Journal of Microscopy, Vol. 190, Pts 1/2, April/May 1998, pp. 238–248.
Received 10 March 1997; accepted 18 August 1997

Retrieval of object information by inverse problems in electron diffraction

K. SCHEERSCHMIDT
Max Planck Institute of Microstructure Physics, Weinberg2, D-06120 Halle, Germany

Key words. Electron crystallography, electron holography, high-resolution electron microscopy, improperly posed problems, inverse solutions, object retrieval.

Summary

The imaging of crystal defects by high-resolution transmission electron microscopy or with the help of the electron diffraction contrast technique is well known and routinely used. However, a direct and phenomenological analysis of electron micrographs is mostly not possible, but requires the application of image simulation and matching techniques. The trial-and-error matching technique is the indirect solution to the direct scattering problem applied to analyse the nature of the object under investigation. Alternatively, inverse problems as direct solutions of electron scattering equations can be deduced using either an invertible linearized eigenvalue system or a discretized form of the diffraction equations. This analysis is based on the knowledge of the complex electron wave at the exit plane of an object reconstructed for the surrounding of single reflections by electron holography or other wave reconstruction techniques. In principle, it enables directly the retrieval of the local thickness and orientation of a sample as well as the refinement of potential coefficients or the determination of the atomic displacements, caused by a crystal lattice defect, relative to the atom positions of the perfect lattice. Considering especially the sample orientation as perturbation the solution is given by a generalized and regularized Moore–Penrose inverse, where the resulting numerical algorithms imply ill-posed inverse problems.

1. Introduction

Electron holography or other reconstruction techniques (Lichte, 1986, 1992; Coene *et al.*, 1992; Van Dyck *et al.*, 1993) permit the determination of the scattered wavefunction at the exit surface of a crystal directly from the hologram or from defocus series up to the microscope information limit owing to the decreasing signal to noise with spatial frequencies in the phases. In particular, the sidebands of a Fourier-transformed hologram represent the Fourier spectrum of the complete complex image wave and its conjugate, respectively, from which the object wave can be reconstructed by separating, centring and applying the inverse Fourier transform including the conjugate complex phase contrast transfer to correct the aberrations. Thus, both the reconstructed amplitudes and phases can be compared to trial-and-error calculations (Lichte, 1991; Lichte *et al.*, 1992), where the wavefunctions are modelled by calculating the interaction process of the electron beam with the almost periodic potential of the matter and fitted to the experiment by varying the object model and the free parameters.

As often occurs in many physical investigations, in the mathematical sense, the direct solution of the diffraction equations implies an inverse problem. Inverse problems are difficult, always fascinating and in most of the cases ill or improperly posed (Tichonov & Arsenin, 1977; Lavrentiev, 1967). Ill or improperly posed means that one or all requirements that usually characterize physics, i.e. existence, uniqueness and stability of a solution, are violated. Although the inverse problems violate especially the existence of unique and continuous solutions to 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, mostly providing a better insight into the basic relations of the physical phenomena.

In previous papers (Scheerschmidt & Hillebrand, 1991; Scheerschmidt & Knoll, 1994, 1995; Scheerschmidt, 1997a,b), it was demonstrated that the local thickness and orientation can be calculated directly from the wavefunction reconstructed at the exit surface of the object instead of using trial-and-error simulation techniques. In principle, the analysis also holds good for the retrieval of the object potential (alternatively to Van Dyck, 1989; Van Dyck & Coene, 1988; Van Dyck & Op de Beeck, 1992; Gribelyuk, 1991), or if only the positions of the atomic scattering centres are evaluated (similar to Head, 1969). The inverse problems, however, generally dealing with insufficiently measured data, always require physically related information *a priori*. It was shown that the knowledge of both the amplitudes and the phases of a sufficiently large number of

plane waves scattered by the object as well as the partial knowledge of the potential of the perfect crystal structure imply the possibility of directly retrieving object information, instead of using trial-and-error simulation techniques. Two approximations are discussed to solve the resulting inverse scattering problem without reconstructing the whole crystal potential.

First, the special problem of retrieving the local sample orientation is solved on the basis of the perturbation approximation for perfect crystals, and by applying regularized and generalized matrices to invert the resulting linearized problem. The corresponding iteration procedure enables the direct analysis of the moduli and phases if a sufficient number of plane wave amplitudes can be separated, yielding local thickness and bending of the object for each image pixel (Scheerschmidt & Knoll, 1995; Scheerschmidt 1997a,b). This approximation is valid solely for sufficiently perfect crystals, otherwise a so-called modelling error occurs and the corresponding atomic displacements due to defects have to be retrieved by the following method.

Second, based on the knowledge of the reconstructed complex electron wave and using a discretized form of the diffraction equations, an alternative method is developed (Scheerschmidt & Knoll, 1994, 1995), yielding an algebraic equation system for the complex amplitudes and the elastic displacements. In principle, this system enables the direct retrieval of the atomic displacements, caused by a crystal lattice defect, relative to the atom positions of the perfect lattice. Here it is assumed that the deformable ion approximation can be applied (see Section 3), which sufficiently locates the defect information in the neighbourhood of the crystal lattice reflections. The equations are invertible provided the completeness relation of the plane waves is valid; this is described by the continuity of the electron current. A special inverse problem of electron scattering is deduced considering only the atomic displacements given by the zeros of a function with an incompletely known Fourier spectrum from the scattered electron wave of which the displacement field of a crystal lattice defect can, in principle, be retrieved.

The present paper outlines the fundamental relations mainly for the first of the two special inverse problems describing some first numerical experiences related to the solution of the direct retrieval of local thickness and orientation. Numerical aspects considered are e.g. the stability of unique inverse solutions in terms of noise, and the regularization of the problem. In addition, the physical reasons for ambiguities in the inverse solutions are discussed.

2. Wave reconstruction by electron holography

Holography with electrons offers one of the possibilities of increasing the resolution by correcting microscope aberrations. It also enables the complete complex object wave to be restored. Image plane off-axis holograms are recorded in a microscope which is equipped with a Möllenstedt-type electron biprism inserted between the back focal plane and the intermediate image plane of the objective lens (Lichte, 1986, 1991, 1992; Lichte *et al.*, 1992). The object is arranged so that a reference wave outside it is transferred through the microscope, and due to a positive voltage of the biprism both waves mutually overlap in the image plane, creating additional interference fringes. The intensity of the latter is modulated by the modulus of the object wave, whereas the fringe position is varied by the phase of the object wave. Thus the recorded interference pattern is an electron hologram from which both the modulus and the phase of the object wave can be reconstructed by optical diffraction or numerical reconstruction. The reconstruction starts with a Fourier transform of the hologram. In addition to two sidebands in the central region of the Fourier spectrum, the zero peak and autocorrelation occur; this is equivalent to a conventional diffractogram. The sidebands represent the Fourier spectrum of the complete complex image wave and its conjugate, respectively, from which the object wave $o(x,y)$ can thus be reconstructed by separating, centring and applying the inverse Fourier transform including a reciprocal Scherzer filter with damping and microscope aberrations (Lichte, 1991; Orchowski *et al.*, 1995).

In the following it is important that, besides the whole sideband, each single reflection of sufficient intensity can be reconstructed separately (Scheerschmidt, 1997a,b). This provides the possibility of noise reduction if suitable windows and filtering are applied and if the pixels are precisely centred to avoid additional phase shifts. The environment of the reflections included in the filtering process has to be chosen such that the information of local distortions folded with the reflections will be transferred to the reconstructed partial waves. The reconstruction of the single reflections causes modulus and phase to be distributed in the partial waves, which is the presupposition of the inverse algorithm discussed in the following.

Figure 1 demonstrates the Fourier spectrum and the single-reflection reconstruction using an experimental hologram of a $\Sigma 13$ (001) tilt grain boundary in gold ($\theta = 22{\cdot}6°$, see Orchowski *et al.*, 1995; Orchowski & Lichte, 1996) and preliminary common work (Orchowski *et al.*, 1993). The Fourier spectrum of the hologram is given in Fig. 1(a); the pairs of the reflections used for the reconstruction are indicated with the corresponding reciprocal lattice vectors. To reconstruct the particular plane waves the single reflections as indicated in the spectrum of the hologram are separated and filtered through a Gaussian mask. Figure 1(b) shows the reconstructed real space wavefunction of the single reflections as the real part (REA) and Fig. 1(c) as the imaginary part (IMA) of the particular plane wave directions, i.e. for the reflections

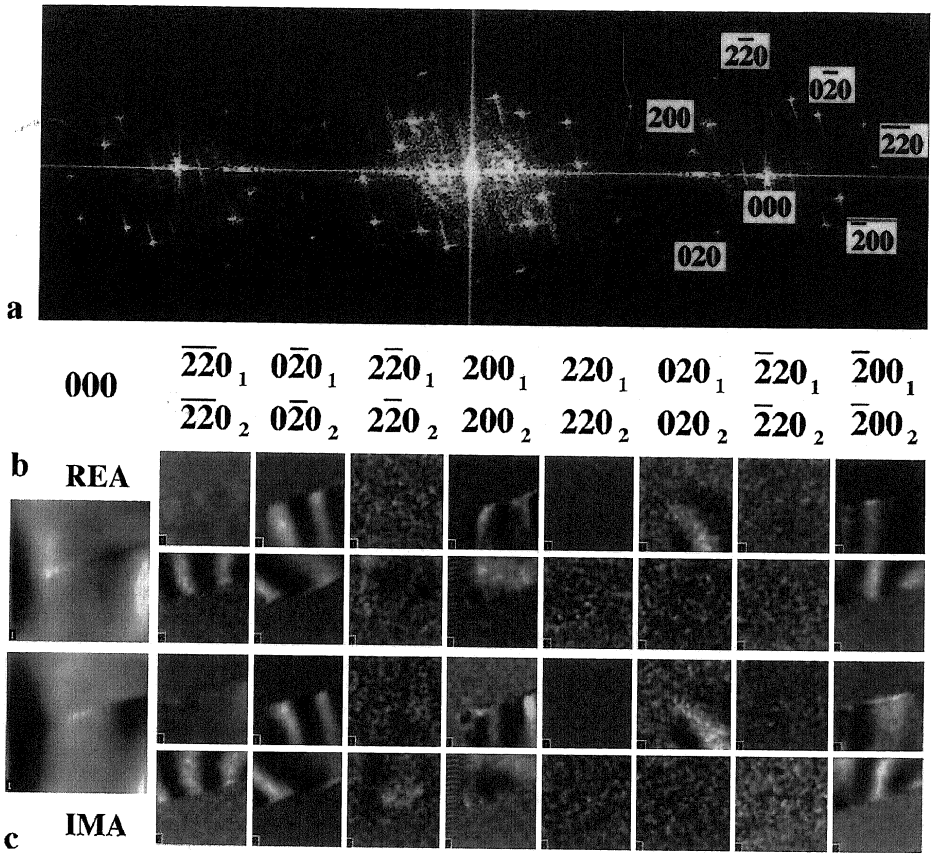

Fig. 1. Reconstruction of single reflections of a Σ13 (100) Au grain boundary, separately for both grains denoted by 1 and 2, respectively: (a) Fourier spectrum of the hologram (0·05 nm fringes, A. Orchowski, University Tübingen (Orchowski & Lichte, 1996); with indices of the reflections and asymmetric intensities in the sideband showing the mistilted orientation; (b) real (REA) and (c) imaginary (IMA) part of the partial waves, separately for grains 1 and 2, respectively, reconstructed from the separated and Gaussian filtered reflections.

chosen of types 000, {200} and {220}, respectively. The single reflections are denoted by 1 and 2 according to grains 1 and 2, respectively; in the upper rows grain 1 is excited, in the lower ones it is grain 2. The reconstructed amplitudes of the reflections can be interpreted directly as bright- and dark-field images of the grain boundary. The real and imaginary part gives the same information as amplitudes and phases without, however, the phase wrapping problem according to the multivalued phases. The reconstruction of the higher-order reflections is impossible here because of the lower intensity of the latter and the mutual overlap of the autocorrelation and the sideband. The modulation by lower frequencies is due to the local bending of the sample or to thickness oscillations, and contains the information needed for the following retrieval procedure.

3. Dynamical electron diffraction as an inverse problem

Assuming that the object wave is reconstructed free of aberrations, as discussed in Section 2, or that the scattering

is performed under diffraction contrast conditions, the influence of the microscope imaging process itself can be neglected for the surrounding of the reconstructed single reflections. Thus the image contrast is determined solely by the interaction of the electrons with the almost periodic object potential. The microscope information limit, however, restricts the possibility of reconstructing higher-order reflections, which results in the incompleteness as well as underdetermined inverse problems discussed later.

The interaction of electrons with a crystalline object is described on the basis of a periodic potential with the electron structure factors such as the expansion coefficients and the Bloch-wave method for solving the high-energy transmision electron diffraction. Different formulations can be given, using Bloch-wave or plane-wave representations of the scattered waves, applying direct or reciprocal space expansion, and direct integration or slice techniques, which, in principle, are equivalent descriptions (van Dyck, 1985, 1989; Spence & Zuo, 1992). The object wave in terms of modified plane waves with complex amplitudes ϕ_g yields

$$o(\mathbf{R}) = \Sigma_g \phi_g e^{2\pi i((\mathbf{K}+\mathbf{g})\mathbf{R}+s_g t)} \quad (1)$$

with reflections **g**, excitations s_g, wave vector **K** (corrected for refraction) and thickness t of a parallel-sided object. The amplitudes ϕ_g are constant with respect to z in the vacuum outside the object, which means that the plane waves are the stationary solutions of the wave equation. Within the crystal, however, the amplitudes of the modified plane waves ϕ_g are z-dependent according to the Ewald pendulum solution as described by the Bloch waves, which are the stationary solution within the periodic potential.

The basic equations of the Bloch-wave presentation in forward scattering approximation are given by the eigenvalue system

$$\Sigma A_{gh} C_h - \gamma C_g = 0, \text{ with } 2K_z A_{gh} = K_z s_g \delta_{gh} - V_{g-h} \quad (2)$$

yielding the amplitudes $C_g^{(j)}$ of the jth partial wave and its 'anpassung' $\gamma^{(j)}$ to the dispersion of the lattice as a function of the lattice potential (Fourier coefficient V_g) as well as the relative orientation of the object with respect to the electron beam incidence **K**. With these eigenvalues and vectors, for a plane parallel perfect crystal of thickness t the complex amplitudes ϕ_g of Eq. (1) are given directly in matrix form by

$$\Phi = \mathbf{C}\mathbf{X}\mathbf{C}^{-1}\theta, \quad (3)$$

where $\Phi = [\phi_g]$ and θ are the vectors of the amplitudes of the exit and the incident waves, respectively, and **X** represents the diagonalized scattering matrix $\exp(2\pi i \mathbf{A} t)$.

Furthermore, using the deformable ion approximation a crystal lattice defect can be included by its elastic displacement field **u** as a phase shift of the Fourier spectrum of the crystal potential. The evaluation of the quantum-theoretical scattering problem using the high-energy forward scattering approximation yields a parabolic differential equation system for vector Φ of the complex amplitudes of the elastically scattered electron waves:

$$\partial\Phi/\partial z = (\nabla^2 + V[e^{i\mathbf{gu}}])\Phi, \quad (4)$$

with $\nabla^2 = \{iK_z\nabla_{xy}{}^2 - 2(\mathbf{K}+\mathbf{g})\nabla_{xy}\}\Phi/2K'_z + 2\pi(s_h - s_g)z$, $\nabla_{xy} = (\partial/\partial x, \partial/\partial y, 0)$; $K'_z = K_z + g_z + s_g$ and the potential $V = V' + iV''$ including the lattice potential V' and the absorption V'' (one electron-optical potential approximation of inelastic scattering).

In addition, boundary and initial conditions have to be applied. The linearized high-energy approximation directly fits $\phi_g(\mathbf{R},t)$ at the crystal exit surface to $\phi_g(\mathbf{R})$ outside, demanding $|\phi_g(\mathbf{R},0)| = \delta_{g0}$ at the entrance surface, whereas the continuity of the derivatives has to be omitted in the linearized case. It enables one, however, to fit the unknown displacements at the exit foil surface by using Eq. (4) without potential outside and inverting Eq. (4) directly at the exit surface. Instead of boundary conditions one can assume a periodic continuation within the slab to describe large extended crystals. The differential Eqs. (4) allow a diffusion-like interpretation and can be discretized using standard difference algorithms. Within the crystal the difference equations are equivalent for backward and forward integration with respect to the beam propagation, thus being insufficient for determining both the wave amplitudes and the elastic displacement field. One of the difference equations, however, can be replaced applying the continuity equation. At the exit surface, a further relation is given applying the forward integration with vanishing potential and the backward integration to ensure the continuity of the derivatives. An algebraic equation system results, which together with the periodic boundary conditions and the initial conditions is exactly determined. Thus, in principle, the retrieval of the displacements is given as a particular inverse problem implying finding the root of a function given by an incomplete Fourier transform (see Scheerschmidt & Knoll, 1994, 1995). Such an inverse problem is ill-posed because only one equation has to be solved for the vectorial displacements and the spectrum considered is incomplete and noisy.

The eigenvalue system for perfect crystals, Eq. (2), can be linearized by applying perturbation methods. Assuming that the eigenvalues γ are nondegenerated, and by analogy with Eq. (3), the perturbation solution may read

$$\phi = \Gamma \Xi \Gamma^{-1}\theta, \quad (5)$$

where the matrices are given by

$$\Gamma = \mathbf{C}(1+\Delta), \ \Xi = \{\exp(2\pi i \lambda t)\} \text{ and}$$
$$\lambda = \gamma + \Delta\{\delta_{ij}\} + \Delta^{-1}\{1/(\gamma_i - \gamma_j)\}\Delta \quad (6)$$

with {} indicating a diagonal matrix here and in the following.

As diagonal elements the perturbation matrix $\Delta_{gh} = (\Delta \mathbf{K}.\mathbf{g})\{\delta_{gh}\} + i\Delta V_{gh}$ contains the deviation $\Delta\mathbf{K} = \mathbf{K} - \mathbf{K}_o$ of the orientation K from that of the original unperturbed eigenvalue system $\mathbf{K}_o$ (starting value for the retrieval). The nondiagonal elements describe a perturbation of the potential as, for example, according to optical absorption.

Starting from approximate values of thickness t_o and beam orientation $\mathbf{K}_o = (K_{xo}, K_{yo})$ gained from *a priori* knowledge or by analysing, for example, the asymmetry of the single reflections reconstructed from the holographically retrieved wavefunction, the perturbation solution is valid within certain intervals around t_o and (K_{xo}, K_{yo}). Equation (5) can then be expanded in a Taylor series yielding

$$\phi(t, K_x, K_y) = \phi(t_o.K_{xo}, K_{yo}) + (t - t_o)\delta\phi/\delta t + (K_x - K_{xo}, K_y - K_{yo})\mathrm{grad}_k\phi. \quad (7)$$

The derivatives can be gained directly from Eq. (6) using equivalent abbreviations:

$$\delta\phi/\delta t = \Gamma.\delta\Xi/\delta t.\Gamma^{-1}\theta \quad (8)$$

and

$$\mathrm{grad}_k\phi = (\mathrm{grad}_k\Gamma.\Xi - \Gamma^{-1}.\mathrm{grad}_k\Gamma.\Xi + \Gamma.\mathrm{grad}_k\Xi).\Gamma^{-1}.\theta. \quad (9)$$

The linearized Eq. (7) together with the analytical expressions (8,9) enable the inverse solution:

$$(t, K_x, K_y) = \mathbf{M}_{inv}.[\phi^{exp} - \phi^{pert}], \quad (10)$$

where the matrix is given, for example, by the Penrose–Moore inverse $\mathbf{M}_{inv} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T$, which is represented analytically using the matrix of the coefficients $\mathbf{M} = (\delta\Phi/\delta t, \mathrm{grad}_k\Phi)$ of Eqs. (8,9). The series expansion (7) as well as the resulting formalism (10) can be extended to include also the derivatives of deviations from potential coefficients, which are omitted here for the sake of simplicity. This means that additional unknown object parameters can be included in the retrieval procedure as far as the problem remains overdetermined with respect to the unknowns.

4. Regularized solutions of the inverse scattering problem

Algorithm (10) is the solution to the inverse problem concerning the local thickness and orientation analysis. The regularized inverse iteration can be applied directly to each pixel in the real space representation of the single reflections reconstructed from the hologram. On an *a-priori* assumption for the basic eigenvalue system Eq. (2) which may describe the experimental situation, i.e. the number of reflections which can be observed, the retrieval starts with suitable local thickness t_o as well as incident beam orientation (K_{xo}, K_{yo}). The resulting values of thickness t and orientation (K_x, K_y) are probably refined if Eq. (10) is applied to the complex measured amplitudes (real and imaginary part of the plane waves) of each image pixel and each reflection g. Figure 2 demonstrates the applicability using the single reflection wave reconstruction of Fig. 1 and based on a nine-beam eigenvalue system to model the diffraction behaviour. Here, no further assumption was made as to the initial thickness t_o. The best fit was revealed by searching the absolute minimum of the defect of the vector norm at an extended thickness interval. Figure 2, upper row, shows two relatively flat thickness regions, corresponding to the object plateau and the border of the hole in the sample. The two regions are separated by a very noisy region, which characterizes the wedge-shaped border itself. The retrieved incident wave vector $\mathbf{K}(i,j)$ for each pixel (i,j) shows oscillations with the pixel numbers (middle and lower rows, x- and y-components, respectively), caused by the bending of the lattice planes, which results from the relaxation of the grain boundary owing to an additional twist component. Different initial orientations of $K_0 = (0{\cdot}51, 0{\cdot}71, 0{\cdot}0)$ in Fig. 2(a) and of $K_0 = (-0{\cdot}28, 1{\cdot}21, 0{\cdot}0)$ in Fig. 2(b) yield very noisy results in thickness t and orientation (K_x, K_y) for the 64×64 pixels retrieved. Nevertheless, both cases show almost the same values, $t = 0{\cdot}77\xi$ and $t = 0$ for the plateau of the object and the hole, respectively. The application of the generalized, but non-regularized, Moore–Penrose inverse results in nonstabilized inverse solutions, as shown. In addition, the eigenvalue solution cannot describe the region between the two grains, i.e. the grain boundary itself. This region will later be included, applying the discretized differential equations to retrieve the displacement field of the interface dislocations. For the retrieval procedure based on the eigenvalue system the grain boundary itself occurs as a modelling error in the mathematical sense.

The inversion proposed is based on the linearization and the fact that the problem is overdetermined with respect to the unknowns but underdetermined if the noise is included, resulting in a least-square minimization of a suitable vector norm of the defect (Louis, 1989; Bertero (1989), e.g.

$$||\Phi^{exp} - \Phi^{pert}|| = \mathrm{Min}. \quad (11)$$

As the iteration procedure seems to be amplifying the noise, regularizations should be used to enhance the stability of the procedure. The most general regularization may be of the Ivanov–Phillips–Tichonov-type (see, for example, Bertero, 1989),

$$||\Phi^{exp} - \Phi^{pert}||^2 + \epsilon||\mathbf{Z}||^2 = (\Phi^{exp} - \Phi^{pert})^+\mathbf{C}_1(\Phi^{exp} - \Phi^{pert}) + \epsilon\,\mathbf{Z}^T\mathbf{C}_2\mathbf{Z} = \mathrm{Min}. \quad (12)$$

While the Moore–Penrose inverse minimizes the defect, an additional constraint here allows one to weight the measured data by $\mathbf{C}_1$ and to smooth the solution $\mathbf{Z} = (t, K_x, K_y)$ by $\mathbf{C}_2$. Using the Moore–Penrose or similar

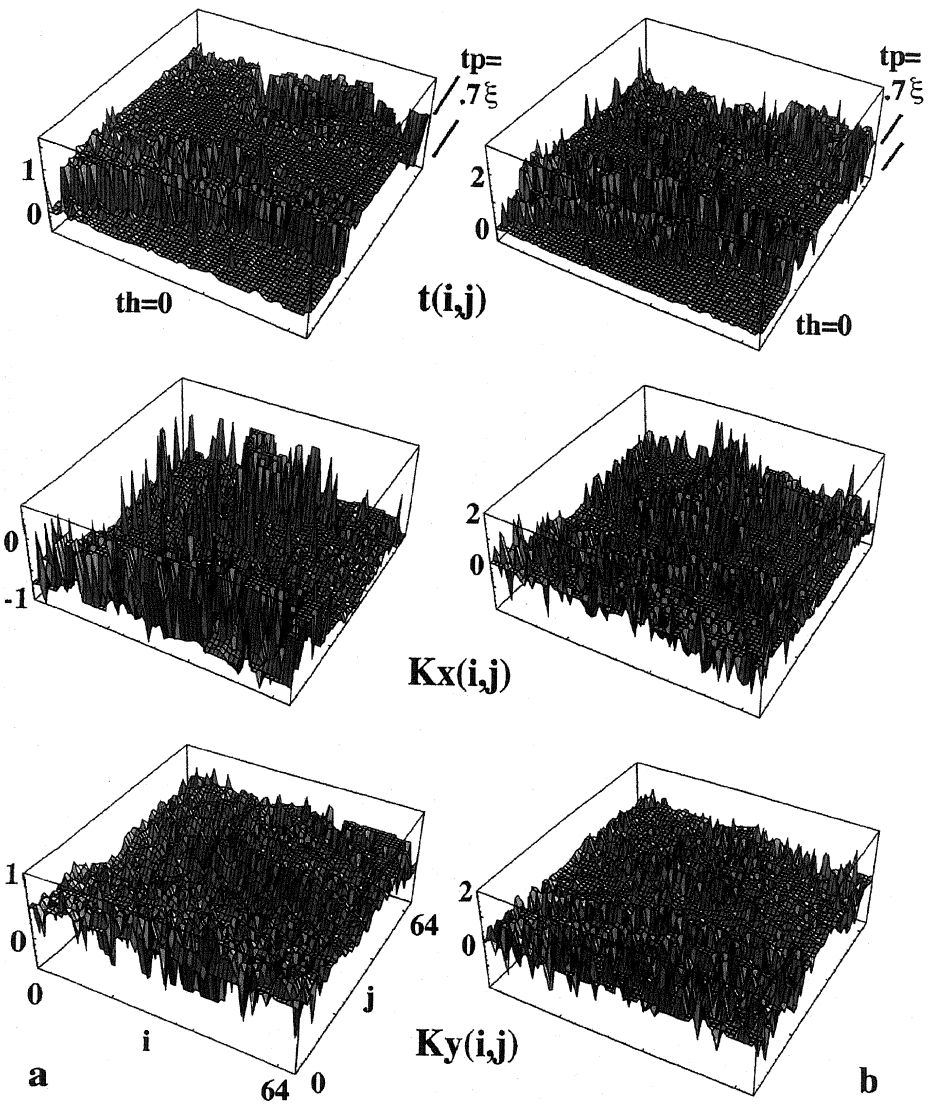

Fig. 2. Nonstabilized iteratively determined local sample thickness t and beam orientation (K_x, K_y) as functions of the pixel coordinates (i,j) retrieved from the reconstructed reflections of the experimental hologram in Fig. 1 for arbitrary starting values of thickness t (resulting in stable solutions $t = 0$ in the hole, and $t = 0{\cdot}77\xi$ on the plateau) and given start values of orientation $\mathbf{K} = (0{\cdot}51, 0{\cdot}71, 0{\cdot}0)$ and $\mathbf{K} = (-0{\cdot}28, 1{\cdot}21, 0{\cdot}0)$ for the left (a) and right (b) columns, respectively.

generalizations always allows ill-posed problems with discrete data to be transformed to well-posed, but mostly ill-conditioned ones. The solution exists and is unique; however, it is mostly unstable. The generalized solution may be considered an average of the true solutions, the resulting generalized inverse matrix including the regularization matrices may be

$$\mathbf{M}_{\text{inv}} = (\mathbf{M}^{\text{T}}\mathbf{C}_1\mathbf{M}+\epsilon\mathbf{C}_2)^{-1}\mathbf{M}^{\text{T}}, \qquad (13)$$

with the suitable regularization factor ϵ and matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively. The iterative solution of Eq. (10) with this generalized inverse (13) yields a self-consistent approach.

The generalized approach represents the maximum-likelihood solutions if the weight matrices $\mathbf{C}_1$ are suitably chosen with respect to the components $\mathbf{g}$ according to the reflections in $\mathbf{\Phi}$. Gaussian distributed noise can be described by unit weights ($\mathbf{C}_1$ equals unit matrix), Poisson distributed noise demands weights inversely proportional to the intensity of the reflections ($\mathbf{C}_1$ is diagonal with reciprocal intensities).

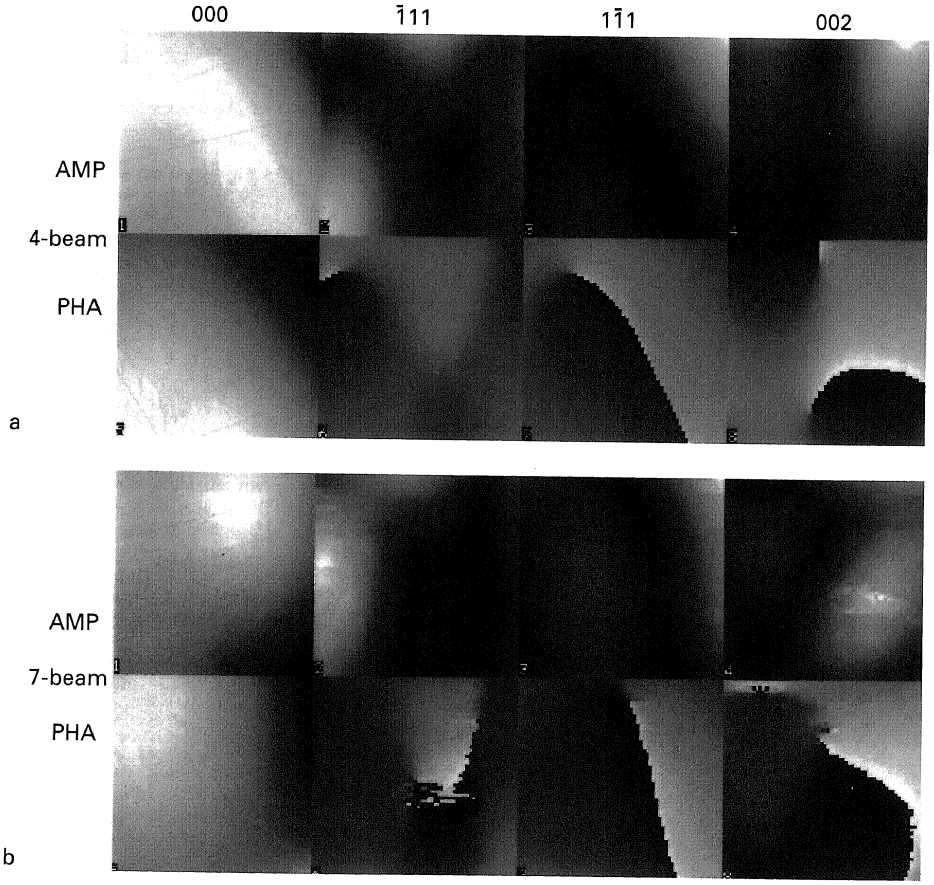

Fig. 3. Simulated plane wave amplitudes (AMP) and phases (PHA) of the reflections 000, $\bar{1}11$, $1\bar{1}1$ and 002 for [110]-orientated GaAs with thickness $t = 2{\cdot}55\xi$, varying beam orientation $\Delta\mathbf{K}_{xy} = 0{\cdot}01(i+20)[001] + 0{\cdot}01(j+20)[110]$ with the pixel coordinates (i,j), and assuming four- and seven-Bloch-wave scattering in (a) and (b), respectively.

In image processing, however, the regularization is described as a procedure smoothing the pixels (i,j) (Huang, 1975). A solution with small second derivatives with respect to neighbouring pixels tends to be more accurate ($\mathbf{C}_2$ has 2 as diagonals and -1 as first off-diagonals). In general, any constraint $\mathbf{C}_2$ which is quadratic may be used to yield a solution resembling Eqs. (10) and (13).

Assuming that the different weights can be separated without a loss of generality, the weighting $\mathbf{C}_1$ is given by $W^{+}_{gh}W_{g'h'}$ with $W|\Phi|$. The smoothing $\mathbf{C}_2$ can be described by matrix filters with respect to the pixels (i,j). A zero-order smoothing ($\mathbf{C}_2$ equals unit matrix) is equivalent to outlier detection or avoiding leaverages (Rousseeuw, 1977).

Figures 3–5 demonstrate the influence of smoothing and regularization to a synthetically generated object wave. The plane waves presented in Fig. 3 by using amplitudes and phases of the reflections 000, $\bar{1}11$, $1\bar{1}1$, 002 are calculated assuming a plane parallel object of thickness $t = 2{\cdot}55\xi$ and four- or seven-Bloch-wave scattering (Fig. 3a,b, respectively). The object is described by the potential coefficients of GaAs and with preferentially [110] orientation, i.e. for a polar orientation. The actual orientation is a linear function of the pixels, i.e. the object is characterized by a bent crystal, and the incident beam varies with respect to the crystal according to $\mathbf{K} = \mathbf{K}_0 + 0{\cdot}01(i+20)[001] + 0{\cdot}01(j+20)[110]$, with the pixel coordinates (i,j) varying from 1 to 64. Figures 4 and 5 show the retrieval of thickness $t(i,j)$ and beam orientation $\mathbf{K}(i,j)$ as functions of the image pixels. The retrieval procedure is applied with a four-Bloch-wave model to the simulated data of Fig. 3, without (Fig. 4a) and with

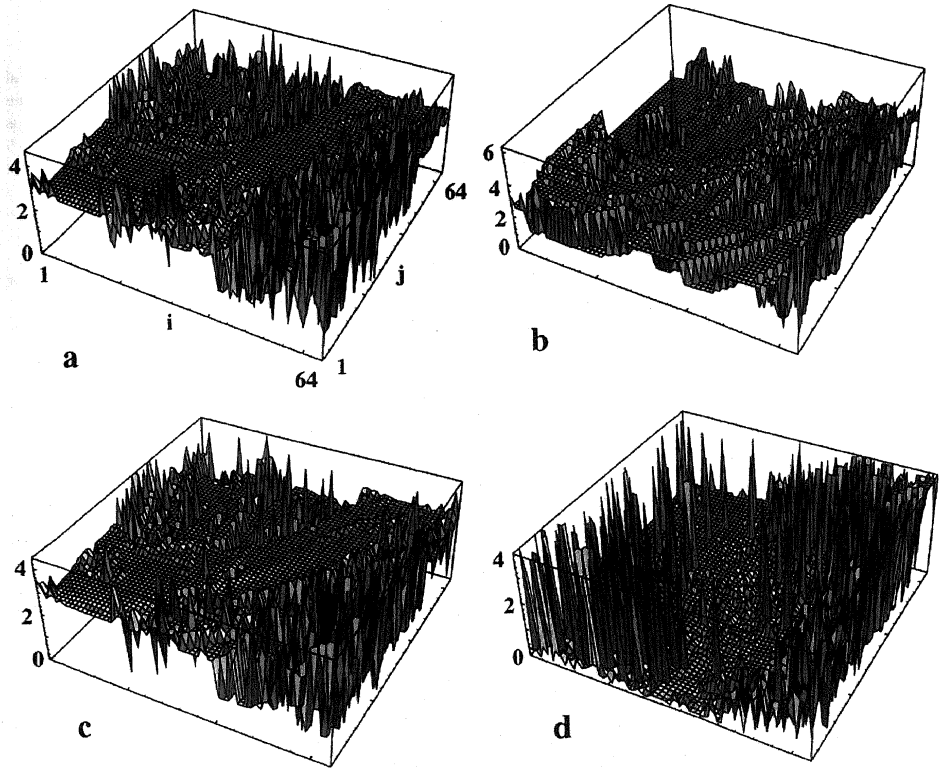

Fig. 4. Comparison of local sample thickness t retrieved from Fig. 3(a) using different smoothing matrices and regularization parameters ϵ: (a) without smoothing, $\epsilon = 10^{-5}$; (b) smoothing the first derivative, $\epsilon = 10^{-5}$; (c) smoothing the second derivative, $\epsilon = 10^{-4}$; (d) as (c) retrieved from Fig. 3(b).

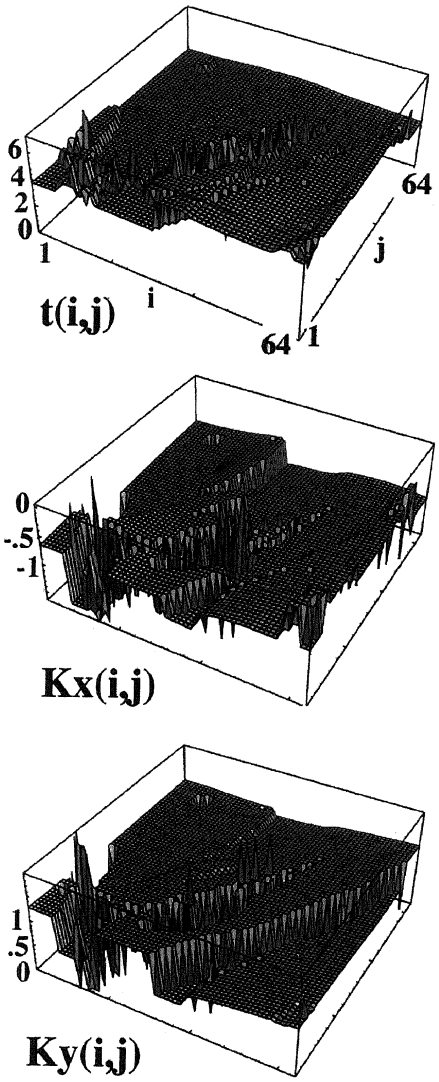

Fig. 5. Local sample thickness t and beam orientation (K_x,K_y) as functions of the pixel coordinates (i,j) retrieved from the simulated plane waves of Fig. 3(a) with smoothing the first derivatives and for a regression parameter of $\epsilon = 10^{-4}$.

different smoothings (Figs. 4b–d and Fig. 5) and different regularizations: $\epsilon = 10^{-5}$ in Fig. 4(a,b) and $\epsilon = 10^{-4}$ in Figs. 4(c,d) and 5. As assumed for the simulation of the waves the thickness is approximately constant and the orientation shows the linear functionality. Nevertheless, the noise is reduced but not removed by the regularization; always sytematic errors additionally occur as stepwise changes of t and $\mathbf{K}$, which may be considered nonunique ambiguous solutions of the inverse problem and transformation errors. The worst case of Fig. 4(d) demonstrates the modelling error because the seven-beam case is insufficiently described by the four-beam model. The smoothing of the first derivatives seems to be better than that of the second ones, which, however, is also a function of the regularization.

4. Ambiguities of the inverse solution and numerical aspects

At least three reasons for ambiguities of the retrieval procedure can be discussed: (i) a periodicity of the eigenvalues which keeps Eqs. (3) or (5) invariant if γ is replaced by $\gamma + n/t$ or λ by $\lambda + n/t$, respectively, with arbitrary integers n; (ii) a quasiperiodicity of the crystal thickness which does not change Eqs. (3) or (5) if t is replaced by $t + n/\mathrm{trace}\{\gamma\}$ or $t + n/\mathrm{trace}\{\lambda\}$ with integer n sufficient large; and (iii) a numerical ambiguity if the matrix of the coefficients $\mathbf{M} = (\delta\Phi/\delta t, \mathrm{grad}_{\mathbf{K}}\Phi)$ according to Eqs. 8 and (9) yields the commutation relation $\mathbf{M}_1\mathbf{M}_2 = \mathbf{M}_2\mathbf{M}_1$ for two different unknowns $\mathbf{X}_i = (t_i, \mathbf{K}_i)$ with $i = 1,2$.

The periodicities of (i) and (ii) in Eq. (3) – and similarly for Eq. (5) – are immediately revealed by replacing γ or t after rearranging the equations as follows:

$$\Phi = \exp(2\pi i \mathbf{A}t)\theta = \exp(2\pi i \mathbf{C}\{\gamma\}\mathbf{C}^{-1}t)\theta$$
$$= \mathbf{C}\{\exp(2\pi i \gamma t)\}\mathbf{C}^{-1}\theta, \qquad (14)$$

which reflects the diagonalization according to Eq. (2). Thus the periodicity of γ in Eq. (14) can be related to a quasiperiodicity of the orientation in $\mathbf{A}$, which can be described by a shift $\delta\mathbf{K}$ because $\mathbf{C}\{n/t\}\mathbf{C}^{-1} = \delta\mathbf{K}.\mathbf{g}/K_z$ can approximately be fulfilled for all reflections if the values of n are equal and sufficiently large (similar to the quasiperiodicity of the thickness in (ii) according to a theorem of Kronecker that a set of real numbers can always be approximated by integers for sufficiently large common factors). Condition (iii) is much more complex and will be analysed in detail in a forthcoming paper; however, the existence of the numerical ambiguities can be followed by symmetry arguments. In all three cases (i–iii) periodicities in the measured data yield ambiguities in the retrieved thicknesses and orientations. Suitable regularizations will hopefully enable us to restrict the solutions to plausible *a priori* data.

The regularization parameter can be bounded (Bertero, 1989), but in the physically relevant problems such bounds are too rough and should be estimated by numerical tests. To study the confidence level of the solutions the retrieved thicknesses and orientations are compared with those used in simulated holograms, which have been performed for either perfect crystals with increasing thickness and linearly varying orientation, or for a theoretical grain boundary (see Scheerschmidt, 1997b). To check the reliability and accuracy by using simulated inputs is advantageous as one can compare directly well-known numbers and thus determine the regularization parameter for the best fit. One can use different distance measures like the squared differences, a χ^2 test or cross correlations. Robust measures are very fast and stable: the simple sign test $\mathrm{sgn}(Z_{ij} - \langle Z \rangle) < c$ of all pixels or the product of neighbouring pixels, for instance, allows one to detect sytematic errors, whereas weights which are controlled by the regression coefficient between retrieved and exact data (Rousseeuw, 1987) enable the finding of outliers and leaverages. No test may be considered perfect or superior, because a large number of differences is always reflected by only one number. Figure 6(a) shows different $\log(\chi^2)$ measures as functions of test parameters but without smoothing; Fig. 6(b) repeats the tests 35–38 as function of the regularization parameter ϵ with weights proportional to the intensities and including pixel smoothing of the second derivative. Here χ_1 reflects the defect of the norm in Eq. (12), i.e. the deviations given by the retrieval procedure itself, χ_2 is the normalized standard χ^2 test, χ_3 the reciprocal regression value, s_1 and s_2 the corresponding sign tests (with and without outliers for upper and lower curves, respectively). Tests 1–18 are used for normalizing the different reflexes and/or applying averages over the pixels and the reflexes for varying normalization weights. In tests 19–38, additionally the regularizations $\epsilon = 10^{-5}$, 10^{-4}, 10^{-3} and 10^{-2} are applied for weights C_1 directly or reciprocally proportional to the amplitudes and intensities of the reflexes or to unity, respectively. In the tests 39–54 similarly the retrieval for constraints including the exact known *a priori* values are given, whereas in the tests 54–60 the C_1, C_2 and *a priori* constraints are varied. The calculations demonstrate that different normalizations and averages have only small influences on the accuracy and smoothing; however, a suitable *a priori* constraint minimizes the error. In addition, it shows clearly that the smoothing in (b) strongly increases the errors and that an optimum regularization exists. Further systematic calculations are necessary to find out the best regularization ϵ, i.e. the compromise between accuracy and stability of the retrieval procedure.

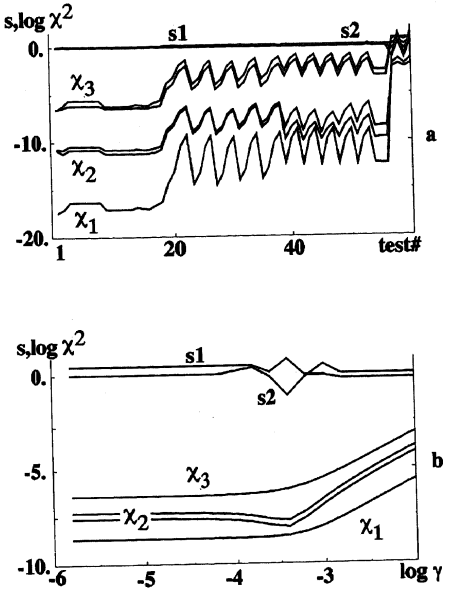

Fig. 6. Confidence tests for different log (χ^2) and robust sign s measures of retrieved test data: (a) as a function of test parameters (see text) without smoothing; (b) as a function of the regularization parameter ϵ and smoothing the second derivative.

Conclusions

The direct solution of Eq. (10), i.e. the explicit evaluation of thickness and orientation from measured data, results in a particular inverse problem of the first kind. Thus, from the mathematical point of view, the retrieval of object parameters from reconstructed plane waves scattered by the object is an ill-posed problem. It enables, however, a local linearization based on perturbation methods which transforms the difficulties of ill-posedness to the mathematical problem of overdetermined equation systems, i.e. to well-posed but ill-conditioned problems. The difficulty with the numerical stability of the direct solution is equivalent to the high calculation necessity in applying optimistic search strategies to find the best fit by image matching techniques in the many-dimensional parameter space. Normalization and regularization require additional information about the periodicity of the object as the basic assumption, and suitable starting values of thickness and orientation, in order to make the process stable and continuous, to enable smoothing and outlier detection, and to restrict the manifold set of solutions possible.

References

Bertero, M. (1989) Linear inverse and ill-posed problems. *Adv. Electron. Electron Phys.* 75, 1–114.