Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 2 September 1999
Accepted 5 January 2000

Reconstruction of the projected crystal potential in transmission electron microscopy by means of a maximum-likelihood refinement algorithm

M. Lentzen* and K. Urban

Institut für Festkörperforschung, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.
Correspondence e-mail: m.lentzen@fz-juelich.de

The projected crystal potential is reconstructed from a nonperiodic high-resolution transmission electron microscopy exit wave function using a maximum-likelihood refinement algorithm. The convergence and the accuracy of the algorithm are investigated using simulated exit wave functions of SiGe, a Shockley partial dislocation in Ge and an area containing randomly distributed Ge columns at different specimen thicknesses. The performance of two different start models for the projected crystal potential is investigated: the weak-phase-object model and a model based on the electron-channelling approximation. The reconstruction is successful even under the strongly nonlinear dynamical diffraction conditions at larger specimen thicknesses, relevant for high-resolution work, and on specimen areas large enough to cover defects in crystalline materials.

1. Introduction

In high-resolution transmission electron microscopy, the scattering of an incident electron wave by the atomic potential of an object is exploited to obtain structural information. The resulting exit wave leaving the object at the bottom surface is imaged by means of an electron optic system. This introduces a number of well known imaging artifacts like, for example, half-spacing contrast from a periodic object, contrast delocalization near defects in crystalline materials, contrast inversion rendering atom columns either bright or dark relative to the image mean intensity, and contrast attenuation of object detail at high spatial frequencies. In recent years, several different techniques, such as electron holography (Lichte, 1986), focal-series reconstruction (Gerchberg & Saxton, 1972; Schiske, 1973; Kirkland, 1984; Kirkland *et al.*, 1985; Saxton, 1978, 1988; Coene *et al.*, 1992; Van Dyck *et al.*, 1993; Thust *et al.*, 1994; Jia & Thust, 1999) and beam-tilt reconstruction (Saxton, 1988; Kirkland *et al.*, 1995), have been developed to remove the imaging artifacts and restore the exit wave up to the resolution limit of the microscope, assuming that a reconstruction provides a more intuitively interpretable view of the structure. In electron holography, an intensity distribution is recorded that originates from the interference between the exit wave function and a known undisturbed reference wave. The exit wave function can then be determined directly from the intensity distribution. In focal-series reconstruction, a series of images of the exit wave function is taken using different focus settings of the microscope. By separating the effects of electron–optic imaging and the exit wave function on the image intensities, the exit wave function can be reconstructed. The same principle is applied in beam-tilt reconstruction, where a series of images of the exit wave function is taken using different beam tilts with respect to the optic axis.

In general, however, the strong interaction of electrons with the atomic potential of the object results in dynamical multiple scattering whose complex nature leads to a loss of a direct one-to-one relationship between the projected potential and the exit wave. This applies particularly to thick objects containing heavy elements and to crystalline objects aligned along a low-index zone axis with respect to the incident electron beam. An example of a property of the exit wave that can give rise to intuitive misinterpretation is the fact that, owing to the *Pendellösung* behaviour of dynamical scattering, beyond a certain specimen thickness a strongly scattering atom column produces a *weaker* local phase shift than a weakly scattering atom column. Another example is the creation of asymmetrical wave components in the presence of specimen tilt with respect to the incident electron beam.

In principle, two ways are possible to obtain information about the projected crystal potential once the exit wave has been determined experimentally by the above-mentioned techniques. On the one hand, one can apply the numerical simulation of dynamical scattering, implemented in standard simulation packages (*e.g.* Stadelmann, 1987), to a number of different test potentials until the simulated exit wave and the reconstructed exit wave match according to a certain set of rating criteria. On the other hand, 'direct' reconstruction techniques have been developed (Henderson *et al.*, 1986; Van Dyck *et al.*, 1989; Op de Beeck & Van Dyck, 1996; Gribelyuk,

1991; Beeching & Spargo, 1993; Spargo *et al.*, 1994; Lentzen & Urban, 1996; Spence, 1998; Allen *et al.*, 1998, 1999; Rez, 1998, 1999) that avoid the potentially tedious process of rating a large number of test potentials and imply the inversion of the dynamical scattering process.

The following enumeration of already elaborated attempts to solve the reconstruction problem gives comments on their use for the practically important case of the restoration of the projected potential at larger specimen thickness and on specimen areas large enough to cover defects in crystalline materials.

In the past few decades, a number of direct reconstruction schemes have been worked out that are based on the one-to-one relation between the strength of the scattering potential and the phase of the exit wave within the framework of the weak-phase-object approximation. An application of this technique is given by Henderson *et al.* (1986). These schemes are suitable for a direct potential restoration, but the weak-phase-object approximation is commonly only valid for the scattering from thin specimens containing light elements.

In recent years, a reconstruction technique has been developed by Van Dyck *et al.* (1989) and Op de Beeck & Van Dyck (1996). It makes use of the fact that, under certain conditions, electron propagation through a crystalline specimen in a low-index orientation can be treated as a channelling phenomenon (Howie, 1966; Kambe *et al.*, 1974; Fujimoto, 1978). This allows substantial simplifications in the treatment of the quantum-mechanical scattering problem. If the intercolumn distance is not too small and the mass per thickness of an atom column is not too high, then the amplitude and phase modulation of the exit wave function is confined to the projected column position. If, furthermore, the resolution is high enough to discern individual adjacent atom columns, then a large-area analysis of the scattering object, column by column, is possible by exploiting the relatively simple relation between the local modulation of the exit wave and the scattering power of an atom column. The approximation works well for specimen thicknesses relevant for high-resolution work, but with increasing thickness or in the presence of heavy-atom columns more and more delocalized components are added to the exit wave, thus hampering a simple scattering analysis.

A third attempt was made by Gribelyuk (1991) based on the refinement of an estimate of the projected potential using the multislice algorithm (Cowley & Moodie, 1957). The first estimate is calculated from the exit wave by assuming single scattering and a subsequent free-space propagation over the estimated specimen thickness. A correction of the model of the scattering potential is then calculated from the difference between the experimental exit wave and an exit wave simulated using the multislice algorithm and the estimate of the potential found so far. The last step is repeated until convergence to a stable solution. Later, a variant of this method was published by Beeching & Spargo (1993), in which the multislice iteration is performed in *reverse* order, and in which the potential correction is determined from the difference between a plane entrance wave and the entrance wave calculated from the experimental exit wave. Since the multislice algorithm is used for the calculation of the wave function, both refinement schemes are suitable for large-area restoration. It was shown (Gribelyuk, 1991; Beeching & Spargo, 1993; Spargo *et al.*, 1994), however, that both variants of the refinement fail to find the correct solution at a larger specimen thickness where the assumptions concerning the first estimate of the potential are not valid.

Our own earlier approach (Lentzen & Urban, 1996) based on a statistical search for the projected potential using the simulated-annealing algorithm (Metropolis *et al.*, 1953) avoids the need for a starting guess for the model potential, however, at the expense of a comparably long calculation time. The rating of test projected potentials is carried out using the multislice algorithm, and therefore this technique can also in principle be used on larger specimen areas. The statistical search works well at small specimen thickness but at larger thicknesses convergence towards a 'false' ambiguous solution may occur because search space seems to become too complex in that case.

Two ingenious inversion schemes have been published recently by Spence (1998) and Allen *et al.* (1998, 1999), who show that a complete solution of the reconstruction problem for a periodic specimen is possible by exploiting the intensity distribution of the Bragg reflections in diffraction patterns of the specimen area under investigation. While the diffraction intensities provide information comparable to that of a Patterson function up to a very high spatial resolution, the missing phase information is either collected by the simultaneous measurement of fringe intensities from partly overlapping diffraction discs recorded under convergent-beam conditions (Spence, 1998) or extracted from the symmetry properties of dynamical scattering imposing a suitable number of mathematical constraints on the solutions of the inversion equations (Allen *et al.*, 1998, 1999). These two inversion schemes work well for periodic objects containing a limited number of spatial frequencies. For more complicated periodic objects, the experimental effort grows with the number of object spatial frequencies to be reconstructed since the number of diffraction peaks to be measured quantitatively increases at the same time. Both schemes require the numerical diagonalization of large matrices, whose dimensions grow with the number of spatial frequencies involved. Inversion from diffraction data has the advantage that resolution is not limited by lenses or diffraction limits.

Recently, it was shown by Rez (1998, 1999) that the projected crystal potential can be determined from two exit waves of the same specimen area measured at different acceleration voltages. A numerical test using simulated exit waves generated from a known input potential showed that the input and the reconstructed Fourier coefficients of the potential were in good agreement concerning the amplitude, while there was some error in the phase (Rez, 1998, 1999). The restoration is carried out using a direct relation between the projected potential and the local change of the exit wave, and is therefore suitable for a potential reconstruction on large specimen areas.

The above enumeration shows that none of the existing reconstruction techniques alone, with the exception of the channelling reconstruction (Van Dyck *et al.*, 1989; Op de Beeck & Van Dyck, 1996) and the voltage variation technique (Rez, 1998, 1999), is suitable for an application at larger specimen thicknesses and at the same time on large specimen areas containing defects in crystalline materials.

In this work, we propose an approach to the reconstruction of the projected crystal potential of a strongly scattering object on large specimen areas based on a maximum-likelihood refinement algorithm. The maximum-likelihood refinement is a powerful improvement of the above refinement approaches (Gribelyuk, 1991; Beeching & Spargo, 1993) to the reconstruction problem with respect to performance at larger specimen thicknesses. It is combined with a start model of the projected potential, which is derived from a given exit wave function using the channelling approximation (Howie, 1966; Kambe *et al.*, 1974; Fujimoto, 1978) of dynamical electron scattering. The proposed refinement technique is suitable for the reconstruction of the projected potential of defects in crystalline materials.

2. Theory of the maximum-likelihood refinement

Within the framework of the maximum-likelihood formalism, the refinement of the projected crystal potential consists of a reiterated three-stage calculation. It comprises the successive simulation of an exit wave function using a model of the projected potential, the rating of the simulated exit wave with respect to the experimental exit wave and the calculation of a correction of the model potential from the difference between the simulated and the experimental exit waves.

2.1. Calculation of the exit wave function

In electron microscopy, the dynamical diffraction of an electron wave by the atomic potential of an object is calculated mainly using either the Bloch-wave formalism (Bethe, 1929) or the multislice technique (Cowley & Moodie, 1957).

Considering the numerical efficiency and the computer memory requirements, in the present implementation the multislice algorithm was chosen for the calculation of an exit wave function. Additionally, the multislice algorithm provides not only the exit wave function but also the wave functions at a number of intermediate specimen thicknesses, which is useful during the fast calculation of the correction of the model projected potential described in §2.4.

2.2. Comparison of exit wave functions

The second step of the maximum-likelihood refinement is the comparison of an exit wave function simulated using a model potential with an experimentally obtained exit wave function. In the maximum-likelihood formalism, the rating between the simulated exit wave function $\psi_{\rm sim}(\mathbf{r})$ and the experimental exit wave function $\psi_{\rm exp}(\mathbf{r})$ is performed by calculating their mean squared difference:

$$S^2 = \int |\psi_{\rm sim}(\mathbf{r}) - \psi_{\rm exp}(\mathbf{r})|^2 \, {\rm d}\mathbf{r}, \tag{1}$$

where $\mathbf{r}$ denotes a vector in real space. Using the Fourier coefficients of both exit wave functions, $\Psi_{\rm sim}(\mathbf{g})$ and $\Psi_{\rm exp}(\mathbf{g})$, the rating can be calculated equivalently in reciprocal space,

$$S^2 = \int |\Psi_{\rm sim}(\mathbf{g}) - \Psi_{\rm exp}(\mathbf{g})|^2 \, {\rm d}\mathbf{g}, \tag{2}$$

where $\mathbf{g}$ denotes a vector in reciprocal space. The reciprocal-space representation (2) gives the same value as the real-space representation (1) *via* Parseval's theorem. The two representations show equal performance in numerical calculations but the reciprocal-space representation (2) is advantageous when analytical calculations concerning dynamical diffraction are performed. For this purpose, it is useful to rewrite (2) in terms of the cross-correlation coefficient of $\mathbf{\Psi}_{\rm exp}$ and $\mathbf{\Psi}_{\rm sim}$:

$$S^2 = 2 - 2{\rm Re}\langle \mathbf{\Psi}_{\rm exp} | \mathbf{\Psi}_{\rm sim} \rangle, \tag{3}$$

where $\mathbf{\Psi}_{\rm exp}$ and $\mathbf{\Psi}_{\rm sim}$ denote the vectors of the Fourier coefficients $\Psi_{\rm exp}(\mathbf{g})$ and $\Psi_{\rm sim}(\mathbf{g})$.

2.3. Calculation of the potential correction

The maximum-likelihood principle requires the determination of the minimum of S^2 with respect to the exit wave function $\psi_{\rm sim}(\mathbf{r})$. This is equivalent to determining the minimum of S^2 with respect to the Fourier coefficients of the projected potential, $U(\mathbf{g})$:

$$\partial S^2 / \partial U(\mathbf{g}) = 0, \tag{4}$$

since the exit wave function is related to the projected potential *via*

$$\mathbf{\Psi}_{\rm sim}(t) = \mathbf{P}\mathbf{\Psi}_0. \tag{5}$$

$\mathbf{\Psi}_{\rm sim}(t)$ denotes the vector of Fourier coefficients of the exit wave function at a specimen thickness t, $\mathbf{\Psi}_0$ the vector of Fourier coefficients of the entrance wave function and

$$\mathbf{P} = \exp(2\pi i \mathbf{A} t) \tag{6}$$

the dynamical scattering matrix (Hirsch *et al.*, 1967). The structure matrix $\mathbf{A}$ contains the Fourier coefficients of the projected potential, the excitation error $s(\mathbf{g})$ of the diffracted beams and the inverse electron wavelength K:

$$A(\mathbf{g}, \mathbf{h}) = U(\mathbf{g} - \mathbf{h})/2K, \quad \mathbf{g} \neq \mathbf{h}, \tag{7}$$

$$A(\mathbf{g}, \mathbf{g}) = s(\mathbf{g}). \tag{8}$$

Inserting (3) and (5) into (4), one can see that determining the minimum of S^2 requires the calculation of the derivative of the dynamical scattering matrix $\mathbf{P}$ with respect to the Fourier coefficients of the projected potential.

An analytical expression for the gradient of S^2 is calculated using the slice expansion (Sturkey, 1957, 1962) of the dynamical scattering matrix and the formalism of quantum-mechanical perturbation theory (Merzbacher, 1961). The slice expansion of $\mathbf{P}$ yields

$$\mathbf{P}(\mathbf{A}, t) = \mathbf{P}^N(\mathbf{A}, \Delta t) \tag{9}$$

via equation (6) if $\mathbf{A}$ is independent of t and the crystal is cut into N slices of thickness $\Delta t = t/N$. If the projected crystal

potential $U(\mathbf{r})$ is changed by a small perturbation $\tilde{U}(\mathbf{r})$, then the structure matrix $\mathbf{A}$ will change to $\mathbf{A}+\tilde{\mathbf{A}}$ with

$$\tilde{A}(\mathbf{g},\mathbf{h}) = \tilde{U}(\mathbf{g}-\mathbf{h})/2K, \quad \mathbf{g}\neq\mathbf{h}, \tag{10}$$

$$\tilde{A}(\mathbf{g},\mathbf{g}) = 0. \tag{11}$$

The scattering matrix will change to

$$\mathbf{P}(\mathbf{A}+\tilde{\mathbf{A}},t) = \mathbf{P}(\mathbf{A},t) + \sum_{k=1}^{N} \mathbf{P}^{N-k}(\mathbf{A},\Delta t)2\pi i\tilde{\mathbf{A}}\Delta t\mathbf{P}^{k}(\mathbf{A},\Delta t) + O\big((\tilde{\mathbf{A}}\Delta t)^2\big), \tag{12}$$

which is obtained by using the slice expansion (9) and expanding the product to first order of the perturbation $\tilde{\mathbf{A}}$. In principle, this equation is equivalent to a result by Speer *et al.* (1990), however, the form (12) provides an easy way to calculate the gradient of S^2. With the limit $N\rightarrow\infty$, the difference between the perturbed and the unperturbed scattering matrix, $\tilde{\mathbf{P}}$, takes the form

$$\tilde{\mathbf{P}}(\mathbf{A},\tilde{\mathbf{A}},t) = \mathbf{P}(\mathbf{A}+\tilde{\mathbf{A}},t) - \mathbf{P}(\mathbf{A},t) = \int_0^t \mathbf{P}(\mathbf{A},t-t')2\pi i\tilde{\mathbf{A}}\mathbf{P}(\mathbf{A},t')\,\mathrm{d}t'. \tag{13}$$

The calculation of S^2 in reciprocal space (3) involves the determination of the expectation value

$$\langle\Psi_{\mathrm{exp}}|\Psi_{\mathrm{sim}}(t)\rangle = \langle\Psi_{\mathrm{exp}}|\mathbf{P}(\mathbf{A},t)|\Psi_0\rangle. \tag{14}$$

The effect of the perturbation of the projected potential, $\tilde{U}(\mathbf{r})$, on this expectation value amounts to the difference

$$\begin{aligned}\langle\Psi_{\mathrm{exp}}|\mathbf{P}(\mathbf{A}+\tilde{\mathbf{A}},t)|\Psi_0\rangle - \langle\Psi_{\mathrm{exp}}|\mathbf{P}(\mathbf{A},t)|\Psi_0\rangle &= \langle\Psi_{\mathrm{exp}}|\tilde{\mathbf{P}}(\mathbf{A},\tilde{\mathbf{A}},t)|\Psi_0\rangle \\ &= \int_0^t \langle\mathbf{P}(\mathbf{A},t'-t)\Psi_{\mathrm{exp}}|2\pi i\tilde{\mathbf{A}}|\mathbf{P}(\mathbf{A},t')\Psi_0\rangle\,\mathrm{d}t'.\end{aligned} \tag{15}$$

From (3), (13), (14) and (15), the change of S^2 due to the perturbation is

$$\tilde{S}^2 = S^2(\mathbf{A}+\tilde{\mathbf{A}}) - S^2(\mathbf{A}) = 4\pi\,\mathrm{Im}\left\{\int_0^t \langle\mathbf{P}(\mathbf{A},t'-t)\Psi_{\mathrm{exp}}|\tilde{\mathbf{A}}|\mathbf{P}(\mathbf{A},t')\Psi_0\rangle\,\mathrm{d}t'\right\}. \tag{16}$$

At this point, it is instructive to study the physical meaning of expression (16). The scattering matrix $\mathbf{P}(\mathbf{A},t'-t)$ induces a back-propagation by a distance $t-t'$, therefore

$$\Psi_{\mathrm{e}}(t') = \mathbf{P}(\mathbf{A},t'-t)\Psi_{\mathrm{exp}} \tag{17}$$

denotes a back-propagation of the experimental exit wave function from the exit plane, $z=t$, to a plane inside the crystal, $z=t'$. On the other hand, the expression

$$\Psi_{\mathrm{s}}(t') = \mathbf{P}(\mathbf{A},t')\Psi_0 \tag{18}$$

denotes the entrance wave at $z=0$ propagated forward to the same plane inside the crystal, $z=t'$. The integrand in (16) is the matrix element of the perturbation evaluated at the crystal plane t' and the integration including all crystal planes performs a thickness averaging of the matrix element.

With the obtained results, a compact description of the gradient of S^2 with respect to the Fourier coefficients of the projected potential can be derived. Now a particular choice of the perturbation is made, where the perturbation of the real potential, $\tilde{U}(\mathbf{r})$, contains only one spatial frequency, $\mathbf{g}$, so that in the Fourier representation only coefficients $\tilde{U}(\mathbf{g})$ and $\tilde{U}(-\mathbf{g})$ are nonzero, which are related by Friedel symmetry. The evaluation of (16) using (17) and (18) yields

$$\frac{\partial S^2}{\partial U(\mathbf{g})} = \frac{2\pi i}{K}\int_0^t \sum_{\mathbf{h}}\big[\Psi_{\mathrm{s}}^*(\mathbf{h}-\mathbf{g},t')\Psi_{\mathrm{e}}(\mathbf{h},t') - \Psi_{\mathrm{e}}^*(\mathbf{h}-\mathbf{g},t')\Psi_{\mathrm{s}}(\mathbf{h},t')\big]\,\mathrm{d}t', \tag{19}$$

the gradient of S^2 in reciprocal space. The Fourier transform of (19) yields

$$\frac{\partial S^2}{\partial U(\mathbf{r})} = \frac{4\pi}{K}\,\mathrm{Im}\left\{\int_0^t \psi_{\mathrm{e}}^*(\mathbf{r},t')\psi_{\mathrm{s}}(\mathbf{r},t')\,\mathrm{d}t'\right\}, \tag{20}$$

the gradient of S^2 in real space, which is the key result to proceed with the determination of the potential correction to a given model projected potential.

In principle, many ways are possible to solve the maximum-likelihood equation (4) or its real-space equivalent

$$\partial S^2/\partial U(\mathbf{r}) = 0 \tag{21}$$

for a projected potential $U(\mathbf{r})$ generating a wave function that matches the experimental wave function.

The refinement algorithm presented in this work follows the standard route of gradient search (Press *et al.*, 1989), which exploits the fact that close to a minimum of S^2 the negative gradient points more or less towards the minimum. In the present case, the refinement of the projected potential is achieved by subtracting a small portion of the gradient of S^2 to a guess at the true projected potential. The whole procedure is repeated with the corrected potential until S^2 converges to a minimum, or to a value of zero for the case of a perfect match.

Since the gradient has a unit of $1/U$ and its magnitude is *a priori* unknown, a suitable real scale factor must be found in order to obtain a useful expression for the potential correction. A good estimate of the scale factor C can be derived from the well known observation that small changes of the projected potential generally cause small kinematical changes of the exit wave function. This particularly holds true for electron scattering from a weak phase object (WPO), where the exit wave function and the projected potential are related by

$$\psi_{\mathrm{WPO}}(\mathbf{r},t) = 1 + \pi i[U(\mathbf{r})/K]t. \tag{22}$$

Let us now suppose that $\psi_{\mathrm{WPO}}(\mathbf{r},t)$ is the experimentally obtained exit wave function $\psi_{\mathrm{exp}}(\mathbf{r})$ and that the first estimate of the projected potential is $U^{(1)}(\mathbf{r})=0$ generating an exit wave function $\psi_{\mathrm{s}}(\mathbf{r},t)=1$. The first refinement step provides an improved model for the projected potential *via*

$$U^{(2)}(\mathbf{r}) = U^{(1)}(\mathbf{r}) - C\frac{\partial S^2}{\partial U(\mathbf{r})}\bigg|_{U(\mathbf{r})=U^{(1)}(\mathbf{r})}. \tag{23}$$

For the weak phase object, the first refinement step is sufficient to restore the correct potential, hence $U^{(2)}(\mathbf{r}) = U(\mathbf{r})$. Now C can be determined from (23) and (20) using the approximation

$$\int_0^t \psi_e^*(\mathbf{r}, t')\psi_s(\mathbf{r}, t')\,dt' \approx \psi_e^*(\mathbf{r}, t)\psi_s(\mathbf{r}, t)\,t, \tag{24}$$

valid for the weak phase object. The solution is

$$C = (K/2\pi t)^2, \tag{25}$$

so that the general nth refinement step reads:

$$U^{(n+1)}(\mathbf{r}) = U^{(n)}(\mathbf{r}) + \delta U^{(n)}(\mathbf{r}) \tag{26}$$

with

$$\delta U^{(n)}(\mathbf{r}) = (K/\pi t^2)\,\mathrm{Im}\left\{\int_0^t \psi_s^*(\mathbf{r}, t')\psi_e(\mathbf{r}, t')\,dt'\right\}, \tag{27}$$

where $\psi_e^*(\mathbf{r}, t')$ and $\psi_s(\mathbf{r}, t')$ are determined for the model potential $U^{(n)}(\mathbf{r})$.

2.4. Numerical implementation

The iterative solution of the maximum-likelihood equation (21) using the refinement formulae (26) and (27) contains two approximations that tend to reduce the convergence speed, and hence the numerical efficiency, of the simple gradient search. Firstly, in many cases, the gradient of S^2 does not precisely point towards the minimum of S^2 (Press *et al.*, 1989). Therefore, a single refinement step will miss the proper minimum position. Secondly, the scale factor C, although rendering the correct unit of the potential correction, is only a rough estimate for objects deviating from the weak-phase-object model. If the resulting potential correction is too large or too small, additional refinement steps are needed to proceed towards the minimum.

Modern numerical implementations of gradient search methods (Press *et al.*, 1989) include two improvements that speed up the calculation and make the numerical refinement more stable. The first improvement concerns a correction of the search direction, which is equal to the gradient direction in (27), by the use of *conjugate gradients*. A detailed description is beyond the scope of this paper but the underlying principle is easy to explain. It is immediately clear that the local gradient of S^2 alone gives only an incomplete model of the shape of the functional dependence of S^2 close to a minimum and additional information on the curvature of S^2 will improve the situation significantly. Within the framework of gradient search, such information can be inferred at a refinement step n by exploiting knowledge of previously calculated gradients at steps $n-1, n-2, \ldots$, since the difference between the individual gradients directly reflects the curvature. In practice, it is sufficient to include the gradient at step $n-1$ for the calculation of the search direction

$$D^{(n)}(\mathbf{r}) = \delta U^{(n)}(\mathbf{r}) + \lambda^{(n)} D^{(n-1)}(\mathbf{r}), \tag{28}$$

with

$$\lambda^{(n)} = \frac{\int[\delta U^{(n)}(\mathbf{r}) - \delta U^{(n-1)}(\mathbf{r})]\delta U^{(n)}(\mathbf{r})\,d\mathbf{r}}{\int[\delta U^{(n-1)}(\mathbf{r})]^2\,d\mathbf{r}}, \tag{29}$$

which is the Fletcher–Reeves formula of the conjugate gradient, applied to the potential correction (27). For the first refinement step, where no previous gradient information is available, the search direction coincides with the gradient,

$$D^{(1)}(\mathbf{r}) = \delta U^{(1)}(\mathbf{r}). \tag{30}$$

The last three expressions, (28), (29) and (30), give a better estimate of the search direction, while the magnitude of the potential correction is still fixed.

The second improvement of the simple gradient search concerns a fine-tuning of the magnitude of the potential correction at refinement step n by introducing a dimensionless 'feed-back' parameter $\gamma^{(n)}$:

$$U^{(n+1)}(\mathbf{r}) = U^{(n)}(\mathbf{r}) + \gamma^{(n)} D^{(n)}(\mathbf{r}). \tag{31}$$

The optimum potential correction is determined by minimizing S^2 with respect to $\gamma^{(n)}$, which is in practice achieved by trying a small number of different feedback parameters and calculating a fit to the minimum. Regarding numerical efficiency, it is sufficient to calculate the optimum feedback parameter from three values of S^2, say for $\gamma^{(n)} = 0$, $\gamma^{(n)} = 1/2$ and $\gamma^{(n)} = 1$, using a parabola fit.

The gain of the operation is generally a higher convergence speed of the refinement, at the expense of the calculation of three exit wave functions related to the respective projected potentials $U^{(n)}(\mathbf{r})$, $U^{(n)}(\mathbf{r}) + D^{(n)}(\mathbf{r})/2$ and $U^{(n)}(\mathbf{r}) + D^{(n)}(\mathbf{r})$. One exit wave function, that related to $U^{(n)}(\mathbf{r})$, has to be calculated in any case, since it is needed to determine the gradient *via* (27), so that the net expense of the parabola fit is the calculation of two additional exit wave functions.

The use of the conjugate-gradient method and the fine-tuning of the feedback parameter speed up the convergence considerably, but still most of the work is spent in the calculation of the integral in (27), which involves the propagation of the entrance wave and the back-propagation of the experimental exit wave function through the crystal. In the present work, the integral is evaluated numerically using the trapezoidal rule,

$$\begin{aligned}\int_0^t \psi_s^*(\mathbf{r}, t')\psi_e(\mathbf{r}, t')\,dt' \approx{}& (\Delta t/2)\psi_s^*(\mathbf{r}, 0)\psi_e(\mathbf{r}, 0) \\ &+ \Delta t \sum_{k=1}^{N-1} \psi_s^*(\mathbf{r}, k\Delta t)\psi_e(\mathbf{r}, k\Delta t) \\ &+ (\Delta t/2)\psi_s^*(\mathbf{r}, t)\psi_e(\mathbf{r}, t),\end{aligned} \tag{32}$$

where $\Delta t = t/N$ denotes the slice thickness and the wave functions at thicknesses $t' = k\Delta t$ are conveniently calculated using the multislice algorithm (§2.1). The numerical back-propagation of the experimental exit wave function is achieved by operating the multislice algorithm with a *negative* slice thickness, which is justified by the fact that $\mathbf{P}(\mathbf{A}, -t)$ is the inverse of the scattering matrix $\mathbf{P}(\mathbf{A}, t)$, holding also for each slice of expansion (9).

In principle, two different layouts of the numerical evaluation of (32) are possible, one optimized for execution speed and another optimized for memory efficiency. In the fast version, the back-propagation of the experimental exit wave function is performed first yielding the wave functions $\psi_e(\mathbf{r}, k\Delta t)$ in descending order, $k = N, N-1, \ldots, 1, 0$, which are each kept in memory. Then the forward propagation of the entrance wave is performed yielding the wave functions $\psi_s(\mathbf{r}, k\Delta t)$ in ascending order, $k = 0, 1, \ldots, N-1, N$. At each slice, the respective wave-function product (32) is calculated and accumulated with the proper weight. In the memory-saving version, the back-propagation of the experimental exit wave function is performed from the exit plane to the entrance plane of the crystal yielding $\psi_e(\mathbf{r}, 0)$ but now without storing the wave functions at intermediate slices. Then two multislice iterations are performed in parallel, one starting with the entrance wave function and one starting with $\psi_e(\mathbf{r}, 0)$. At each slice, the respective wave-function product (32) is calculated and accumulated with the proper weight. The trade-off between execution speed and memory efficiency of the two layouts is immediately clear: The fast layout needs additional memory to store N wave functions, the memory-efficient layout needs one additional multislice iteration for the back-propagation. For the present work, the fast layout was implemented since nowadays standard workstations are equipped with sufficiently large memory.

3. Start models for the projected potential

The vital element of any successful refinement method is the choice of a good start model. The better the model, the faster the refinement will converge towards the 'true' solution. Beyond a certain mismatch between the start model and the true projected potential, refinement algorithms are even unable to restore the true solution (Gribelyuk, 1991; Beeching & Spargo, 1993; Spargo *et al.*, 1994) and they tend to converge to local minima of S^2, representing 'false' or unphysical projected potentials. Therefore, it is highly desirable to collect as much relevant physical information as possible for the first guess at the true projected potential.

The major source of information is, of course, the experimentally measured exit wave function, which in many cases directly provides structural detail, like atom column positions or hints at the scattering power of individual atom columns. In the following subsections, two useful start models are described that can be directly derived from the experimental exit wave function.

3.1. The weak-phase-object model

The relation between the projected potential and the exit wave function for a weakly scattering object imposing only a phase modulation on the incoming entrance wave has already been described in §2.3, equation (22). From the arguments given for the derivation of the scale factor C in equation (25), it is clear that the first step of the maximum-likelihood refinement is compatible with the weak-phase-object restoration of the projected potential if $U(\mathbf{r}) = 0$ is chosen as a first guess. In other words, a separate processing of the exit wave function in order to derive a first guess of $U(\mathbf{r})$ is not needed, since it is already automatically included in the refinement.

3.2. The electron channelling model

In general, it is much more difficult to set up a starting guess for the projected potential for the case of dynamical scattering from a thick specimen than for the weak-phase-object case. The exit wave function contains amplitude and phase modulations that are generally not directly related to the projected crystal potential. One important exception is the case of dynamical scattering in crystalline materials where the amplitude and phase modulations of the electron wave function remain localized close to the projected column positions (Howie, 1966; Kambe *et al.*, 1974; Fujimoto, 1978). This special scattering situation is that of *electron channelling*.

The electron wave function inside the crystal is conveniently described by a linear superposition of Bloch waves $b_k(\mathbf{r})$,

$$\psi(\mathbf{r}, z) = \sum_k \alpha_k b_k(\mathbf{r}) \exp(2\pi i \gamma_k z), \tag{33}$$

which fulfil the Schrödinger equation

$$\nabla^2 b_k(\mathbf{r}) + 4\pi^2 [U(\mathbf{r}) - 2K\gamma_k] b_k(\mathbf{r}) = 0, \tag{34}$$

with α_k being the excitation coefficient of Bloch wave k and $2K\gamma_k$ the respective energy eigenvalue. The boundary condition at the entrance plane of the crystal requires the continuity of the entrance wave and the wave inside the crystal:

$$\sum_k \alpha_k b_k(\mathbf{r}) = 1 \tag{35}$$

for a plane entrance wave.

It is quite generally known that in crystalline materials for electrons incident parallel to a low-index crystal direction the wave function inside the crystal can be well approximated by a superposition of very few Bloch waves only. If the interatomic distances are large enough and the atoms light enough, then the number of important Bloch waves may even be reduced to two, for example for electron channelling along the [110] direction in silicon and germanium at an accelerating voltage of 400 kV. In that case, the excitation coefficients of the two strongest Bloch waves have values of $|\alpha_1|^2 + |\alpha_2|^2 = 0.92$ and $|\alpha_1|^2 + |\alpha_2|^2 = 0.99$.

For the case of two Bloch waves excited, (33) is reduced to

$$\psi(\mathbf{r}, z) = \alpha_1 b_1(\mathbf{r}) \exp(2\pi i \gamma_1 z) + \alpha_2 b_2(\mathbf{r}) \exp(2\pi i \gamma_2 z). \tag{36}$$

Introducing $\Delta\gamma = \gamma_1 - \gamma_2$, $\bar{\gamma} = (\gamma_1 + \gamma_2)/2$ and $\varphi = -2\pi\bar{\gamma}z$ together with the boundary condition (35) results in

$$\psi(\mathbf{r}, z) \exp(i\varphi) = \cos(\pi\Delta\gamma z) + i[2\alpha_1 b_1(\mathbf{r}) - 1] \sin(\pi\Delta\gamma z). \tag{37}$$

From the Schrödinger equation and the boundary condition (35), an approximate relation between the projected crystal potential $U(\mathbf{r})$ and the Bloch wave $b_1(\mathbf{r})$ can be derived, with

$\xi = 1/\Delta\gamma$ being the *extinction distance* for the two-Bloch-wave case,

$$U(\mathbf{r})\xi/2K \approx \alpha_1 b_1(\mathbf{r}) - |\alpha_1|^2, \tag{38}$$

which is exact if $\nabla^2(\alpha_1 b_1(\mathbf{r}) + \alpha_2 b_2(\mathbf{r})) = 0$. Inserting the approximation (38) in (37) gives a direct relation between the projected crystal potential and the exit wave function,

$$\psi(\mathbf{r}, z)\exp(i\varphi) \approx \cos\frac{\pi z}{\xi} + i\left(\frac{U(\mathbf{r})\xi}{K} + 2|\alpha_1|^2 - 1\right)\sin\frac{\pi z}{\xi}, \tag{39}$$

which can be solved for the projected potential,

$$U(\mathbf{r}) \approx A\,\mathrm{Im}\{\psi(\mathbf{r}, z)\exp(i\varphi) - \langle\psi(\mathbf{r}, z)\exp(i\varphi)\rangle\}, \tag{40}$$

with $\langle\ldots\rangle$ denoting the spatial average with respect to $\mathbf{r}$ and

$$A = K/[\xi\sin(\pi z/\xi)]. \tag{41}$$

Equation (40) is the definition of the channelling model of the projected potential with the unknown parameters A and φ. In principle, an optimum choice of the parameters for an experimental exit wave function can be determined by trying different pairs (A, φ), calculating the respective projected potentials, simulating the related exit wave functions and finally rating these using S^2.

Within the framework of the channelling model, an optimum choice for $\exp(i\varphi)$ can be derived directly from the experimental exit wave function. Equation (39) indicates that a proper value of the unknown phase yields a complex function $\psi(\mathbf{r}, z)\exp(i\varphi)$ whose real part is constant with respect to $\mathbf{r}$. There is, of course, an ambiguity by an angle of π, which can be simply resolved since a reasonable projected potential (40) exhibits positive peaks at the atom column sites. In practice, the real part of (39) is not strictly constant with respect to $\mathbf{r}$ since (38) is an approximation and therefore the optimum phase is determined by minimizing the variance of $\mathrm{Re}\{\psi(\mathbf{r}, z)\exp(i\varphi)\}$. A complementary criterion is used by Op de Beeck & Van Dyck (1996), which exploits the 'peakedness' of the resulting projected potential map.

With the optimum choice for the phase, $\sin(\pi z/\xi)$ in (41) can be determined using the relation

$$\cos(\pi z/\xi) \approx \mathrm{Re}\{\langle\psi(\mathbf{r}, z)\exp(i\varphi)\rangle\} \tag{42}$$

up to a sign, an ambiguity that is resolved as described above. Therefore, the search for the parameter A is reduced, together with the known inverse wavelength, K, to an estimate of the extinction distance, ξ.

The electron-channelling model, although being put in a different but equivalent form from equations (39) to (41), has already been used experimentally to derive projected potential maps from exit wave functions (Op de Beeck & Van Dyck, 1996) or to add the missing phase information in dynamical diffraction intensities (Sinkler *et al.*, 1998; Sinkler & Marks, 1999).

4. Application to simulated exit wave functions

4.1. Test structures

Three different test structures are chosen in order to investigate the convergence and the accuracy of the maximum-likelihood refinement algorithm: (i) SiGe in the [110] zone-axis orientation with random Si and Ge occupancies (Fig. 1*a*); (ii) the core structure of a 90° Shockley partial dislocation terminating a stacking fault in Ge [110] (Fig. 1*b*); and (iii) an area containing randomly distributed Ge columns (Fig. 1*c*). The first two structures have a disordered zinc-blende and a diamond structure, respectively, which in the [110] projection exhibit the well known 'dumb-bell' structure of closely spaced atom columns. The three structures represent different degrees of nonperiodicity and the SiGe alloy structure allows a study of the performance of the algorithm for a reconstruction area containing columns of different scattering power.

For each structure, a thickness series of exit wave functions, $t = 4$, $t = 52$ and $t = 100$ Å, was simulated using the multislice routine implemented in the *EMS* package (Stadelmann, 1987), with an accelerating voltage of 400 kV and a slice thickness of 4 Å. In all cases, a frame size of 34×32 Å was used, sampled at a rate of 512×512 pixels. The Doyle–Turner parameterization of the electron scattering factors and a thermal vibration parameter of $B = 0.6$ Å^2 were used for the Ge and Si atoms, and absorption was neglected. The amplitude and phase of the exit wave functions at $t = 100$ Å belonging to the three test structures are displayed in Fig. 2.

The respective thicknesses were chosen in order to cover three very different scattering situations. At $t = 4$ Å, the three test objects can be described well by the weak-phase-object model where the transmitted beam still has an intensity close to 1 and the diffracted beams are only weakly excited. At $t = 52$ Å, electron scattering is strongly dynamical for the second and third test structures with the transmitted beam adopting a relatively low intensity while most diffracted beams have a relatively large intensity. At $t = 100$ Å, dynamical scattering leads to a situation where the transmitted beam regains a relatively high intensity while the diffracted beams have low intensity. The extinction distances of Ge [110] and Si [110] are $\xi_{\mathrm{Ge}} = 150$ and $\xi_{\mathrm{Si}} = 320$ Å, and the large difference between the two extinction distances leads to different local amplitude and phase modulations of the exit wave function at the Si and Ge columns of the SiGe alloy test structure. The most prominent related contrast feature is the phase wrap of the exit wave function at a specimen thickness of $t = 100$ Å, which is displayed in Fig. 2(*a*), where the local phase at the Ge columns advances the mean phase by more than π while the local phase at the Si columns advances the mean phase by less than π.

A comparison of the projected structures displayed in Fig. 1 and the respective exit wave functions displayed in Fig. 2 shows that the wave functions resemble the column structures, but in general the dynamical scattering leads to contrast features that are not directly related to the scattering power of individual columns.

4.2. Reconstruction of the projected crystal potential

The reconstruction experiments using the maximum-likelihood refinement algorithm consist of three steps. Firstly, an 'input' exit wave function is generated using the known projected crystal potential as described in the previous section. Secondly, the reconstruction of the projected crystal potential is started by calculating a first guess at the potential from the input exit wave alone using either the weak-phase-object model (§3.1) or the channelling model (§3.2). Thirdly, the maximum-likelihood refinement algorithm is applied to the previously calculated first guess in order to reconstruct the projected potential from the input wave function. Finally, the reconstructed projected crystal potential is compared with the 'generating' projected crystal potential.

In total, 18 reconstructions have been performed, for each of the three test structures at the three specimen thicknesses listed before and for the two different start models of the projected crystal potential. The maximum-likelihood refinement was successful in 15 cases and the resulting maps of the projected crystal potential match the true 'generating' projected potential extremely well. The respective graphic representations are virtually identical and Fig. 3 displays the result of the reconstruction for the three test structures at a specimen thickness of 100 Å, where the refinement was started from the channelling model. In three cases, the reconstruction algorithm was not successful, namely for the refinements starting from the weak-phase-object model at a specimen thickness of 100 Å.

Within the framework of the maximum-likelihood formalism, success or failure of a reconstruction is judged by measuring the final mismatch between the input exit wave function and the exit wave function corresponding to the reconstructed projected potential using the respective value of S^2. The reconstructions at small specimen thicknesses resulted in values of S^2 around 7×10^{-6}, whereas the successful reconstructions at larger specimen thickness resulted in larger values of around 8×10^{-4}. The reason for this slight residual mismatch is numerical round-off errors, which tend to be larger at higher specimen thickness since in that case the number of slices and hence the number of multislice steps needed is larger.

The convergence behaviour of the maximum-likelihood refinement algorithm is depicted in the plot in Fig. 4. The plot shows the value of S^2 as a function of the refinement step for the SiGe alloy test structure at a specimen thickness of 100 Å using the channelling model. The refinement starts at $S^2 = 5.0 \times 10^{-2}$, which indicates a crude fit of the first guess at the projected potential derived from the channelling model. During the refinement, the trial projected crystal potential is modified more and more towards a potential whose corresponding exit wave function is better correlated to the input exit wave function. Finally, the refinement is stopped at a value of $S^2 = 8.0 \times 10^{-4}$ since no further substantial improvements occur.

The calculation times needed for individual potential reconstructions ranged from 1 to 20 min on a 250 MHz DEC Alpha workstation with a SPECfp95 rating of 8.4, depending on the specimen thickness, the number of refinement cycles and the choice of the start model.

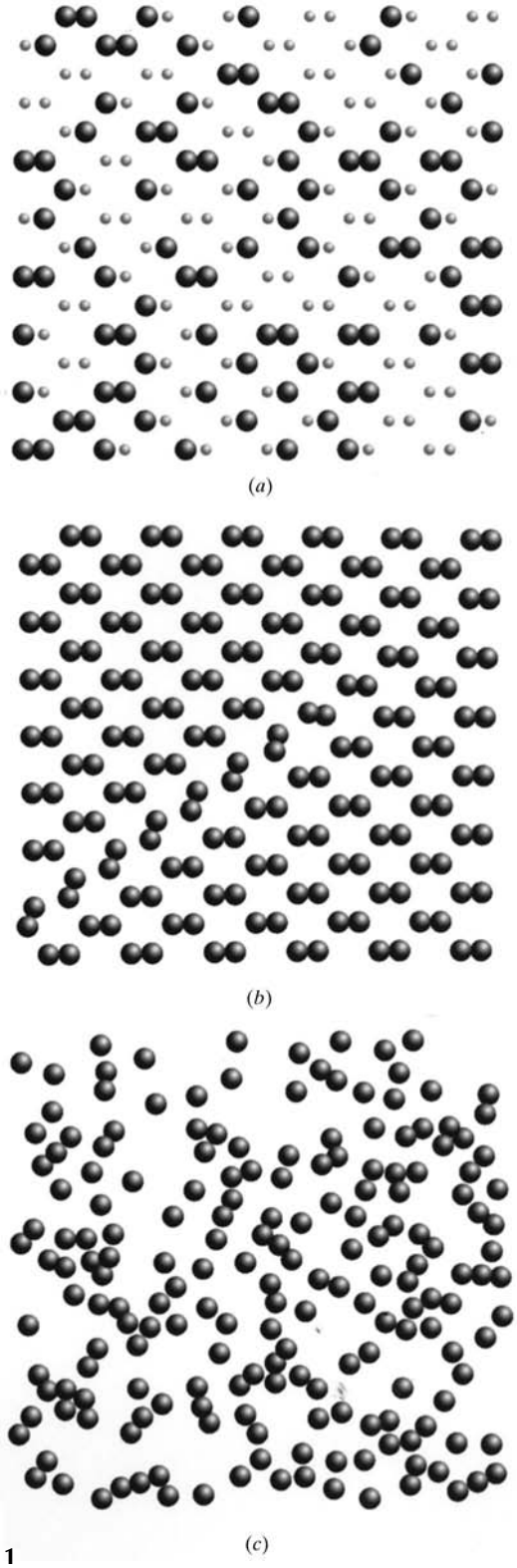

Figure 1
Projection of three test structures. Large spheres indicate Ge atom columns and small spheres Si atom columns. (*a*) SiGe [110]; (*b*) 90° Shockley partial dislocation terminating a stacking fault in Ge [110]; (*c*) randomly distributed Ge columns.

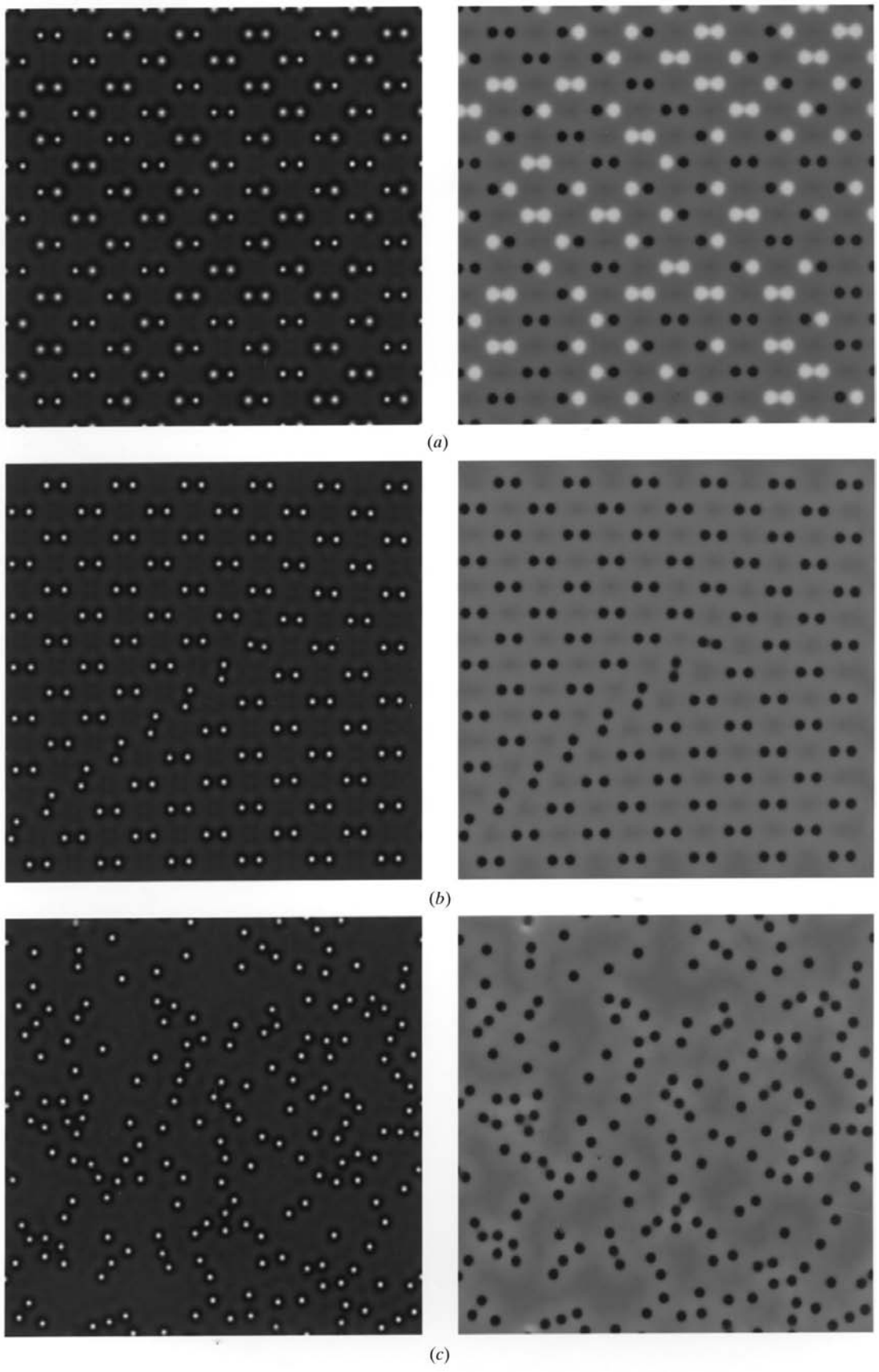

Figure 2
Simulated exit wave functions at a specimen thickness of 100 Å. The left column shows the amplitude and the right column the phase. (*a*) SiGe [110]; (*b*) 90° Shockley partial dislocation terminating a stacking fault in Ge [110]; (*c*) randomly distributed Ge columns.

5. Discussion

The maximum-likelihood refinement of the projected crystal potential from an exit wave function works very well on areas large enough to cover, for example, defects in crystalline materials. The three test structures investigated have an area of 34 × 32 Å each, which completely covers the core structure of a 90° Shockley partial dislocation terminating a stacking fault in Ge [110]. The range of specimen thicknesses achieved by refining start models derived from the channelling approximation of dynamical scattering also makes the refinement strategy suitable for practical high-resolution work in materials research.

In principle, three factors contribute to the computing power needed for the refinement. Firstly, the size of the area makes the numerical effort grow proportionally to $N_1 N_2 \log_2 N_1 N_2$, with N_1 and N_2 denoting the sampling of rows and columns of a respective rectangular frame *via* the use of fast Fourier transforms during the multislice operations. Secondly, the effort grows with the specimen thickness and it is proportional to the number of slices needed to perform the individual multislice iterations. The third contribution is proportional to the number of refinement steps needed to achieve convergence towards a low value of S^2, which clearly depends on the quality of the start model. Generally, the performance of the weak-phase-object model as well as the channelling model deteriorates at larger specimen thicknesses since in the former case the error of neglecting the amplitude modulation of the electron wave grows and in the latter case the wave function inside the crystal must be approximated by a growing number of Bloch waves.

In our fully numerical example, the precision of the reconstructed projected crystal potential is in principle as high as that of the input exit wave function. In practice, however, two limits are imposed, one by the numerical round-off error of the multislice iterations, which has already been mentioned in §4.2, and the other by the noise contained in an experimentally obtained exit wave function. The influence of noise is generally difficult to assess. For the restoration from an exit wave function of a thin specimen, it is immediately clear that a noise component will be transferred to the potential map almost directly. At larger specimen thicknesses, the transfer of weak noise from the exit wave function into the potential map will no longer be direct but still small. At a value of the specimen thickness close to the extinction distance, a reconstruction may be difficult if the nearly extinct signal component is overwhelmed by noise.

The failure of the weak-phase-object start model at a large specimen thickness can be conveniently analysed within the framework of the Bloch-wave theory of dynamical scattering. Two of the three test structures investigated, the Shockley partial dislocation and the area containing randomly distributed Ge columns, have atom columns of the same kind where the Ge-atom repeat distance along the incident electron beam is large enough to excite mainly two Bloch waves during dynamical scattering. The relation (40) between the projected crystal potential and the exit wave function takes full account of the effects of dynamical scattering by the amplitude A and the phase φ, in contrast to the weak-phase-object relation (22),

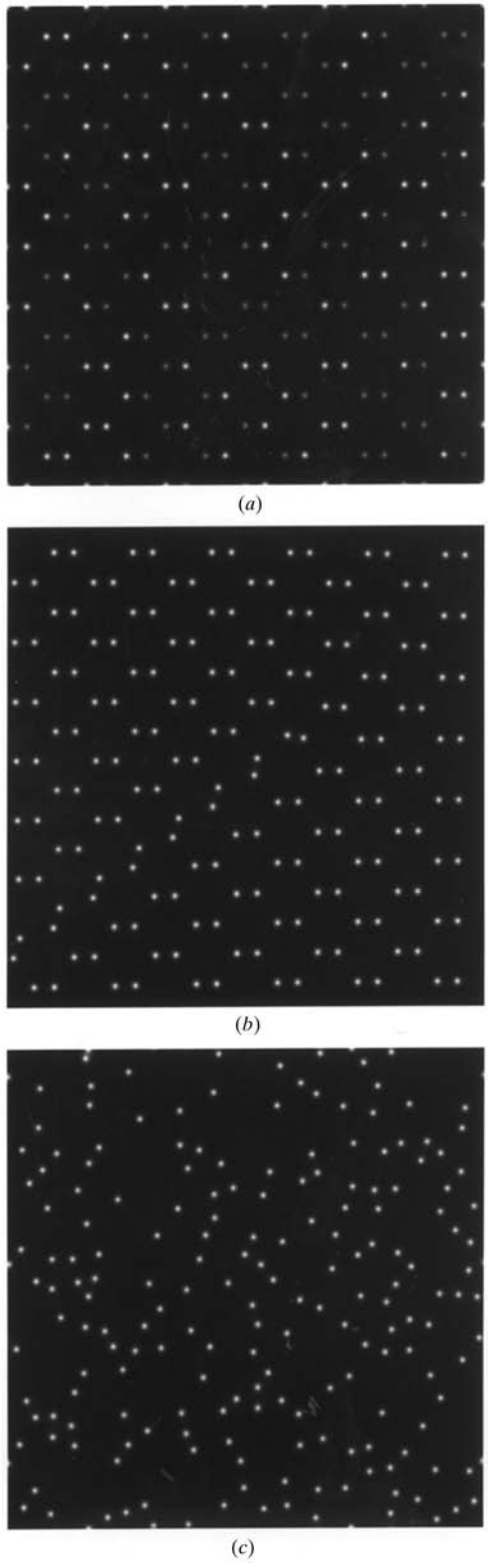

Figure 3
Projected crystal potentials reconstructed from simulated exit wave functions at a specimen thickness of 100 Å using the channelling start model. (*a*) SiGe [110]; (*b*) 90° Shockley partial dislocation terminating a stacking fault in Ge [110]; (*c*) randomly distributed Ge columns.

which 'blindly' restores the projected potential from the imaginary part of the exit wave function.

Neglecting the 'dynamical' phase φ at large specimen thickness is a serious error, which becomes clearer using an estimate of the parameters introduced in §3.2. In many materials with widely spaced atom columns, the energy eigenvalue of the second Bloch wave is close to zero, hence $\gamma_2 \approx 0$, $\xi \approx 1/\gamma_1$ and $\varphi \approx -\pi z/\xi$. For small specimen thickness, $z \ll \xi$, the dynamical phase also becomes small and the 'dynamical' amplitude (41) approaches the value $K/\pi z$. Relation (40) is then equivalent to the weak-phase-object model (22). At a specimen thickness approaching half the extinction distance, however, the dynamical phase turns to $-\pi/2$ thereby interchanging the role of the real and the imaginary part of $\psi(\mathbf{r}, z)$ in equation (40). In this situation, the channelling model correctly extracts the real part of the exit wave function while the weak-phase-object model erroneously extracts the imaginary part. At a specimen thickness just below the extinction distance, the weak-phase-object model even tries to restore an inverted version of the potential map correctly derived by the channelling model. In any case, the wrong or unphysical potential maps are too misleading to be refined successfully by the maximum-likelihood algorithm towards the correct solution.

The success of the maximum-likelihood refinement algorithm in the other test case, the SiGe alloy with random Si and Ge occupancies, can also be conveniently analysed within the framework of the Bloch-wave theory of dynamical scattering. At first sight, it seems strange that the channelling model (40), considering only two Bloch waves, is capable of setting up a reasonable start projected potential, especially at the larger specimen thickness of $t = 100$ Å. Since the SiGe alloy contains Si and Ge columns of different scattering power, at least three Bloch waves are required to properly describe dynamical scattering. Within the framework of the channelling formalism, scattering by single Si and Ge columns is modelled by two different extinction distances, and hence two different sets of the dynamical amplitude A and the dynamical phase φ, at the same specimen thickness.

Different strategies can be used to set up a proper start projected potential for this and also more complicated situations. The first one exploits the common idea of cutting the exit wave function under investigation into many small areas each containing one isolated atom column. Since high-energy electron diffraction produces local modulations of the wave function, it is possible to restore the local column potential *via* the channelling model with a set of dynamical amplitudes A and phases φ adapted for each single column. Finally, the set of the restored column potentials can be patched together to form the first guess at the projected potential for the whole area under investigation. A second and more simple approach exploits the situation that in many materials containing atom columns of different scattering power the related extinction distances do not differ too much. The extinction distance ξ enters the channelling model *via* the dynamical amplitude and phase in the form t/ξ, with t being the specimen thickness. Therefore, the start model for the whole area merely needs to represent a proper compromise of the different dynamical amplitudes and phases. The test on the SiGe alloy structure, where the latter approach was used, demonstrates that an 'overall' dynamical amplitude and phase can be successfully found even for the very differently scattering Si and Ge columns at a specimen thickness of $t = 100$ Å.

An application of the latter approach to a material comprising columns of medium scattering power and very small scattering power, such as oxygen columns, is also very likely to be successful. At specimen thicknesses where the medium scattering columns are not extinct, the channelling model, as above, will restore a good guess at the local projected potential with a certain set of the dynamical amplitude and phase. On the one hand, these values will not be well suited to restore the local projected potential at the weakly scattering columns. Since at these column sites the modulation of the exit wave function is small anyway and deviates only weakly from a plane wave, the improperly restored local projected potential is essentially close to zero, $U(\mathbf{r}) \approx 0$. On the other hand, $U(\mathbf{r}) = 0$ is a perfectly good start model for weakly scattering columns at a specimen thickness below half the extinction distance (§3.1). In the situation mentioned, the channelling model will therefore automatically produce a projected potential map ready for a successful maximum-likelihood refinement.

Besides the exit wave function, the most important input parameter of the refinement algorithm is the specimen thickness. It is *a priori* unknown and up to now there is still no structure-independent way to determine it precisely in a real high-resolution experiment. In the present work on the test structures, the specimen thickness, of course, was known in every instance and the respective proper values were transferred to the refinement algorithm.

For real experimental cases, the specimen thickness can be simply determined within the framework of the maximum-likelihood formalism by running the refinement algorithm for different guesses at the specimen thickness and finally choosing the one with the smallest value of S^2. More elegantly, the maximum-likelihood refinement algorithm can be extended by a procedure to refine a first guess at the specimen

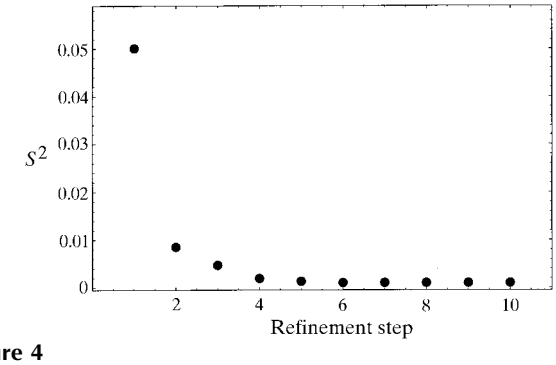

Figure 4
Convergence plot of the maximum-likelihood refinement algorithm for the reconstruction of the projected crystal potential of SiGe [110] at a specimen thickness of 100 Å based on the channelling start model. The mean squared difference S^2 is plotted *versus* the number of refinement steps.

thickness parallel to the projected potential. Following the route of §§2.3 and 2.4, the gradient of S^2 with respect to the thickness,

$$\frac{\partial S^2}{\partial t} = -2\,\mathrm{Re}\left\{\left\langle \boldsymbol{\Psi}_{\mathrm{exp}} \middle| \frac{\partial \boldsymbol{\Psi}_{\mathrm{sim}}}{\partial t} \right\rangle\right\} = 4\pi\,\mathrm{Im}\left\{\langle \boldsymbol{\Psi}_{\mathrm{exp}} | \mathbf{A} | \boldsymbol{\Psi}_{\mathrm{sim}} \rangle\right\}, \quad (43)$$

is needed to calculate the nth thickness correction

$$\delta t^{(n)} = -\frac{K^2}{2\pi^2 \langle U^2(\mathbf{r}) \rangle} \left.\frac{\partial S^2}{\partial t}\right|_{t=t^{(n)}}. \quad (44)$$

Considering numerical efficiency, it is advantageous to evaluate the first term in (43) since the multislice iterations of the maximum-likelihood algorithm provide the wave functions at thicknesses t and $t - \Delta t$, allowing a direct calculation of the gradient of the exit wave function.

In principle, the effects on the exit wave function by a change of the specimen thickness or a change of the projected potential can be separated. This can be seen from the mathematical description of dynamical diffraction using equations (6), (7) and (8), where the specimen thickness t enters *via* $s(\mathbf{g})t$ and $U(\mathbf{g})t$. Hence, the free-space propagation, described by $s(\mathbf{g})t$, contains information on the thickness independently of the projected potential. Ambiguities arise when free-space propagation can be neglected, for example for scattering from a weak phase object, where projected potential and thickness enter as a product. Another ambiguity occurs if the dynamical diffraction can be well described by the two-Bloch-wave model, where the specimen thickness can only be determined up to a multiple of the extinction distance. In practice, these ambiguities are not important since in many cases high-resolution investigations are made, on the one hand, in thin specimen areas below the first extinction distance. On the other hand, the case of a weak phase object, exhibiting a pure phase modulation of the exit wave, is rarely met for crystalline materials of medium or strong scattering power.

A general problem of all techniques determining the projected crystal potential from experimental exit wave functions is the missing information on high spatial frequencies owing to the information limit of the instrument, independent of whether it is caused by an objective aperture or by the virtual aperture imposed by the dampening envelopes of the contrast transfer function. Since, for a proper description of dynamical diffraction, Fourier coefficients of the projected potential at high and at low frequencies are equally important, the potential reconstruction is in principle under-determined. Extending the information limit will improve the situation on the one hand. On the other hand, the use of additional structural knowledge during the refinement procedure can compensate for the loss of high-frequency structure information. The latter comprises various methods of constrained optimization including the popular maximum-entropy method, which has recently been applied in high-resolution scanning transmission electron microscopy (Nellist & Pennycook, 1998).

6. Conclusions

The maximum-likelihood refinement algorithm is able to reconstruct precisely the projected crystal potential from a high-resolution exit wave function. The reconstruction can be performed on a field of view large enough to cover defects in crystalline materials and at specimen thicknesses relevant for high-resolution transmission electron microscopy. The success of the refinement algorithm is based on two major improvements: the use of start models of the projected potential for the dynamical scattering along atom columns as well as a calculation of the gradient of S^2 needed to solve the maximum-likelihood equations for a large specimen thickness. The numerical implementation of the refinement iteration using the multislice algorithm and a compact gradient calculation procedure leads to execution times of a few minutes on today's workstations.

The authors sincerely thank Dr A. Thust for valuable discussions. ML gratefully acknowledges financial support by the Volkswagen Stiftung (grant No. I/71 109).

References

Allen, L. J., Josefsson, T. W. & Leeb, H. (1998). *Acta Cryst.* A**54**, 388–398.
Allen, L. J., Leeb, H. & Spargo, A. E. C. (1999). *Acta Cryst.* A**55**, 105–111.
Beeching, M. J. & Spargo, A. E. C. (1993). *Ultramicroscopy*, **52**, 243–247.
Bethe, H. A. (1929). *Ann. Phys. (Leipzig)*, **87**, 55–128.
Coene, W., Janssen, G., Op de Beeck, M. & Van Dyck, D. (1992). *Phys. Rev. Lett.* **69**, 3743–3746.
Cowley, J. M. & Moodie, A. F. (1957). *Acta Cryst.* **10**, 609–619.
Fujimoto, F. (1978). *Phys. Status Solidi A*, **45**, 99–106.
Gerchberg, R. W. & Saxton, W. O. (1972). *Optik (Stuttgart)*, **35**, No. 2, 237–246.
Gribelyuk, M. A. (1991). *Acta Cryst.* A**47**, 715–723.
Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. (1986). *Ultramicroscopy*, **19**, 147–178.
Hirsch, P., Howie, A., Nicholson, R., Pashley, D. W. & Whelan, M. J. (1967). *Electron Microscopy of Thin Crystals.* London: Butterworths.
Howie, A. (1966). *Philos. Mag. Lett.* **14**, 223–237.
Jia, C. L. & Thust, A. (1999). *Phys. Rev. Lett.* **82**, 5052–5055.
Kambe, K., Lehmpfuhl, G. & Fujimoto, F. (1974). *Z. Naturforsch. Teil A*, **29**, 1034–1044.
Kirkland, A. I., Saxton, W. O., Chau, K.-L., Tsuno, K. & Kawasaki, M. (1995). *Ultramicroscopy*, **57**, 355–374.
Kirkland, E. J. (1984). *Ultramicroscopy*, **15**, 151–172.
Kirkland, E. J., Siegel, B. M., Uyeda, N. & Fujiyoshi, Y. (1985). *Ultramicroscopy*, **17**, 87–104.
Lentzen, M. & Urban, K. (1996). *Ultramicroscopy*, **62**, 89–102.
Lichte, H. (1986). *Ultramicroscopy*, **20**, 293–304.
Merzbacher, E. (1961). *Quantum Mechanics.* New York: John Wiley and Sons.
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1091.
Nellist, P. D. & Pennycook, S. J. (1998). *J. Microsc.* **190**, 159–170.
Op de Beeck, M. & Van Dyck, D. (1996). *Ultramicroscopy*, **64**, 153–165.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1989). *Numerical Recipes.* Cambridge University Press.
Rez, P. (1998). *Proceedings of the 14th International Congress on Electron Microscopy*, edited by H. A. Calderon Benavides & M. J. Yacamán, Vol. 1, pp. 617–618. Bristol/Philadelphia: Institute of Physics Publishing.
Rez, P. (1999). *Acta Cryst.* A**55**, 160–167.
Saxton, W. O. (1978). *Computer Techniques for Image Processing in Electron Microscopy.* London: Academic Press.
Saxton, W. O. (1988). *Scanning Microsc. Suppl.* **2**, 213–224.
Schiske, P. (1973). In *Image Processing and Computer Aided Design in Electron Optics*, edited by P. W. Hawkes. London: Academic Press.
Sinkler, W., Bengu, E. & Marks, L. D. (1998). *Acta Cryst.* A**54**, 591–605.
Sinkler, W. & Marks, L. D. (1999). *Ultramicroscopy*, **75**, 251–268.
Spargo, A. E. C., Beeching, M. J. & Allen, L. J. (1994). *Ultramicroscopy*, **55**, 329–333.
Speer, S., Spence, J. C. H. & Ihrig, E. (1990). *Acta Cryst.* A**46**, 763–772.
Spence, J. C. H. (1998). *Acta Cryst.* A**54**, 7–18.
Stadelmann, P. A. (1987). *Ultramicroscopy*, **21**, 131–146.
Sturkey, L. (1957). *Acta Cryst.* **10**, 858–859.
Sturkey, L. (1962). *Proc. Phys. Soc.* **80**, 321–354.
Thust, A., Lentzen, M. & Urban, K. (1994). *Ultramicroscopy*, **53**, 101–120.
Van Dyck, D., Danckaert, J., Coene, W., Selderslaghs, E., Broddin, D., Van Landuyt, J. & Amelinckx, S. (1989). In *Computer Simulation of Electron Microscope Diffraction and Images*, edited by W. Krakow & M. O'Keefe. Warrendale, PA: The Minerals, Metals and Materials Society.
Van Dyck, D., Op de Beeck, M. & Coene, W. (1993). *Optik (Stuttgart)*, **93**, No. 3, 103–107.