1982) 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 estimator. If the outcome of the experiment is the set $\{n_i\}$, the likelihood function (L) is found with the aid of multinomial 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 case of HREM, $p(x_i/\{a_n\})$ represents the probability that the electron hits the pixel 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 i. In practice, it is more convenient to use the logarithm of the likelihood function, defined as $\log(L)$. Since \log is a monotonic function, $\log(L)$ yields the same maxima. We find for $\log(L)$, using (2),

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

The base of the logarithm is not important for maximization. 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 maximum likelihood 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 fitness function. In principle the search for the best parameter set is then reduced to the search for optimal fitness in parameter space. Different optimisation methods exist (e.g., hill climbing, genetic algorithms, Tabu search, simulated annealing....), but they all fail if the dimension of the search space is too high.

As described above, the dimension of the search space can be reduced drastically by using reconstruction schemes that undo the imaging process so as to uncouple the model parameters. In case of HREM, the reconstruction would be ideal if all individual atom columns could be isolated.

Resolution

In order to study the aspects of resolution in the framework of parameter estimation, we will use a very simple

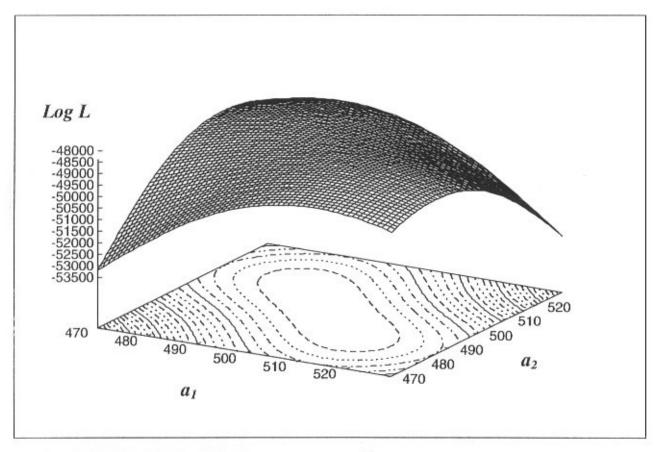


Figure 14. Log likelihood function log(L) as a function of the position parameters a_1 and a_2 for the simulated experiment of Figure 13

example. Consider an experiment that consists in localising 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:

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

Figure 6 shows an example with a Gaussian object with position 300 and standard deviation 70 (in pixel units). Figure 7 now shows a simulated experiment using Eqn. (4) with N=10000 samples. From this experiment one now has to estimate the position of the Gaussian object. The loglikelihood function $\log(L)$ for this experiment is shown in Figure 8. The maximum likelihood 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 of a is given by

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

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. Since the form of the object is known, the figure of merit is now the standard deviation on the estimated position of the object 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

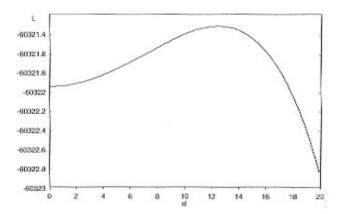


Figure 15. Log likelihood function $\log(L)$ as a function of the distance parameter $d=a_1$ - a_2 . In this case the two peaks can be resolved.

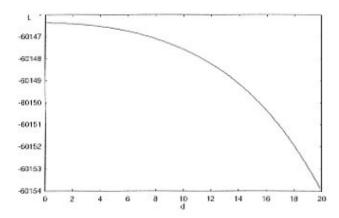


Figure 16. Log likelihood function $\log(L)$ as a function of the distance parameter $d = a_1 - a_2$. In this case the two peaks cannot be resolved.

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

Figure 9 shows an example of two Gaussian objects with standard deviation 70 at respective positions 300 and 700 (in pixel units). Figure 10 shows a simulated experiment with N=10000 samples (counts). The parameter space is now two-dimensional. The log likelihood function $\log(L)$ of this experiment is shown in Figure 11. (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 Figure 12 with the corresponding experiment in Figure 13, the two maxima of L will merge as shown in Figure 14. The peaks are now unresolvable, due to a degeneracy of parameter space.

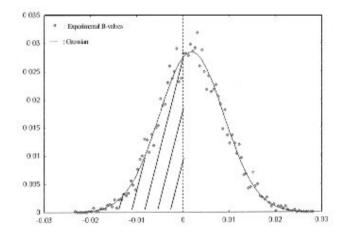


Figure 17. Distribution of the curvature factor B. If B < 0 the two peaks cannot be resolved.

Resolving two objects (peaks) has now become a yes or no problem. Figure 15 shows log(L) as a function of d, i.e., the parameter describing the distance between the peaks. 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 16 shows the same log(L) in case of degeneracy. Here the maximum occurs at d = 0, i.e., the objects are not resolved. In Van den Bos (1987) this degeneracy has been described using catastrophe theory. The problem has been studied in more detail in Van den Bos (1992), Van den Bos and Den Dekker (1995), Den Dekker (1996) and Den Dekker and Van den Bos (1997). 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 17 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, i.e., the probability given by the unshaded area in Figure 17. The probability that the objects will not be resolved is given by the shaded area in Figure 17. This probability is a function of d, σ_0 and N and can be calculated explicitly (Bettens et al., 1998, 1999):

$$P(B>0) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left[\frac{\sqrt{N}}{8} \left(\frac{d}{\sigma_0} \right)^2 \right]$$
 (7)

with erf the error function.

Conclusion

Indirect and direct methods for interpreting HREM images can be seen as different ways of matcfhing model

parameters from an experiment. In 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 non-resolution of objects is due to a degeneracy in parameter space. It leads to the definition of probability of resolution that can be calculated explicitly.

References

Bettens E, Den Dekker AJ, Van Dyck D, Sijbers J (1998) Ultimate resolution in the framework of parameter estimation. In: Proceedings of the IASTED International Conference Signal and Image processing, Las Vegas 1998. Namazi NM, ed. IASTED/ACTAPRESS, Anaheim, pp. 229-233.

Bettens E, Van Dyck D, Den Dekker AJ, Sijbers J, Van den Bos A (1999) Model-based two-object resolution from observations having counting statistics. Ultramicroscopy **77**: 37-48.

Bierwolf R, Hohenstein M (1994) Remise free reconstruction of the exit-surface wave function in HREM. Ultramicroscopy **56**: 32-45.

Buist A., Van den Bos A, Miedema MAO (1996) Optimal experimental design for exit wave reconstruction from focal series in TEM. Ultramicroscopy **64**: 137-152.

Coene W, Janssen G, Op de Beeck M, Van Dyck D (1992) Phase retrieval through focus variation for ultraresolution in field-emission transmission electron microscopy. Phys. Rev. Lett. **69**: 3743-3746.

Den Dekker AJ (1996) Model based optical two-point resolution. In: Proceedings of the ICASSP 96, Atlanta 1996. Hayes MH, ed. IEEE Service Center, Piscataway, pp. 2395-2398 (ISBN 0-7803-3192-3).

Den Dekker AJ, Van den Bos A (1997) Resolution: a survey. J Optical Soc America **A14**: 547-557.

Jansen J, Tang D, Zandbergen HW, Schenk H (1997) *MSLS*, a least squares procedure for accurate crystal structure determination from dynamical electron diffraction patterns. Acta Cryst. **A54**: 91-101.

Kirkland EJ, Siegel BM, Uyeda N, Fujiyoshi F (1980) Digital reconstruction of bright field phase contrast images from high resolution electron micrographs. Ultramicroscopy **5**: 479-503.

Kirkland EJ (1984) Improved high resolution image processing of bright field electron micrographs. Ultramicroscopy **15**: 151-172.

Lentzen M, Urban K (1996) Reconstruction of the projected crystal potential from a periodic high-resolution electron microscopy exit plane wave function. Ultramicroscopy **62**: 89-102.

Op de Beeck M, Van Dyck D (1996) Direct structure reconstruction in HRTEM. Ultramicroscopy **64**: 153-165.

Saxton WO (1986) Focal series restoration in HREM. In: Proceedings of the XIth International Congress on Electron Microscopy, Kyoto, Imura T, Maruse S, Suzuki, T, eds.. The Japanese Society of Electron Microscopy, Tokyo, Vol. PDP, pp. 1-4.

Schiske P (1973) Focal series reconstruction. In: Image Processing and Computer Aided Design in Electron Optics. Hawkes PW, ed. Academic Press, London/New York. pp. 82-90.

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

Van den Bos A (1982) Parameter estimation. In: Handbook of Measurement Science, Volume 1. Sydenham PH, ed. John Wiley & Sons, Chichester, pp. 331-377.

Van den Bos A (1987) Optical resolution: an analysis based on catastrophe theory. J Optical Soc America **A4**: 1402-1406.

Van den Bos A (1992) Ultimate resolution: a mathematical framework. Ultramicroscopy **47**: 298-306.

Van den Bos A, Den Dekker AJ (1995) Ultimate resolution in the presence of coherence. Ultramicroscopy **60**, 345-348.

Van Dyck D, Op de Beeck M (1992) Direct methods in high resolution electron microscopy. Scanning Microscopy Supplement 6: 115-120.

Van Dyck D, Op de Beeck M (1996) A simple intuitive theory for electron diffraction. Ultramicroscopy **64**: 99-107.