Electron microscope object reconstruction: Retrieval of local variations in mixed type potentials. Part I: Theoretical preliminaries

Kurt Scheerschmidt*

Max Planck Institute of Microstructure Physics, Weinberg 2, D-06120 Halle, Germany

**Abstract**

The direct object retrieval via the linearized inversion of the dynamical scattering matrix is extended using a second order perturbation theory and including mixed type potentials. The higher order perturbation increases the confidence region extracting object thickness and bending directly out of amplitude and phase of an electron wave without using trial-and-error iterative matching. Applying parameterization of a mixed type total scattering potential as *a priori* information enables a simple extension of the structure retrieval procedure to reconstruct local variations of the object potential, too. © 2009 Elsevier B.V. All rights reserved.

1. Introduction

The retrieval of the object structure in transmission electron microscopy (TEM) is in mathematical sense an inverse problem, as it is the case for the most mathematical problems in science, technology and medicine [1,2]. One has to determine model parameter or even a model itself out of experimental data where the direct problem, e.g. the physical link between a model and its parameter to the resulting data is known solely. Thus, all difficulties arising for inverse problems, especially their ill-posedness in mathematical sense, occur in TEM object retrieval, too. The ill-posedness is the main difficulty irrespective which way of solution is preferred: the direct solution of the inverse problem, suitable restriction to isolated parameter reconstruction, or trial-and-error solutions. The latter is up to now applied in nearly all applications of TEM structure analysis and is called image matching, if the intensity is fitted. A trial-and-error analysis in TEM is simply the repeated solution of the direct problem by varying all the modeling parameter up to correspondence with the experiment. Structure retrieval in TEM needs at least a solution of the phase problem, image matching based on the dynamical theory, and as much as possible additional *a priori* information. Then combining any inverse solution with tomography methods a complete structural retrieval should be possible.

The phase problem in the imaging process alone may in principle be solved using electron holography [3] or defocus wave reconstruction techniques [4], and in the following is assumed, that the exit wave of the object is already determined in amplitude and phase. Then, for instance, as described earlier in details [5–7], the retrieval of local object information can be performed directly from the electron microscope exit wave function without using trial-and-error iterative matching. The algorithm allows the direct analysis of variations within the lateral object extension of object thickness and beam orientation or equivalently local bending of the object. In principle, extensions are possible also to include changes of the scattering potential, local structural variations and special lattice defects into the reconstruction algorithm. Always the object retrieval requires the solution of the inverse scattering problem, which can be gained by linearizing the solution of the dynamical theory and constructing regularized and generalized inverse matrices.

In general, inverse problems are difficult, always fascinating, and in most cases ill- or improperly posed [1,2]. Ill- or improperly posed means that one or all of the requirements are violated usually characterizing physics, i.e. the existence, uniqueness and stability of a solution. The reason is a mathematical one, and even if this is not of general interest in physics, one can find in mathematical textbooks (already from J. Hadamard around 1902), that any operator equation of the first kind $k \times f = g$ with a linear compact operator between Banach spaces cannot have a solution $f$ which continuously depends on the right hand side $g$. The non-continuity makes the inverse instable. And much more difficult, a lack of information cannot be remedied by any mathematical...
trickery, but all important decisions must be made on the basis of insufficient information. Thus dealing with insufficient measured data requires always a priori physical related information. Although inverse problems violate especially the existence of unique and continuous solutions for arbitrary data, they are of great practical importance if the trial-and-error solution demands a large variety of possible solutions and models to be tested, as they provide a better insight into the basic relations of the physical phenomena.

The present paper is the first part of a reformulation of the object retrieval using linearized direct inversion of the scattering matrix; solely the theoretical background is given. A second part will check the consequences of reformulation and present applications. The paper is organized as follows: A short description of the problem in Chapter 2 and resuming some typical inverse solutions given in the literature. Chapter 3 gives a short repetition of the inverse solution by linearization and the extension to a 3rd order perturbation theory as well as the new approach to the very important regularization problem, an ill-posed problem which is made well-posed, loses stability and has the consequence of reformulation and presentation of the problem in Chapter 2 and resuming some typical applications. The most important techniques to solve the direct scattering problem Eqs. (1)-(3) are the diagonalization of the eigenvalue problem (1) for the whole structure or the transformation of Eq. (3) to the so-called multi-slice algorithm. The diagonalization makes difficult to include crystal defects directly and yield to additional differential equations to be solved for the varying amplitudes or excitations, but yield to a better physical insight and enables the linearization as discussed below.

The direct problem consists in assuming parameter sets \( \mathbf{I}_1 = (t, \mathbf{K}, \mathbf{V}_g, \ldots) \), which describe the scattering and structural data, solving Eqs. (1)-(3) and comparing the image intensity \( I = |\phi|^2 \) or the wave \( \phi \) with TEM or holographic experiments. The trial-and-error solution varies the parameters \( \mathbf{I} \), calculating repeatedly \( \phi \) up to experimental image matching. Uniqueness and stability is presumed, but only guaranteed, if the same analyses are done as discussed below for the inverse solution.

Fig. 1 schematically shows all necessary steps of the object analysis by TEM: anticlockwise the trial-and-error technique, clockwise the direct solution of the inverse problem. It should be mentioned only that some object information immediately can be retrieved out of the images, if a suitable reduction of the data under special imaging conditions is performed, e.g. strain and composition analysis using special reflections, shift of contrast features, Fourier filtering of the images or the geometrical phase analysis. 

2. Object retrieval

Different techniques, as mentioned in the introduction, are used to retrieve object data out of the TEM exit wave function. The basis for the analysis is the dynamical theory of electron scattering which is given in terms of the modified plane waves with complex coefficients \( C_g \), excitations \( \phi \), and reflections \( \phi^* \).

The basis for the analysis is the dynamical theory of electron scattering which is given in terms of the modified plane waves with complex coefficients \( C_g \), excitations \( \phi \), and reflections \( \phi^* \).

The direct problem consists in assuming parameter sets \( \mathbf{I}_1 = (t, \mathbf{K}, \mathbf{V}_g, \ldots) \), which describe the scattering and structural data, solving Eqs. (1)-(3) and comparing the image intensity \( I = |\phi|^2 \) or the wave \( \phi \) with TEM or holographic experiments. The trial-and-error solution varies the parameters \( \mathbf{I} \), calculating repeatedly \( \phi \) up to experimental image matching. Uniqueness and stability is presumed, but only guaranteed, if the same analyses are done as discussed below for the inverse solution.

Fig. 1 schematically shows all necessary steps of the object analysis by TEM: anticlockwise the trial-and-error technique, clockwise the direct solution of the inverse problem. It should be mentioned only that some object information immediately can be retrieved out of the images, if a suitable reduction of the data under special imaging conditions is performed, e.g. strain and composition analysis using special reflections, shift of contrast features, Fourier filtering of the images or the geometrical phase analysis. Another possibility to extract direct information out of images is the interpretation in terms of special contrast features, e.g. s-state model interpreting high resolution TEMs in terms of column channeling of atomic row projections (cf. [9] and the discussion of its limitations [10]). The “brute force” method [11], where for each reflection of holographic reconstructed micro-diffractions amplitudes and phases for a sufficient parameter space are simulated in advance as basis to find the best fit to the experiment, is a bridge between trial-and-error and inversion avoiding solely the repeated calculation. The combination of trial-and-error simulations with an electron based direct structure analysis method [12] is a step towards inverse solution, too.

Whereas the first step of inversion is solved, replacing the image by a hologram or a defocus series, which makes the problem linear, and find the exit object wave by inverse Fourier transform, the step from the exit wave to the object structure is the real challenge of inversion.

The present paper is the first part of a reformulation of the object retrieval using linearized direct inversion of the scattering matrix; solely the theoretical background is given. A second part will check the consequences of reformulation and present applications. The paper is organized as follows: A short description of the problem in Chapter 2 and resuming some typical inverse solutions given in the literature. Chapter 3 gives a short repetition of the inverse solution by linearization and the extension to a 3rd order perturbation theory as well as the new approach to the very important regularization problem, an ill-posed problem which is made well-posed, loses stability and has the consequence of reformulation and presentation of the problem in Chapter 2 and resuming some typical applications. The most important techniques to solve the direct scattering problem Eqs. (1)-(3) are the diagonalization of the eigenvalue problem (1) for the whole structure or the transformation of Eq. (3) to the so-called multi-slice algorithm. The diagonalization makes difficult to include crystal defects directly and yield to additional differential equations to be solved for the varying amplitudes or excitations, but yield to a better physical insight and enables the linearization as discussed below.

The direct problem consists in assuming parameter sets \( \mathbf{I}_1 = (t, \mathbf{K}, \mathbf{V}_g, \ldots) \), which describe the scattering and structural data, solving Eqs. (1)-(3) and comparing the image intensity \( I = |\phi|^2 \) or the wave \( \phi \) with TEM or holographic experiments. The trial-and-error solution varies the parameters \( \mathbf{I} \), calculating repeatedly \( \phi \) up to experimental image matching. Uniqueness and stability is presumed, but only guaranteed, if the same analyses are done as discussed below for the inverse solution.

Fig. 1 schematically shows all necessary steps of the object analysis by TEM: anticlockwise the trial-and-error technique, clockwise the direct solution of the inverse problem. It should be mentioned only that some object information immediately can be retrieved out of the images, if a suitable reduction of the data under special imaging conditions is performed, e.g. strain and composition analysis using special reflections, shift of contrast features, Fourier filtering of the images or the geometrical phase analysis. Another possibility to extract direct information out of images is the interpretation in terms of special contrast features, e.g. s-state model interpreting high resolution TEMs in terms of column channeling of atomic row projections (cf. [9] and the discussion of its limitations [10]). The “brute force” method [11], where for each reflection of holographic reconstructed micro-diffractions amplitudes and phases for a sufficient parameter space are simulated in advance as basis to find the best fit to the experiment, is a bridge between trial-and-error and inversion avoiding solely the repeated calculation. The combination of trial-and-error simulations with an electron based direct structure analysis method [12] is a step towards inverse solution, too.

Whereas the first step of inversion is solved, replacing the image by a hologram or a defocus series, which makes the problem linear, and find the exit object wave by inverse Fourier transform, the step from the exit wave to the object structure is the real challenge of inversion.

The first step of an inverse TEM problem was proposed by Head [13] using the cross section formulation of diffracton contrast for lattice defects with plane strain or stress to reconstruct the whole displacement field of the defect out of the component registered within the image plane. It reflects the basic idea: use as much as possible a priori information and restrict the problem to find only the deviations in parameter space of a known operator instead of an unknown operator itself. And it is a direct inversion: the image contrast and the displacement field gradient of lattice defects are related directly via a series expansion using the a priori information of plane stresses which couples depth oscillations and lateral contrast modulations.

The different inverse solutions may be characterized as follows: solutions based on the inversion of the multi-slice formulation [14–21]; inline holography to directly structure
factor evaluation [22]: inverting the dynamical matrix by using additional relations between its entries [23–26] or reducing the unknowns by applying systematic row excitation [27] or series expansion of the scattering matrix [28] to overcome the problem that each 2-dimensional pattern fills only one column of the scattering matrix; Pade approximation of the scattering matrix or ptychography [29–31]; applying additional information by using rocking beam in convergent diffraction [32], heavy atoms or two different thicknesses [33,34] to overcome the scattering phase problem; transform the inverse problem via an algebraic discretization to include defect reconstruction [35,36]; and by ‘‘ in (10) by the experimentally measured wave the inversion is corrected by the perturbation are given by

\[
\Phi(\Pi) = (\Pi(\Pi_0) + (\Pi - \Pi_0) \text{grad}_\Pi \Phi) + \text{more explicit}
\]

\[
\Phi(\Pi) = \Phi(\Pi_0) + (t - t_0) \delta \Phi / \delta t + \delta k_y - k_0 \text{grad}_\Phi \Phi + \cdots
\]

The perturbation matrix \( \Lambda \) reads explicitly \( \Lambda_{ij} = \mathcal{C}^{-1}(\Delta \mathbf{K}_g)(\delta \mathbf{g}_0) + \mathbf{A} \mathbf{V}_g \mathbf{C} \) and contains as diagonal elements the deviation of the orientation \( \Delta \mathbf{K} \) from that of the original non-perturbed eigenvalue system \( \mathbf{K} \). The non-diagonal elements describe a perturbation of the potential as, e.g. according to optical absorption or varying structures.

Still Eq. (4) is a highly nonlinear relation between the parameters \( \Pi \) and the waves \( \Phi \); however, it can be expanded in a Taylor series in 1st order yielding

\[
\Phi(\Pi) = \Phi(\Pi_0) + (\Pi - \Pi_0) \text{grad}_\Pi \Phi
\]

and

\[
\delta \Phi / \delta t = \Gamma \delta \mathbf{X} / \delta t \Gamma^{-1} \theta
\]

and

\[
\text{grad}_\Phi \Phi = \left( \text{grad}_\mathbf{g} \mathbf{K} \mathbf{X} - \mathbf{I} \right)^{-1} \text{grad}_\mathbf{g} \mathbf{K} \mathbf{X} + \Gamma \text{grad}_\mathbf{g} \mathbf{X} \Gamma^{-1} \theta,
\]

t etc.

Thus, the linearized Eq. (6) has the form \( \Phi(\Pi) = \Phi(\Pi_0) + (\Pi - \Pi_0) \mathbf{M} \) with analytically known theoretical wave function \( \Phi(\Pi_0) \) and matrix of derivatives \( \mathbf{M} \). Substituting \( \Phi(\Pi_0) \) by the experimentally measured wave the inversion is straightforward for the unknown parameter \( \Pi \) as discussed in the following Chapter 3.

3. Higher order perturbation and mixed type potential

Based on the perturbation and the linearization the retrieval procedure may now be summarized as follows. Starting e.g. from an electron hologram, where all reflections \( \mathbf{g} \) are separately reconstructed, the moduli and phases for each partial wave of reflex \( \mathbf{g} \) of the experimental exit plane wave \( \Phi^{\text{exp}} \) are determined as function of the lateral pixel position \( (ij) \). Moduli and phases up to the maximum resolution are necessary to get sufficient a priori
data. Theoretical waves $\Phi^{th}$ are then calculated using the dynamical scattering matrix $\mathbf{M}$ for an a priori model characterized by the number of beams and the scattering potential represented by the potential coefficients $V_i^q$. With a suitable experimentally predetermined trial average beam orientation $\mathbf{K}_0$ and a sample thickness $t_0$ as a free parameter, a perturbation approximation yields both $\Phi^{th}$ and $\mathbf{M}$ as linear functions of parameters to be retrieved.

The analytic form of the equations enables the inverse solution

$$\Pi = [t, \mathbf{K}, V_i, \ldots] = [t_0, \mathbf{K}_0, V_i^0, \ldots] + \mathbf{M}_{inv}(\Phi^{exp} - \Phi^{th})$$

(8)

thus yielding directly for each image pixel $(i,j)$ the local thickness $t(i,j)$, the local beam orientation $\mathbf{K}(i,j)$, and the variation of the potential $\mathbf{V}$ as well as further data included into the parameter space $\Pi$.

An enhancement of the reconstruction algorithm is possible: The application of mixed type potentials simplifies the reconstruction algorithm and allows overcoming some limitations using a local variable $\mathbf{V}$. The optical potential matrix $\mathbf{V}$ is replaced by a mixture of different but constant matrices $\mathbf{V}^k$ representing different structures or composites or defect regions, etc., irrespective whether the absorption is included as additional imaginary part or more complex. Important is, that additional parameter $q_k$ are introduced to describe the local variation. With the parameterized mixed type potential

$$\mathbf{V}(i,j) = \sum_k q_k(\mathbf{V}_i^k)^j \mathbf{V}^k$$

(9)

the inverse solution is then replaced by

$$[t, \mathbf{K}, q_1, q_2, \ldots] = [t_0, \mathbf{K}_0, q_01, q_02, \ldots] + \mathbf{M}_{inv}(\Phi^{exp} - \Phi^{th})$$

(10)

or equivalently in short

$$\Pi = \Pi_0 + \mathbf{M}_{inv}(\Phi^{exp} - \Phi^{th})$$

(10')

where the new $V_i^k$ as coefficients of $\mathbf{V}^k$ describe only additional a priori information, but the $q_i$ increase the space of the unknown parameter $\Pi = [t, \mathbf{K}, q, \ldots]$ to be reconstructed for each pixel $(i,j)$. The $q_k(i,j)$ should thus be able to retrieve a locally variable potential by weighting $\mathbf{V}$ via Eq. (9). The mixed type potential replaces the deviation matrix as given above by the new one $d_{gh} = C^{-1}((\Delta \mathbf{K} \mathbf{g})(\mathbf{v}_g + \Delta q_k V_i^K) \mathbf{K}_0$ and thus all the derivatives via Eq. (5). The remaining equations of the reconstruction formalism keep the same form.

As a second enhancement the perturbation may be extended to higher order which increases the confidence region of the approximation by linearization. The perturbed eigenvalues up to 3rd order reads

$$\lambda = \gamma + A(d_0)D + A^{-1}(1/(\gamma_i - \gamma_i))_{00}D - A^{-1}(A_{n}/(\gamma_i - \gamma_i)^2)_{00}D$$

$$+ A^{-1}(1/(\gamma_i - \gamma_i))_{00}D(1/(\gamma_i - \gamma_i))_{00}D$$

(11)

and similar terms for the eigenvectors $\mathbf{G}$ which is straightforward and thus not written here explicitly.

4. Stability and confidence of the direct inverse

In mathematical sense the inverse problem is ill-posed and needs special techniques to get well-posed. A generalized and regularized inverse matrix, as e.g. the extended Moore–Penrose $\mathbf{M}_{inv} = (\mathbf{M}^\dagger \mathbf{C} \mathbf{M} + \varepsilon \mathbf{C})^{-1} \mathbf{M}^\dagger$, provides the least square or the likelihood solution of the direct problem extended by the constraint $\gamma|I - \Pi_0| = \min$ with a suitable matrix norm. The generalization of the inverse matrix avoids the ill-posedness of the mathematical problem, but the generalized solution is now ill-conditioned and thus instable. As pointed out in different previous analyses (cf. [38–40] and references therein), a suitable regularization of the retrieval procedure via the regularization parameter $\gamma$ and the smoothing matrices $\mathbf{C}_k, \mathbf{C}_r$ requires the control of the confidence and stability region, as well as the avoiding of modeling errors. The confidence region may be discussed considering the error of the fit for synthetic models using a likelihood measure, showing that the thickness is an uncritical parameter. The linearization smooths the solution, which is of advantage for increasing the stability of the algorithm, however, it increases the fit error, which reduces drastically the confidence region. The problem may be solved by an additional iteration process varying the a priori start configuration whenever the retrieved data go beyond the limits of the confidence region. This holds true also for the new parameter space including the $q_k$ of the mixed type potential. In addition, the difficulty in the retrieval of the potential is avoided which arise because the thickness is coupled with the mean absorption potential and a tilt offset couples with the mean scattering potential. Due to these couplings an artificial degeneracy of the solution occurred, the mixed type potential, however, removes this coupling degeneracy via the restricted freedom in applying the a priori information.

5. Conclusions

Structure retrieval at an atomic level directly from a reconstructed electron wave at the exit surface of an object, results in particular inverse problems of the first kind, viz. the analysis of object parameters from measured data. Different possible solutions are discussed; no one is sufficient stable up to now. The direct solution of the inverse problem by linearization using perturbation theory techniques is described in detail. To reduce the difficulties, especially the instabilities of the inverse solution due to the restricted confidence of the linearization, the perturbation is extended to 3rd order and a simplified model of the object structure is applied via a mixed type of potential.

The difficulties result from the mathematical point of view that the retrieval procedure is an ill-posed inverse problem requiring additional information, e.g. the periodicity of the object, the mean thickness, the global orientation, the average potential or the unperturbed atomic positions in order to make the process stable and continuous, i.e. to avoid singularities, and to restrict the manifold set of solutions. The procedure described has transformed the ill-posed problem to a well-posed but instable one. Whether uniqueness and stability of the extended solutions are enhanced and can now handled by the regularization methods applied, should be discussed in more detail in a forthcoming paper.

References