Application of Direct Methods to Dynamical Electron Diffraction Data for Solving Bulk Crystal Structures

W. Sinkler,* E. Bengu and L. D. Marks

*Northwestern University, Department of Materials Science and Engineering, Evanston, IL 60208-3108, USA.
E-mail: sinkler@apollo.numis.nwu.edu*

Abstract

The crystal structures of two previously unknown bulk phases in the Ga–In–Sn–O system have recently been solved using direct methods applied to electron diffraction intensities. In both cases, phasing of dynamical diffraction intensities clearly indicated the positions of O atoms in the crystal structures. It is shown here that a correlation between the dynamical diffraction amplitudes and the Fourier components of $|1 - \psi(\mathbf{r})|$ enables direct methods using dynamical intensities to restore structural information present in $|1 - \psi(\mathbf{r})|$. Both the presence of atom-like peaks in $|1 - \psi(\mathbf{r})|$ as well as the emphasis of light atoms are explained using electron channeling theory. Similar results can be expected for any structure consisting of well resolved atomic columns parallel to the zone-axis direction for which data are recorded. With $(Ga, In)_2SnO_5$ as a model structure, it is shown that the combination of strongly dynamical electron diffraction with direct methods is a powerful technique for detecting light-atom positions in bulk inorganic crystal structures without the need to grow single crystals.

1. Introduction

Direct methods for crystal structure determination are unique in their ability to accurately determine the positions of atoms within a crystalline unit cell using only measured diffraction amplitudes. The power of this approach is evident in the long list of organic and inorganic structures that have been solved using direct methods. The theoretical basis of direct methods has been developed under the assumption that the diffraction amplitudes are purely kinematical and the prior application of direct methods has thus largely been to cases in which complete kinematical diffraction data sets are available from single-crystal X-ray measurements. Such measurements are associated with significant experimental effort and are dependent on the ability to obtain a single crystal. This has been a motivation for recent work applying direct methods to electron diffraction data for solving organic structures (Dorset, 1996; Gilmore et al., 1993) and surface reconstructions

(Marks et al., 1997; Landree et al., 1998; Collazo-Davila et al., 1997; Gilmore et al., 1997). While surface diffraction data are rather well established to be very close to kinematical in character (Tanishiro & Takayanagi, 1989; Marks et al., 1991), there is always a worry with bulk applications. One cannot argue that the diffraction is kinematical based solely upon the fact that only light atoms are present in moderately thin samples. The existence of kinematical conditions may at best be asserted if pseudo-random atomic positions lead to cancellation of dynamical effects (Marks, 1988).

Recently, we have reported the solution of a bulk inorganic $(Ga, In)_2SnO_5$ ceramic oxide structure using direct methods applied to electron diffraction data (Sinkler et al., 1998). This represents an extension of direct methods to a case where dynamical effects cannot be treated as an error of limited significance, but rather dominate the data causing them to deviate severely from the kinematical case. The procedure followed in solving the $(Ga, In)_2SnO_5$ structure was quite straightforward: phasing of electron diffraction amplitudes was performed using a minimum relative entropy algorithm described elsewhere (Marks & Landree, 1998). The direct method was not performed *a priori* but rather the phases of some low-order beams were fixed using information from the Fourier transform of a high-resolution transmission-electron-microscope (HRTEM) image. One of the top-ranked phasing solutions is shown in Fig. 1 (Sinkler et al., 1998). Superimposed on the figure is a ball-and-stick projection of the $(Ga, In)_2SnO_5$ structure, showing O-atom positions that were subsequently confirmed using powder neutron diffraction (Sinkler et al., 1998; Edwards et al., 1998). As is apparent from Fig. 1, strong peaks occur in the phasing solution at the positions of the O atoms in the structure. The detection of O-atom positions by direct methods is evidently a dynamical effect, as it is contrary to what is expected in the kinematical case, namely an emphasis of the heavier cation atoms in the structure. This indicates that, rather than constituting an undesirable error source in the data, dynamical effects can in fact be useful in direct methods. This contrasts with earlier work using direct methods of electron diffraction intensities (Fan et al., 1991; Fu et al., 1994; Hu et al., 1992; Dong et al.,

1992), in which special care was taken to avoid dynamical effects by using thin specimen regions and the results were interpreted within a kinematical approximation. We have recently solved an additional structure in the Ga–In–Sn–O system using the same approach (Edwards *et al.*, 1998), adding further weight to the viability of direct methods using dynamical diffraction data as a technique for solving inorganic structures, and in particular for locating light-element atom positions in ceramics.

The present work explores the physical basis for the ability of direct methods in combination with dynamical electron diffraction to locate light atoms. We show that direct methods do not restore the complex exit wave $\psi(\mathbf{r})$ in real space or the crystal potential but instead the modulus of the Babinet, or $|1 - \psi(\mathbf{r})|$, which corresponds to the most probable distribution, a reasonably good approximation to the diffracted wave for moderately thin samples (< 200 Å). The enhanced detectability of the light atoms arises as a consequence of dynamical diffraction and is modeled using channeling theory. This also rationalizes previous successes in determining organic structures *via* direct methods using electron diffraction.

The outline of this paper is as follows. First, we will briefly review some of the key features of the channeling theory of electron diffraction, which leads to the conclusions that on a zone axis the wavefunction is flat, except around the sites of the atomic strings. In §3, it is demonstrated that the information content inherent in the dynamical intensities is to a good approximation that visible as peaks in $|1 - \psi(\mathbf{r})|$. In addition, the viability of incorporating information from HRTEM images to allow fixing of some phases is discussed. Following this, calculated $|1 - \psi(\mathbf{r})|$ for the (Ga, In)$_2$SnO$_5$ structure are shown and the tendency for peaks to appear in $|1 - \psi(\mathbf{r})|$ at atom positions, and more particularly at light-atom positions, in oxide structures is explained in terms of electron channeling theory. Finally, phasing solutions are presented from both simulated and experimental diffraction data, and a correspondence is shown between peaks in the solutions and in $|1 - \psi(\mathbf{r})|$.

2. Theory I. Interpretation of dynamical exit waves using channeling theory

The central characteristics of the electron wavefunction $\psi(\mathbf{r})$, the occurrence of oscillations at atom positions as a function of crystal thickness and the dependence of the frequency and amplitude of the oscillations on atomic number can be explained using electron channeling theory (Berry, 1971; Kambe *et al.*, 1974; Tamura & Ohtsuki, 1974; Fujimoto, 1978; Gemmell, 1974; Kambe, 1982; Ohtsuki, 1983; Marks, 1985; Van Dyck, 1985; Van
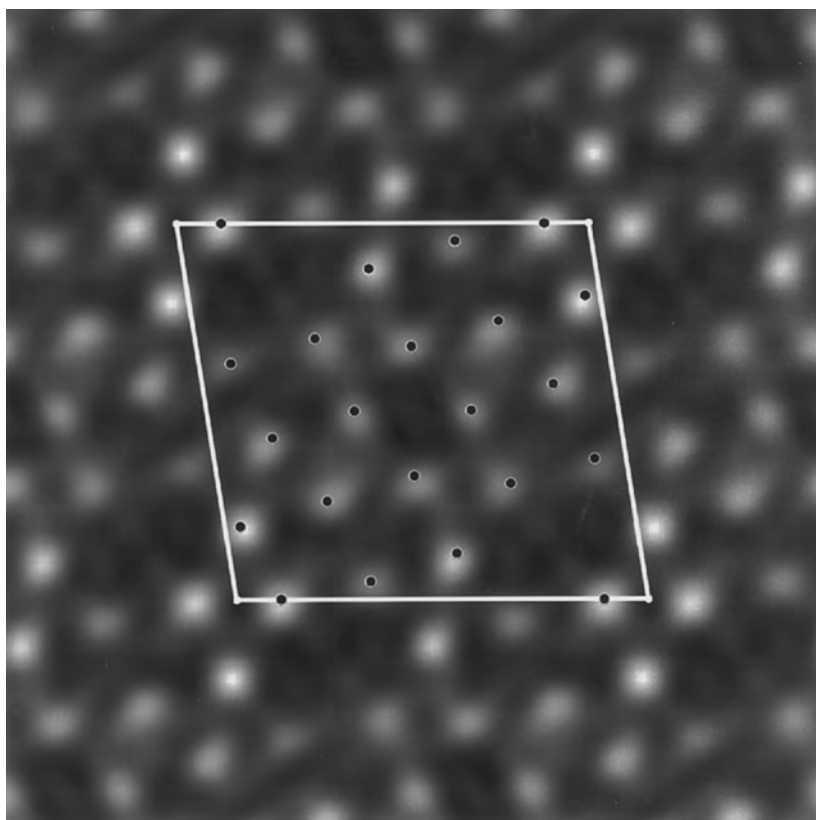

Fig. 1. Phasing solution from experimental data for (Ga, In)$_2$SnO$_5$ (Sinkler *et al.*, 1998). Superimposed is a model of the neutron-refined structure showing O-atom positions.

Dyck & Op de Beeck, 1996). Channeling theory starts from the premise that all diffraction contributing to the exit wave is from the zero-order Laue zone for a given zone-axis orientation. This approximation significantly reduces the complexity of the Bloch-wave solution to the Schroedinger equation for fast electrons, while preserving in quantitative form many of its features. By making the channeling approximation, the total wave function is reduced to

$$\psi(\mathbf{R}, z) = \sum_n C_n \Phi_n(\mathbf{R}) \exp[-i\pi(E_n/E_0)kz], \quad (1)$$

in which $\mathbf{R}$ is a two-dimensional vector perpendicular to the incident-beam direction z, E_0 is the incident-beam energy and $k = [(2mE_0)/h]^{1/2}$ is the relativistically corrected incident-beam wavevector. Equation (1) expresses the total wave as a superposition of two-dimensional Bloch waves, each having a characteristic oscillation along $\mathbf{z}$. The strength of the oscillation is given by the value of E_n. In terms of the exact Bloch-wave theory (Hirsch *et al.*, 1977), E_n is given by $E_n = E - (h^2/2m)k_z^2$, with k_z being the z component of the wave (parallel to the incident-beam direction) and E the total Bloch-wave energy. The E_n are therefore the quantized transverse or in-plane energies of the wave. By direct substitution of (1) into the three-dimensional Schroedinger equation for fast electrons (Hirsch *et al.*, 1977; Spence, 1988), one obtains an equation for the two-dimensional Bloch functions $\{\Phi_n(\mathbf{R})\}$ and the corresponding energies $\{E_n\}$:

$$\nabla_R^2 \Phi + (8\pi^2 m/h^2)[E - \rho(\mathbf{R})]\Phi = 0. \quad (2)$$

In this equation, ∇_R^2 is the Laplace operator in two dimensions and $\rho(\mathbf{R})$ is the specimen potential averaged along the beam direction (Lindhard or string potential) (Ohtsuki, 1983; Gemmell, 1974) given by

$$\rho(\mathbf{R}) = (1/d) \int_{-\infty}^{\infty} \rho(x, y, z) \, \mathrm{d}z, \quad (3)$$

in which d is the repeat distance along the electron beam.

Equation (2) has been investigated previously (Berry, 1971; Kambe *et al.*, 1974; Fujimoto, 1978; Van Dyck, 1985; Van Dyck & Op de Beeck, 1996) and, for crystalline projections in which the atoms are located in straight well separated atomic columns aligned with the electron beam, there are generally only one or two localized bound-state solutions Φ_i to (2). When the overlap between neighboring columns is minimal, the bound states are radially symmetric analogous to atomic $1s$ and $2s$ states. In addition to these bound states, there is a 'continuum' of unbound states with higher transverse energy. The unbound states, which dominate in the flat interstitial regions, are orthogonal to the bound states and may be rigorously obtained using the orthogonalized plane-wave method (OPW) (Ziman, 1972). Rather than interacting with the full potential, the OPW

states interact with the standard 'pseudopotential' with substantial cancellation of the deep well at the atomic sites – as such they are almost kinematical for a reasonably thin sample. Bearing this in mind, an approximation to the unbound states may be obtained by making use of the boundary condition at $z = 0$, which is $\sum_i C_i \Phi_i = 1$ (for unit plane-wave incidence). One can then rewrite (1) as (Van Dyck & Op de Beeck, 1996)

$$\psi(\mathbf{R}, z) = 1 + \sum_n C_n \Phi_n(\mathbf{R})\{\exp[-i\pi(E_n/E_0)kz] - 1\}. \quad (4)$$

If the sum over all states is now replaced by a single bound state Φ_b, one obtains

$$1 - \psi(\mathbf{R}, z) = C_b \Phi_b\{1 - \exp[-i\pi(E_b/E_0)kz]\}. \quad (5)$$

This satisfies the boundary condition and is equivalent to approximating the OPW states with a single unbound state Φ_0 with $E = 0$ and $C_0 \Phi_0 = 1 - C_b \Phi_b$. The same equation has also been derived by Van Dyck & Op de Beeck (1996), using different arguments. It represents the Babinet function $1 - \psi(\mathbf{r})$ for well separated atomic columns in the case in which a single channeling state Φ_b is dominant and will work well for moderately thin crystals, far better than a simple kinematical model. (For thicker crystals, the unbound states cannot be ignored.) For atomic number $Z \leq 35$ and $d \simeq 3$ Å, a single $1s$ channeling state is dominant for foils less than 200 Å thick because either there is no $2s$ state or the $2s$ energy is very small.

Equation (2) can be readily solved numerically for the case of an isolated atomic column using the separation of variables employed by Tamura & Ohtsuki (1974) and atomic potentials obtained by inverting the electron scattering factors of Doyle & Turner (1968).† Fig. 2 shows a plot of the eigenvalues of the $1s$ and $2s$ states for $d = 3.0$ Å and 300 kV accelerating voltage. A trend of linearly increasing eigenvalues with increasing Z is observed, as also reported by Van Dyck & Op de Beeck (1996). The channeling theory gives a satisfactory explanation for the oscillations with depth that characterize the electron exit wave and the main dependencies of these oscillations. This is illustrated in Fig. 2 by comparing the $1s$ channeling eigenvalues with $2E_0\lambda/z'$, where z' is the period of the oscillations in $|1 - \psi(\mathbf{r})|$ as a function of crystal thickness in multislice simulations for an isolated atomic column. A quantitative agreement with the multislice simulations is found for atomic numbers less than approximately 26. For larger-Z elements, the excitation errors, which are neglected in the channeling approximation, become more important and cause the channeling eigenvalues to underestimate the actual oscillation frequency. The role

---

† Fortran code for calculating channeling eigenvalues for an atomic column can be downloaded at www.numis.nwu.edu/ftp/direct/channeling.

of excitation errors in causing disagreement at large Z is consistent with improved agreement for both increasing accelerating voltage and decreasing interatomic spacing d.

In the case that the separation of neighboring atomic columns is sufficient to minimize overlap of the potentials (and thus the bound-state wavefunctions), one may consider the total wavefunction of a crystal as made up of the separate contributions of individual atomic column states. This has been suggested by Van Dyck & Op de Beek (1996) and results in a Babinet wavefunction given by

$$1 - \psi(\mathbf{R}, z) = \sum_i C_i \Phi_i (R - R_i)\{1 - \exp[-i\pi(E_i/E_0)kz]\} \tag{6}$$

where Φ_i is the appropriate bound-state eigenfunction for the atomic column centered at R_i. Experimental electron diffraction data consist of the amplitudes $|\Psi(\mathbf{h})|$, which are the same in their modulus (except the direct beam) as the Fourier coefficients of the Babinet. Because the Babinet oscillates rapidly with depth, the diffraction data will depend sensitively on thickness and will not represent a fixed characteristic of the sample structure, such as the electrostatic potential $\rho(\mathbf{r})$ in the case of kinematical scattering. In spite of the complex oscillatory dependence on crystal thickness (with Z-dependent frequency), channeling theory and in particular equation (6) permit one to generalize that $1 - \psi(\mathbf{r})$ will have a form that is relatively constant and close to zero in the interatomic regions provided that the unbound states can be ignored. The Babinet will be

Fig. 2. Plot of channeling eigenvalues for single atomic columns for 300 kV incident electrons and atomic spacing $d = 3$ Å along the columns. For comparison, the values of $2E_0\lambda/z'$ are also plotted, where z' is the oscillation periodicity in $|\psi(\mathbf{r}) - 1|$ at atomic columns obtained from multislice calculations.

peaked near the projected positions of atomic columns and the phase will vary linearly as a function of thickness.

3. Theory II. Interpretation of phasing results from strongly dynamical amplitudes

Dynamical electron diffraction amplitudes correspond to a real-space electron exit wave $\psi(\mathbf{r})$ that is in general complex. This contrasts with the kinematical case in which

$$F(-\mathbf{h}) = F^*(\mathbf{h}) \tag{7}$$

(F being the structure factor, $\mathbf{h}$ a reciprocal-lattice vector). Because of (7), the Fourier transform (FT) of $F(\mathbf{h})$ is real in the kinematical case and is the electrostatic potential $\rho(\mathbf{r})$ (for electrons) or electron density (for X-rays). In applying direct methods to electron diffraction, the complex nature of the wave in real space is suppressed by insisting that $F(-\mathbf{h}) = F^*(\mathbf{h})$ in the solutions. While this can be justified by a kinematical approximation, the presence of strong dynamical effects in the diffraction data combined with (7) causes a departure of the phasing solutions from any actual physical distribution.

From the above, one would presume that direct methods cannot be applied to dynamical diffraction but we will show here that this is not correct. To see why, one needs to look rather deeper at what direct methods actually do, not at a simple level of Σ_1 or Σ_2 relationships but in the general case. The basic theory of direct methods arises from probabilistic relationships obeyed by a set of (pseudo-random) 'atomistic' peaks in real space, ideally δ functions, surrounded by regions of essentially zero amplitude. Using any of a number of different multisolution techniques, plausible phases are obtained for the diffracted amplitudes that are highly probable based upon statistical relationships. If the problem is well posed, $e.g.$ approximately correct removal of the atomic scattering factors, information to about 1 Å^{-1} and small measurement errors, the phases will approach the level where a restoration of the wavefunction [or $\rho(\mathbf{r})$] is achieved.

When one has dynamical diffraction, at least for a moderately thin sample (< 200 Å), one still has features in the channeling eigenfunctions that are localized about projected atomic columns. However, the electron wave is complex in real space rather than real. Also, in the regions away from the atoms, instead of having zero amplitude the wavefunction is essentially constant. This constant term is removed by considering, instead of $\psi(\mathbf{r})$, the Babinet $1 - \psi(\mathbf{r})$, which as shown above is essentially zero away from the atoms, and has a peaked form near atomic columns. The Babinet's Fourier coefficients are identical in their modulus to those of $\psi(\mathbf{r})$ (except the transmitted beam). If we consider the probability of a given Babinet in an information-theory
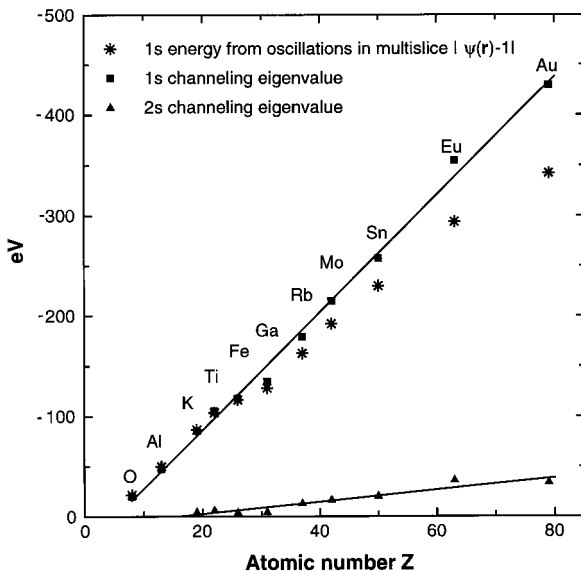
sense, this probability is independent of the phase. Using classical maximum-entropy arguments, the most probable distribution consistent with the data will maximize the entropy S, given by

$$S = - \sum_r |1 - \psi(\mathbf{r})| \ln \left[|1 - \psi(\mathbf{r})| \Big/ \sum_r |1 - \psi(\mathbf{r})| \right]. \quad (8)$$

A distribution consistent with this may also be obtained by minimizing the relative entropy $S_r(\mathbf{r})$ given by (Marks & Landree, 1998)

$$S_r(\mathbf{r}) = |1 - \psi(\mathbf{r})| \ln[|1 - \psi(\mathbf{r})| / e\langle|1 - \psi(\mathbf{r})|\rangle] \\ + \langle|1 - \psi(\mathbf{r})|\rangle. \quad (9)$$

The use of (9) to minimize relative entropy, and the relationship between (8) and (9) are described in more detail in Appendix A. Note that the actual phase of $1 - \psi(\mathbf{r})$ does not come into either of the above expressions. This suggests that direct methods will restore the function $|1 - \psi(\mathbf{r})|$, provided that the Fourier coefficients of the latter are comparable to the measured $|\Psi(\mathbf{h})|$.
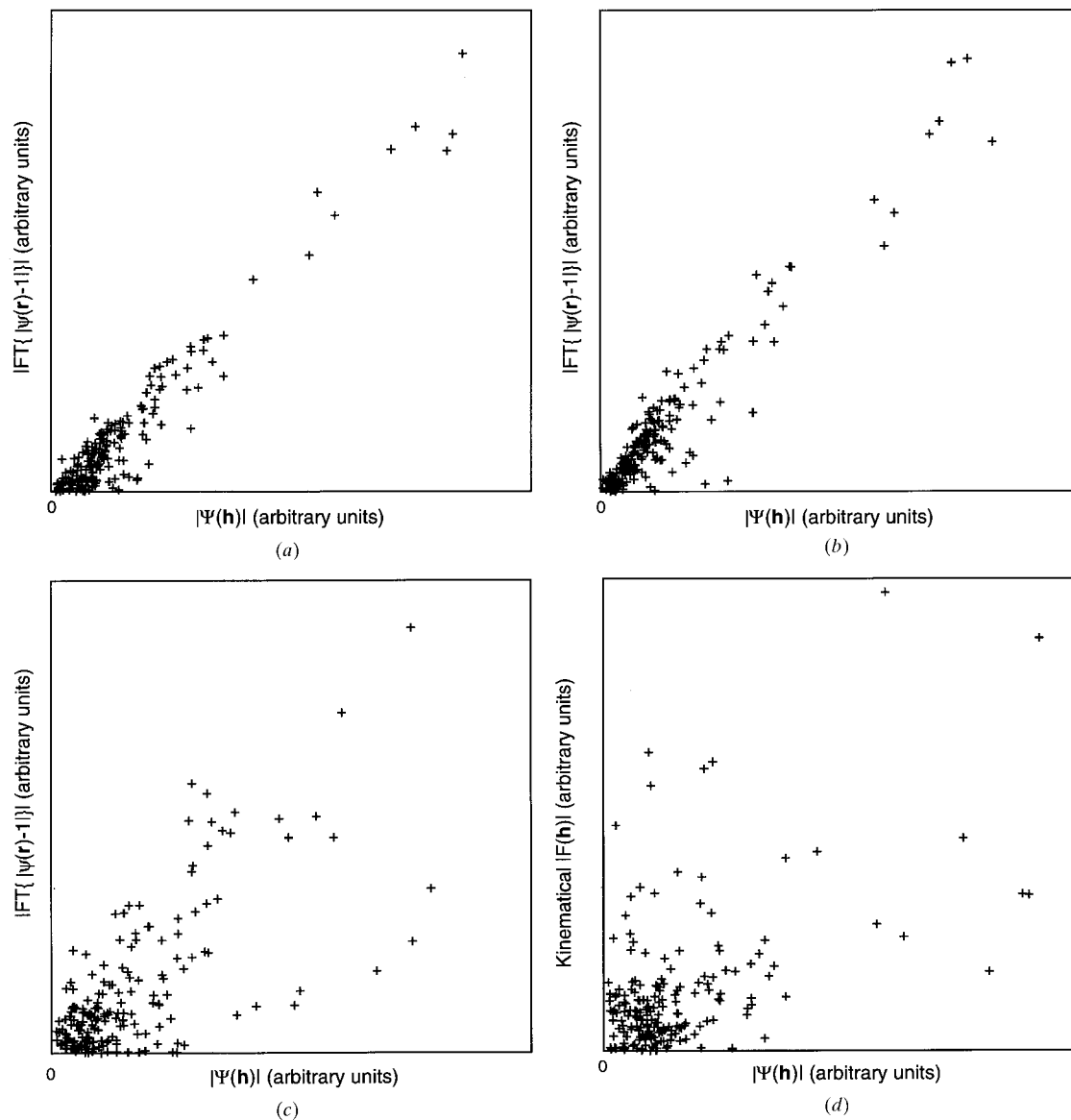

Fig. 3. (a)–(c) Plots of Fourier coefficients of $|\psi(\mathbf{r}) - 1|$ against diffraction amplitudes $|\Psi(\mathbf{h})|$ for all unique spots $\mathbf{h} = (h0l)$ [except (000)] in the [010] zone axis of the $(Ga, In)_2SnO_5$ structure, with $|\mathbf{h}| < 1$ Å^{-1}. All quantities are calculated using the multislice method. (a) 100 Å thickness. (b) 200 Å thickness. (c) 300 Å thickness. (d) For comparison with (a)–(c), a plot of the kinematical structure-factor amplitudes $|F(\mathbf{h})|$ against $|\Psi(\mathbf{h})|$ at 100 Å thickness is shown.

In the kinematical case, $|1 - \psi(\mathbf{r})|$ is proportional to the electrostatic potential $\rho(\mathbf{r})$ and the Fourier coefficients of $|1 - \psi(\mathbf{r})|$ are therefore exactly proportional to the structure factors $|F(\mathbf{h})|$. Whereas for a dynamical exit wave the Fourier coefficients of $|1 - \psi(\mathbf{r})|$ are not identical to the diffraction amplitudes $|\Psi(\mathbf{h})|$, the two sets tend nevertheless to be correlated over thicknesses well beyond those for which the kinematical approximation, *i.e.* $|\Psi(\mathbf{h})| \simeq \mathrm{FT}(\rho(\mathbf{r}))$, breaks down. This is illustrated for the example of $(\mathrm{Ga}, \mathrm{In})_2\mathrm{SnO}_5$ in Figs. $3(a)$–(c), which show plots of the diffraction amplitudes $|\Psi(\mathbf{h})|$ calculated using multislice *versus* the Fourier components of $|1 - \psi(\mathbf{r})|$, for thicknesses of 100, 200 and 300 Å. The two sets of Fourier amplitudes are closely correlated to at least 200 Å particularly for the stronger beams and a significant loss of correlation is found only for thicknesses near 300 Å. For comparison, Fig. $3(d)$ shows that there is little or no correlation between $|\Psi(\mathbf{h})|$ and $|F(\mathbf{h})| = \mathrm{FT}(\rho(\mathbf{r}))$ at 100 Å: simple restoration of the specimen potential using data for 100 Å thickness is out of the question. However, because of the significant correlation between the amplitudes $|\Psi(\mathbf{h})|$ and $\mathrm{FT}(|1 - \psi(\mathbf{r})|)$, structural information contained in the peaks of $|1 - \psi(\mathbf{r})|$ represent structural features that may be found using direct methods.

An additional real-space distribution that is of importance to direct methods in the present work is that of a HRTEM image, as it is this distribution $I(\mathbf{r})$ at Scherzer defocus that is used to fix the phases of some beams input into the direct-methods algorithm. The use of HRTEM information to fix the phases of some diffracted beams for direct methods has been performed previously (Fan *et al.*, 1991; Fu *et al.*, 1994; Hu *et al.*, 1992; Dong *et al.*, 1992), but always within the framework of a kinematical approximation. On first consideration, it may seem surprising that image phases can be used towards restoring $|1 - \psi(\mathbf{r})|$ without special precautions to ensure that both diffraction and image data represent the same thickness. Phases from HRTEM taken at Scherzer defocus can however be incorporated, as long as the phases chosen are from strong beams and are restricted to small spatial frequency to minimize dynamical and non-linear imaging effects. This can again be understood based on the example of $(\mathrm{Ga}, \mathrm{In})_2\mathrm{SnO}_5$. From the centrosymmetry of the structure, the $\mathrm{FT}(|1 - \psi(\mathbf{r})|)$ has all phases equal to either 0 or 180°. The incorporation of phases from the image toward a restoration of $|1 - \psi(\mathbf{r})|$ depends on the beams chosen having as few as possible phase switches with depth. The number of phase switches in a sample of 200 Å thickness is plotted for all $(h, 0, l)$ Fourier components of $|1 - \psi(\mathbf{r})|$ with $|\mathbf{h}| \leq 1 \ \mathring{\mathrm{A}}^{-1}$ against the kinematical diffraction amplitude $|F(\mathbf{h})|$ in Fig. $4(a)$ and against the spatial frequency $|\mathbf{h}|$ in Fig. $4(b)$. It can be seen from the plots that the tendency for a Fourier component of $|1 - \psi(\mathbf{r})|$ to switch phase increases with both decreasing amplitude and increasing spatial frequency. If one

regards only those beams that are in the top 20% of intensities and which have spatial frequencies less than $0.45 \ \mathring{\mathrm{A}}^{-1}$, only 5 of 17 or 30% show any phase switch in a 200 Å thickness. These beams are therefore pseudo-kinematical in the sense that, while their intensities are dynamical, their phases are kinematical or static with depth. For beams with frequencies below the Scherzer resolution limit, typically in the range of $0.5 \ \mathring{\mathrm{A}}^{-1}$, the effect of the lens optics on the image phases is simply the addition of a factor of π. If in obtaining phase information from a HRTEM image one is careful to choose strong beams of small spatial frequency, these phases can be used with relative confidence as representative of $|1 - \psi(\mathbf{r})|$ over a broad range of thickness. (Of course, one also has to be careful to avoid non-linear imaging effects which tend to be more important at higher spatial frequencies.)

(a)
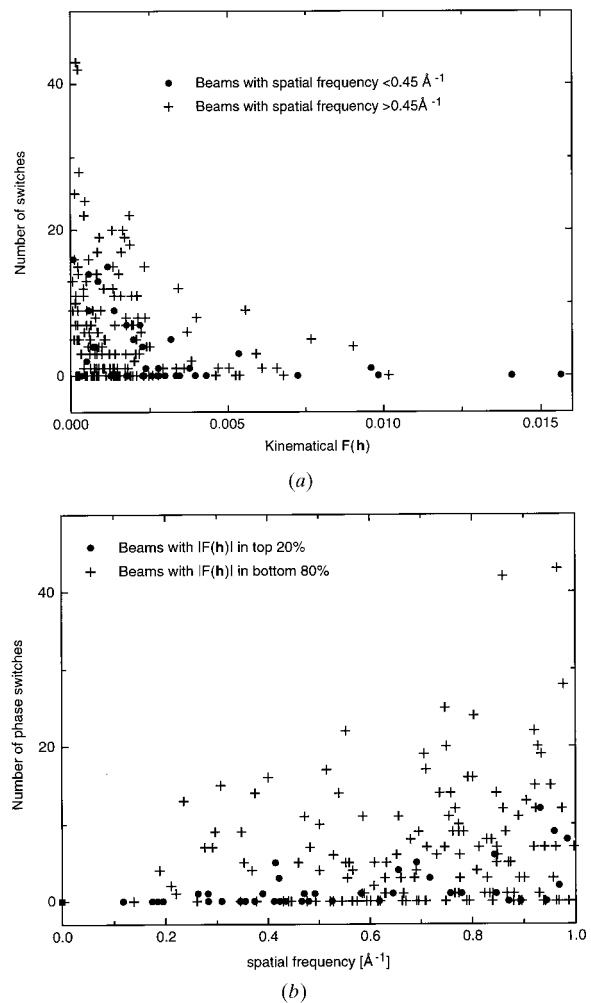

(b)

Fig. 4. Plots of the number of phase switches with depth (0 to 180° or *vice versa*) in the Fourier coefficients of $|\psi(\mathbf{r}) - 1|$ for points $\mathbf{h} = (h0l)$ in the [010] zone axis of the $(\mathrm{Ga}, \mathrm{In})_2\mathrm{SnO}_5$ structure. (a) Dependence on kinematical scattering amplitude $|F(\mathbf{h})|$. (b) Dependence on spatial frequency $|\mathbf{h}|$.

Table 1. *Neutron-refined crystal structure data for* $(Ga, In)_2SnO_5$ *(Sinkler et al., 1998)*

Space group $P2/m$ (No. 10), $a = 11.68919$ (18), $b = 3.16734$ (5), $c = 10.73068$ (15) Å, $\gamma = 98.9997$ (15)°. $R_p = 0.0495$, $wR_p = 0.0761$, $\chi^2 = 2.164$.

Atom	x	y	z	$U_i \times 100$ (Å)
Sn1	0.0	0.0	0.0	1.2 (30)
Sn2	0.5	0.0	0.0	−0.04 (21)
Sn3	0.5918 (6)	0.5	0.3112 (7)	0.35 (16)
In/Ga1†	0.3281 (6)	0.0	0.3859 (7)	0.09 (25)
In/Ga2‡	0.0756 (8)	0.5	0.3053 (9)	1.10 (30)
Ga1	0.1500 (5)	0.0	0.6022 (6)	0.94 (18)
Ga2	0.2624 (5)	0.5	0.0869 (5)	0.41 (17)
O1	0.3320 (7)	0.0	0.0472 (8)	1.09 (24)
O2	0.0365 (6)	0.0	0.1941 (8)	1.14 (22)
O3	0.4789 (7)	0.0	0.3245 (8)	1.37 (24)
O4	0.7096 (6)	0.0	0.3076 (6)	0.63 (17)
O5	0.1490 (7)	0.0	0.4287 (9)	1.77 (27)
O6	0.1081 (7)	0.5	0.0003 (7)	0.24 (17)
O7	0.5515 (7)	0.5	0.1209 (8)	0.96 (23)
O8	0.2570 (6)	0.5	0.2617 (8)	0.41 (19)
O9	0.3586 (6)	0.5	0.4987 (7)	1.17 (19)
O10	0.0768 (7)	0.5	0.6272 (8)	0.65 (21)

† Indium occupancy 0.48.   ‡ Indium occupancy 0.72.

Based on the important role of $|1 - \psi(\mathbf{r})|$ as the function upon which direct methods act, $|1 - \psi(\mathbf{r})|$ was investigated in the model case of the $(Ga, In)_2SnO_5$ structure shown in Fig. 5. The atomic coordinates for $(Ga, In)_2SnO_5$ obtained by neutron refinement are listed in Table 1 (Sinkler *et al.*, 1998). Multislice calculations for obtaining $|1 - \psi(\mathbf{r})|$ were performed using *NUMIS* multislice software developed at Northwestern University. The calculations were performed for 300 kV incident radiation and did not include absorption effects. Thermal vibration was included in the calculation corresponding to Debye–Waller factors of $B = 0.3$ Å^2 for the metal atoms and 0.5 Å^2 for O atoms. Examples of $|1 - \psi(\mathbf{r})|$ calculated for several thicknesses are shown in Figs. 6(*a*)–(*f*). By comparison with Fig. 5, it can be seen that up to thicknesses greater than 200 Å the peaks in $|1 - \psi(\mathbf{r})|$ are located at atom positions. As predicted from channeling theory, the peaks show a tendency to oscillate with thickness such that for a given thickness they are located at a subset of the atoms in the structure. For instance, in Fig. 6(*b*), the Ga atoms appear particularly bright, Sn atoms are weak and the O atoms have peaks roughly equal to those at Ga/In positions, while, in Fig. 6(*c*), the Sn atoms are weak, Ga and Ga/In atoms are virtually extinct and the strongest peaks are at O-atom positions. At all thicknesses between about 40 and 200 slices (63 to 320 Å), strong peaks were found to occur at O-atom positions. Fig. 7 investigates peak strength *versus* thickness in more detail, showing a plot of $|1 - \psi(\mathbf{r})|$ for $\mathbf{r}$ at the positions of several atomic columns in the structure as a function of depth. There is a well defined oscillation in the wave amplitude at the atom positions and the period of the oscillation

decreases as the atomic number Z increases. In addition, the heights of the peaks tend to increase with decreasing atomic number, consistent with the observed strong peaks at O-atom positions in $|1 - \psi(\mathbf{r})|$ over a broad range of thicknesses.

The emphasis of light-atom positions in $|1 - \psi(\mathbf{r})|$ results from a combination of two effects, one of which emerges from the channeling formalism and the second being the resolution dependence of $|1 - \psi(\mathbf{r})|$. In the channeling approximation for an atomic column for which a single bound state is dominant, $|1 - \psi(\mathbf{r})|$ has the form

$$|1 - \psi(\mathbf{r})| = C_b \Phi_b(\mathbf{R})(2 - 2\cos 2\pi K'z)^{1/2}, \qquad (10)$$

as can be obtained directly from (6) using $K' = E_b k / 2E_0$. The limiting form of (10) as z approaches zero is

$$|1 - \psi(\mathbf{r})|_{z \to 0} = C_b \Phi_b(\mathbf{R}) 2\pi K'z. \qquad (11)$$

One can obtain an approximate form for $\Phi_b(\mathbf{R})$ by comparing the slope of $|1 - \psi(\mathbf{r})|$ as z approaches zero with the expression

$$|1 - \psi(\mathbf{r})| = \pi[\rho(\mathbf{R})/E_0]kz$$

obtained from the weak-phase-object approximation. This leads to

$$C_b \Phi_b(\mathbf{R}) = \rho(\mathbf{R})/E_b. \qquad (12)$$

The form of the 1*s* wave thus mimics that of the projected atomic potential. Using (12) to replace $C_b \Phi_b(\mathbf{R})$ in (10), one arrives at an oscillating $|1 - \psi(\mathbf{r})|$ with period $2E_0\lambda/E_b$, which peaks with depth at an amplitude of $2\rho(\mathbf{R})/E_b$. The variation of the amplitude with Z is a second-order effect, as E_b tends to scale with the depth of the potential $\rho(\mathbf{R})$. Nevertheless, within the channeling approximation, the amplitude does decrease
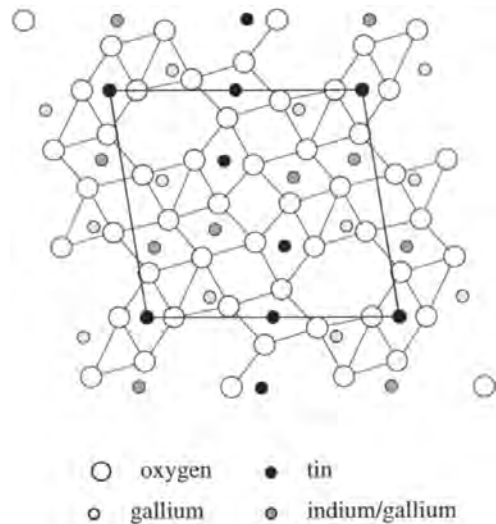

Fig. 5. View of the $(Ga, In)_2SnO_5$ structure in [010] projection.

somewhat with increasing Z. This is shown in Table 2, which compares the predicted amplitude at $\mathbf{R} = 0$ (center of the atomic column) with that obtained from multislice calculations at both 6 and 1 Å^{-1} resolutions, where resolution is defined here and in the following as

$2 \sin \theta_{max}/\lambda$. While the channeling theory significantly overestimates the amplitude due to its neglect of excitation errors, the trend as well as the relative amplitudes for different Z are in good agreement with the multislice values for 6 Å^{-1} resolution. From Table 2, one can see
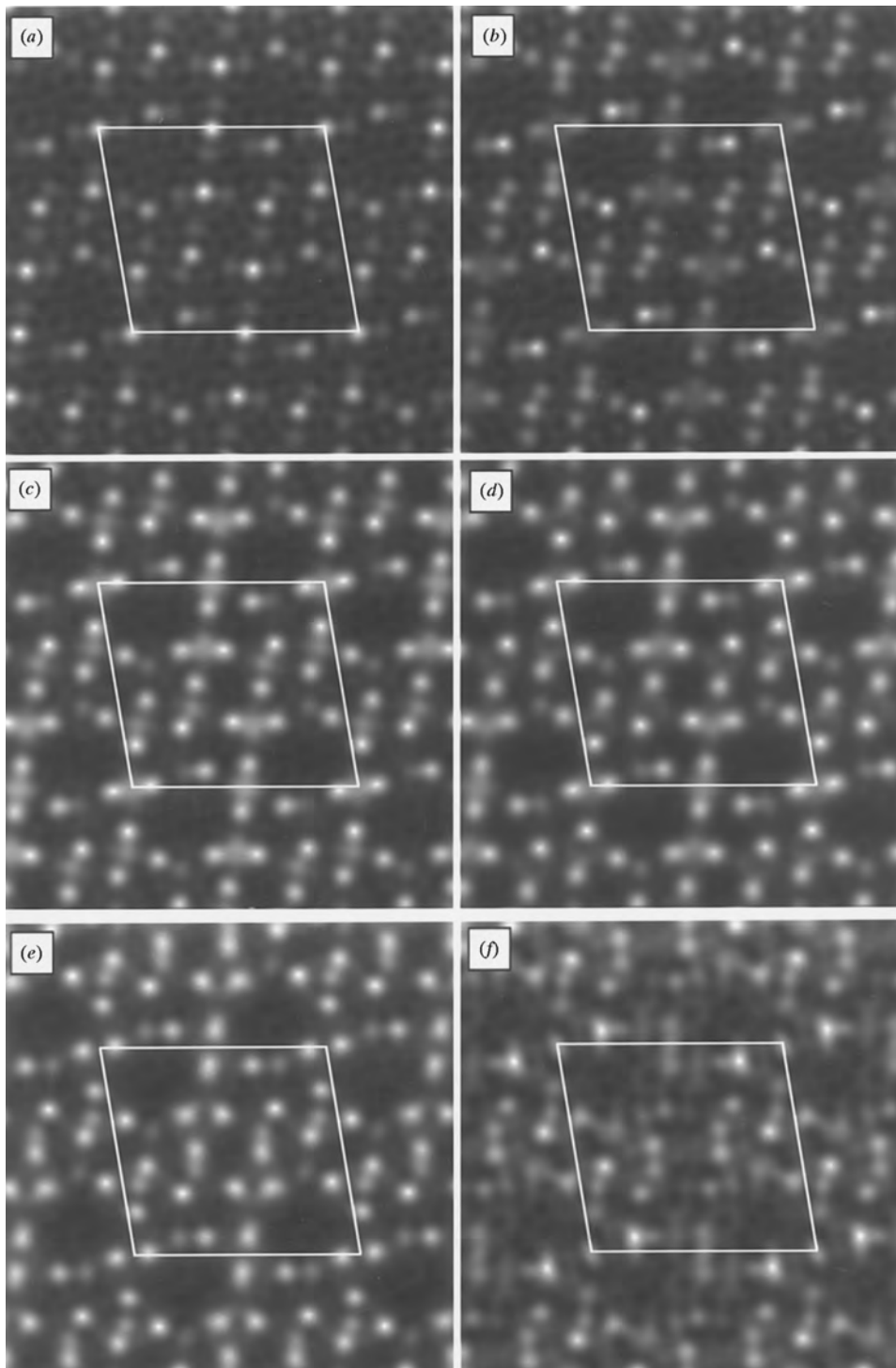

Fig. 6. Exit waves $|\psi(\mathbf{r}) - 1|$ calculated using multislice for the $(Ga, In)_2 SnO_5$ structure. (a) 8 slices. (b) 24 slices. (c) 48 slices. (d) 72 slices. (e) 128 slices. (f) 224 slices. The slice thickness is 1.58 Å.

Table 2. *Amplitude of the first maximum with depth in*
$|\psi(r) - 1|$ at the center of an atomic column

Atom	$2\rho(0)/E_1$	Multislice amplitude, 6 Å^{-1} resolution	Multislice amplitude, 1 Å^{-1} resolution
O	8.77	4.97	4.46169
Al	6.74	4.85	3.25662
K	5.62	4.49	2.64991
Ti	5.35	4.29	2.44803
Fe	5.57	4.13	2.26356
Ga	5.77	3.93	2.09908
Rb	5.32	3.66	1.96289
Mo	5.09	3.46	1.87577
Sn	5.13	3.33	1.79342
Eu	4.74	2.95	1.63094
Au	4.88	2.61	1.49058

that the Z dependence is further enhanced by reducing the resolution. The damping effect of limiting the resolution to 1 Å^{-1} is much more pronounced for heavy atoms, for which a larger proportion of the scattering is to large angles. This causes an additional reduction in the amplitude at the positions of heavy cations, relative to that at oxygen.
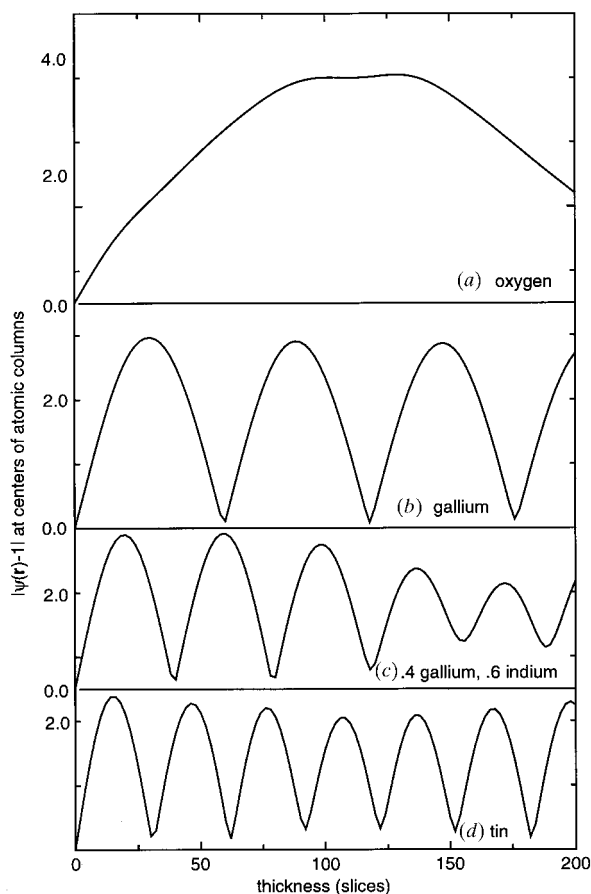
Fig. 7. Amplitude of $|\psi(\mathbf{r}) - 1|$ as a function of depth at the centers of several atomic columns in the (Ga, In)$_2$SnO$_5$ structure.

Table 3. *Experimental electron diffraction amplitudes for*
(Ga, In)$_2$SnO$_5$ [010] zone axis (arbitrary scaling)

h,k,l	$|\Psi(h,k,l)|$	h,k,l	$|\Psi(h,k,l)|$
4,0,2	1.0000	−1,0,3	0.1347
−5,0,3	0.8399	−5,0,2	0.1323
−1,0,5	0.7977	2,0,2	0.1316
0,0,3	0.7058	3,0,3	0.1223
−2,0,1	0.6684	−3,0,9	0.1206
−5,0,0	0.5020	−7,0,4	0.1191
−2,0,4	0.4525	−5,0,6	0.1177
−4,0,1	0.3977	−3,0,5	0.1175
−6,0,2	0.3881	1,0,4	0.1158
3,0,4	0.3316	−7,0,0	0.1147
−1,0,1	0.3220	−5,0,4	0.1127
−2,0,0	0.3122	−2,0,5	0.1094
−4,0,4	0.3070	7,0,6	0.1093
1,0,1	0.2381	8,0,4	0.1063
−3,0,3	0.2378	−4,0,5	0.1055
1,0,5	0.2262	7,0,3	0.1044
2,0,3	0.2221	2,0,4	0.1026
−3,0,2	0.2194	6,0,5	0.1005
3,0,1	0.2192	5,0,1	0.1003
−4,0,2	0.2183	−6,0,4	0.1001
−3,0,1	0.2182	−1,0,4	0.0980
−1,0,2	0.2150	−4,0,8	0.0977
−6,0,5	0.2065	2,0,5	0.0964
−7,0,7	0.2020	7,0,2	0.0953
−3,0,0	0.1989	−6,0,0	0.0942
−2,0,3	0.1964	0,0,6	0.0942
−2,0,7	0.1897	−8,0,5	0.0936
2,0,6	0.1847	−6,0,3	0.0910
−6,0,8	0.1828	−7,0,3	0.0879
6,0,2	0.1730	4,0,3	0.0868
4,0,5	0.1704	−6,0,1	0.0829
8,0,1	0.1703	4,0,6	0.0787
6,0,1	0.1606	−7,0,1	0.0784
1,0,2	0.1491	0,0,5	0.0749
1,0,3	0.1488	0,0,2	0.0748
3,0,7	0.1445	5,0,6	0.0743
−3,0,4	0.1414	1,0,8	0.0721
−3,0,6	0.1402	−4,0,0	0.0716
4,0,1	0.1401	2,0,1	0.0687
−8,0,2	0.1373	7,0,1	0.0675
0,0,4	0.1364		

4. Application of direct methods to calculated and experimental dynamical amplitudes

Direct methods were performed using both experimental electron diffraction amplitudes from the (Ga, In)$_2$SnO$_5$ structure (Table 3) as well as diffraction amplitudes for the same structure calculated using the multislice method. The direct methods employed a minimum-relative-entropy functional (Marks & Landree, 1998), but similar results may be expected with other direct-methods algorithms. The basis set consisted of phases fixed based on the HRTEM image (see below), plus an additional set of 16 fixed phases that were optimized using a genetic algorithm (Holland, 1975; Landree *et al.*, 1997). This results in multiple solutions because each of the possible 2^{16} sets of initial phases yields a different solution. The direct methods were in

Table 4. *Comparison of phases obtained from HRTEM with calculated phases of Fourier components of $|1 - \psi(r)|$ for a number of thicknesses (in slices)*

$(h0l)$ beams of $(Ga, In)_2SnO_5$ $(+ = 0°; - = 180°)$.

h, l	HR	2	8	16	24	32	40	48	56	64	72	96	128	160	192	224	256	288	320	352	384
4,−1	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
0,3	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
2,−1	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
4,2	−	+	+	+	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	−
2,0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+
5,0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	+	+	+	+
3,−1	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
1,3	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
1,−2	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
3,−3	−	−	−	−	−	−	−	−	+	−	−	−	+	+	+	+	−	−	−	+	+
1,−1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
1,−3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3,0	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	−
1,1	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−	−
4,−2	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
0,4	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
2,−3	+	+	+	+	+	+	−	−	+	−	−	−	−	−	−	−	+	+	+	−	−
3,−2	+	+	+	+	+	+	−	−	−	−	−	+	+	+	+	+	+	+	+	−	−

their main features identical to those employed with kinematical diffraction intensities and minimize the relative entropy by expanding (9) as the series (with iteration n)

$$|1 - \psi(\mathbf{r})|_{n+1} = |1 - \psi(\mathbf{r})|_n \ln\{|1 - \psi(\mathbf{r})|_n / \langle|1 - \psi(\mathbf{r})|_n\rangle\}. \quad (13)$$

The inverse FT of the basis set thus provided the initial estimate $|\psi(\mathbf{r}) - 1|_0$, and new phase estimates corresponding to $|\psi(\mathbf{r})-1|_1$ were obtained by FT following application of (13). This was then iterated in order to assign phases to all non-basis reflections. While a more rigorous justification of (13) would be based upon the arguments mentioned earlier in connection with (8) and (9), a simpler interpretation is to consider this as a sharpening function that finds real-space distributions that contain peaked or atom-like features and are consistent with the diffraction data.

In the normalizations of the diffracted amplitudes $|\Psi(\mathbf{h})|$ to obtain unitary and normalized structure factors $[|U(\mathbf{h})|$'s and $|E(\mathbf{h})|$'s], the kinematical atomic scattering factors were used for lack of a more accurate dynamical analog. (In general, while accurate normalization is desirable it is not critical.) For the phase extension *via* (13), the raw data were normalized to obtain windowed unitary structure factors as described elsewhere (Marks & Landree, 1998). Nevertheless, in order to counteract the uncertainty in normalization, the raw dynamical structure factors (plus phases) rather than the $U(\mathbf{h})$ were used in calculating the figure of merit (FOM).

$$\text{FOM} = \sum_{\mathbf{h}}{}' |\Psi_n(\mathbf{h}) - \beta\Psi_{n+1}(\mathbf{h})| \Big/ \sum_{\mathbf{h}}{}' |\Psi_n(\mathbf{h})|. \quad (14)$$

In (14), the prime on the sums indicates exclusion of $\mathbf{h} = 0$ and β is a constant that is chosen to minimize the FOM. The FOM is calculated at each iteration of (13) and is a test of the self-consistency of the solution. The greater the departure of the diffraction set $\Psi_{n+1}(\mathbf{h})$ following application of (13) from a rescaled version of the prior $\Psi_n(\mathbf{h})$, the larger the FOM will be. Iteration of (13) is continued as long as the FOM decreases. The FOM also serves as a parameter to optimize the basis set using the genetic algorithm.

A final measure taken to assist in convergence to solutions reflecting peaks of $|1 - \psi(\mathbf{r})|$ was the fixing of some phases from a HRTEM image. (Direct methods in the absence of additional phase information from HRTEM were found to be considerably less reliable for correctly indicating atomic positions.) The phases fixed were taken from an experimental image of $(Ga, In)_2SnO_5$ along the monoclinic b axis, which was used in solving the structure (Sinkler *et al.*, 1998). Image phases were obtained, after lattice averaging, from an image taken at Scherzer defocus. An origin refinement was used to minimize the deviation from 0 and 180° consistent with centrosymmetry. The phases fixed were chosen to satisfy:

(i) phase deviating less than 25° from 0 or 180°;
(ii) spatial frequency less than 0.45 Å^{-1};
(iii) amplitude greater than 5% that of the strongest beam in the image Fourier transform.

This allowed fixing of 17 of the strongest beams in the diffraction data set. One additional beam, the (402), was fixed based on a strong Σ_1 relationship. Table 4 shows the beams whose phases were fixed, as well as the evolution in the phases of these beams in the calculated $|1 - \psi(\mathbf{r})|$ with thickness; in agreement with Fig. 4, the

phases of these beams are reasonably stationary with depth.

Direct methods were employed on simulated data sets including 193 beams to a 1.0 Å resolution. The top 20 solutions based on the FOM optimization were stored for inspection. Fig. 8 shows solutions from calculated dynamical diffraction intensities for $(Ga, In)_2SnO_5$ corresponding to the same thicknesses for which $|1 - \psi(\mathbf{r})|$ is shown in Fig. 5. In plotting the solutions, the $|\Psi(\mathbf{h})|$ from the multislice calculation is provided with
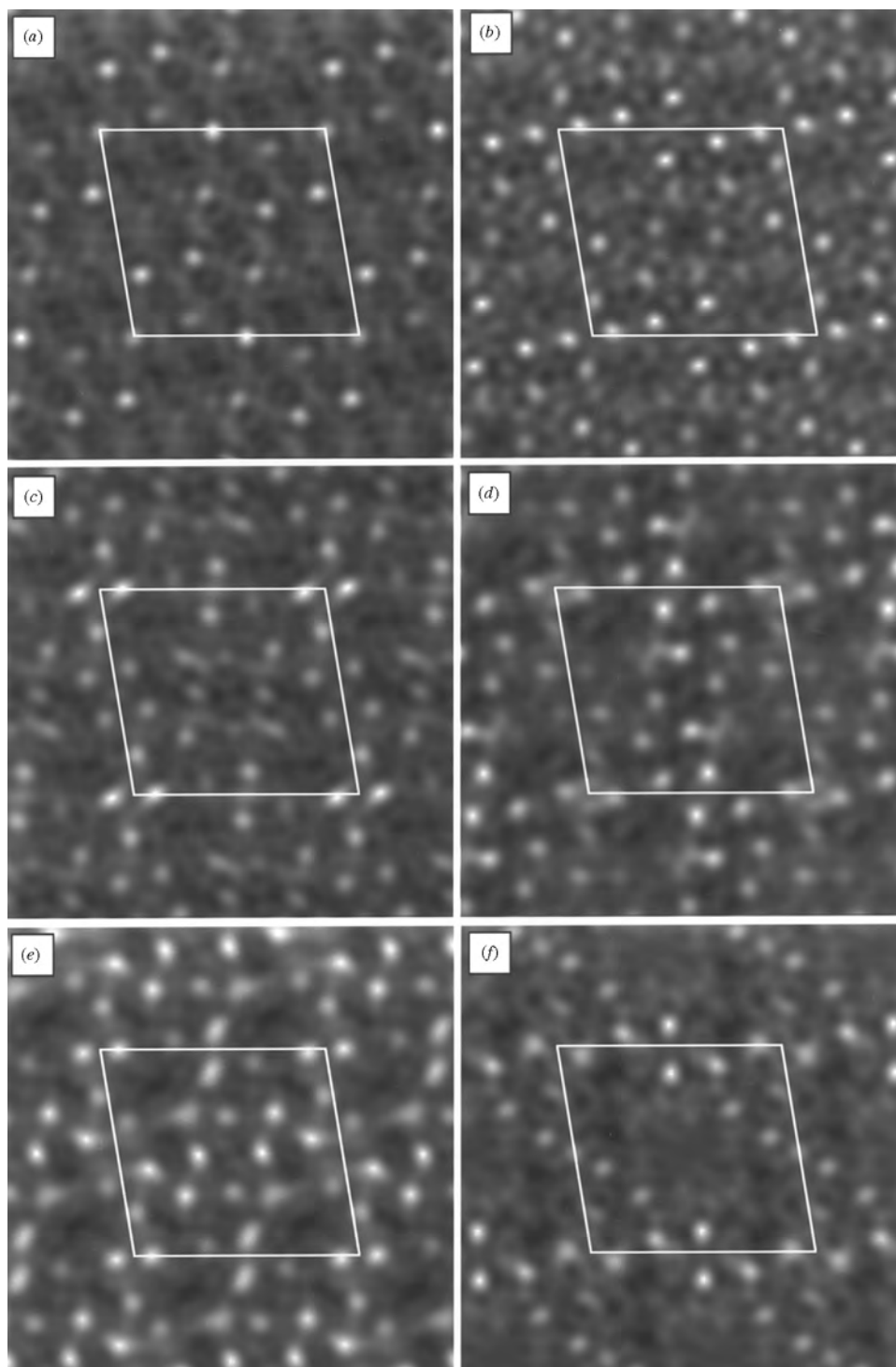

Fig. 8. Phasing solutions obtained using calculated diffraction intensities for thicknesses corresponding to Fig. 6. The rankings of the solutions by FOM is given in Table 5.

Table 5. *Analysis of phasing solutions for (Ga, In)$_2$SnO$_5$*

Thickness (slices)	Thickness (Å)	Rank of solution	Peaks with height >40% of max.	Peaks at cations	Peaks at O atoms	Spurious peaks
8	12.6	2	12	12	0	0
24	37.9	1	20	2	18	0
48	75.8	1	24	3	20	1
72	113.8	1	22	2	18	2
128	202.2	4	20	0	20	0
160	252.8	1	18	0	18	0
192	303.4	2	18	0	16	2
224	353.9	16	17	0	10	7

phases from direct methods and reverse Fourier transformed. The choice of the solutions from among the top 20 is unavoidably subjective, as the structure is now known [it was not when this work was first performed (Sinkler *et al.*, 1998)]. Nevertheless, it was found that, over a broad range between 24 and 192 layers (38–300 Å) thickness, a solution clearly showing at least 16 of the 20 O-atom positions was among the top four solutions. This is shown in Table 5, which presents an analysis of the solutions in Fig. 8, as well as of several other solutions for additional thicknesses. Table 5 lists the number of peaks in the solution whose heights are within 40% of the maximum, as well as the number of these that were within 0.6 Å of either a cation or an O atom in the structure. To counter the subjectivity in selecting the solutions, it should be noted that those that do not accurately reflect the structure can often be picked out simply by a more streaked appearance. Additionally, in many cases in which some of the O atoms were not located near a strong enough peak in the solution to appear in Table 5, they were indicated by weaker peaks.

Based on comparison of Fig. 6 and Fig. 8, the appearance of $|1 - \psi(\mathbf{r})|$ is an excellent guide to the structural features that phasing may be expected to find.
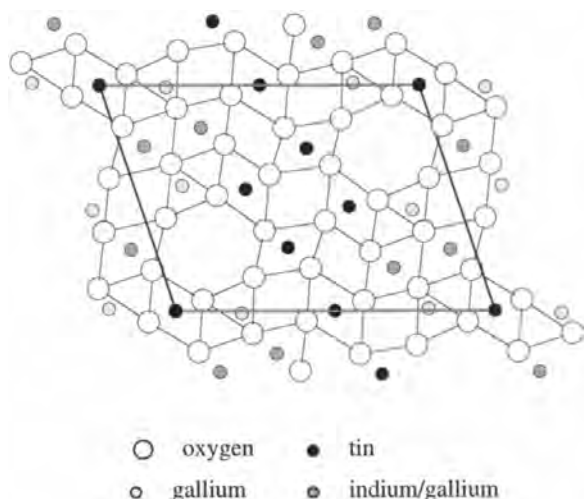
Fig. 9. View of (Ga, In)$_4$Sn$_3$O$_{12}$ structure in [010] projection.

For a thickness of 12.6 Å, the phasing solution correctly locates cations (all but two of the Ga atoms, which are nevertheless present as weaker peaks in the phasing map). This is consistent with the dominance of the cations in $|1 - \psi(\mathbf{r})|$ at this thickness. For thicknesses between 76 and 253 Å (48 to 160 slices), peaks in the solutions are consistently found at 18 or more of the 20 O-atom positions, reflecting the dominance of peaks at O-atom positions in $|1 - \psi(\mathbf{r})|$ throughout this range. Finally, $|1 - \psi(\mathbf{r})|$ begins to show strong features not related to the structure at thicknesses greater than 300 Å; presumably the unbound states can no longer be ignored. This again is clearly reflected in the phasing solutions for thicknesses beyond 300 Å in a pronounced drop in the number and ranking of solutions that have peaks in agreement with atom positions.

The availability of phasing maps such as shown in Fig. 8 is clearly of great value towards solving a crystal structure. An example of a phasing solution obtained using experimental electron diffraction intensities from (Ga, In)$_2$SnO$_5$ has already been shown as Fig. 1 (Sinkler *et al.*, 1998). While the solution was ranked seventh, it was easily selected on inspection based on the occurrence of sharp round peaks. The solution clearly shows all 20 O atoms of the (Ga, In)$_2$SnO$_5$ structure. Subsequent Rietveld refinement of neutron diffraction data confirmed to within 0.27 Å the accuracy of the O-atom positions taken from the phasing map in Fig. 1 (Sinkler *et al.*, 1998; Edwards *et al.*, 1998). Fig. 10 shows two phasing solutions from experimental data on the (Ga, In)$_4$Sn$_3$O$_{12}$ structure (Edwards *et al.*, 1998), which is shown in a ball-and-stick model in Fig. 9. In that case, no single solution showed all the O-atom positions, perhaps because diffraction data were only collected to 0.77 Å^{-1}. Fig. 10 shows the 3rd- and 13th-ranked solutions. Fig. 10(*a*) has 16 peaks and two additional peaks that are weaker than 40% of maximum, for a total of 18 (crosses in Fig. 10*a*). All agree well with 18 of the 24 O atoms in Fig. 9. Fig. 10(*b*) has most peaks in agreement with Fig. 10(*a*) but has four additional peaks at the positions marked with crosses. Combining these two solutions thus provides approximate positions for 22 of the 24 O atoms.

The ability of direct methods to locate the O atoms using experimental electron diffraction demonstrates

○ oxygen   ● tin
○ gallium   ◉ indium/gallium

that the technique does not depend on ideal modeled intensities but is also tolerant of errors associated with collection of the data. Such errors arise not only due to limitations of the collection medium [photographic film – for details on data collection see Xu *et al.* (1994)] but also from imperfect crystal tilt and from some averaging over thickness that is unavoidable even using very small spot sizes. With regard to averaging over thickness, it is possible that this actually enhanced the sensitivity of the technique for O atoms in the experimental cases. This might be expected because the frequencies of oscillations in $|1 - \psi(\mathbf{r})|$ at cation positions are faster than at O-atom positions, as shown in §3, Fig. 7. Averaging over thickness would thus tend to decrease the intensity of beams reflecting the cation distribution relative to those emphasizing O atoms. Preliminary attempts at modeling this averaging to investigate its effect on phasing did not show any effect on the success rate of the phasing at finding O-atom positions, but this lack of a clear effect may be due to the rather high success rate without averaging.

5. Discussion

The present work has indicated that a combination of HRTEM and direct methods of strongly dynamical electron diffraction intensities can constitute a powerful means for detecting light atoms in bulk inorganic materials. The physical basis for the detection of light atoms has been investigated in this work and a satisfactory explanation has been found using electron channeling theory. Channeling effects are responsible for the localization of peaks in the function $|1 - \psi(\mathbf{r})|$ and the emphasis of light-atom columns in $|1 - \psi(\mathbf{r})|$ can

be attributed to a combination of dynamical channeling and resolution-dependent effects. Two factors are responsible for the ability of direct methods to restore these peaks. The first is the rough scaling between the diffraction amplitudes, the $|\Psi(\mathbf{h})|$, and the Fourier components of $|1 - \psi(\mathbf{r})|$, as shown in §2. Secondly, $|1 - \psi(\mathbf{r})|$ is a peaked function containing atom-like features and is thus consistent with the direct-methods approach. Because of the relationship between $|1 - \psi(\mathbf{r})|$ and direct-methods solutions using dynamical intensities, the appearance of $|1 - \psi(\mathbf{r})|$ may be used with confidence as a basis for estimating whether direct methods can provide useful structural information in a given case.

From channeling theory, the localization of intensity in $|1 - \psi(\mathbf{r})|$ near to the projected positions of atomic columns depends on the degree to which the structure resolves into well separated atomic columns so it is clear that the technique in its present form will work best for such cases. This nevertheless includes a large number of important ceramic materials, including for instance superconductors, for which modeling has shown similar success rates as for the Ga–In–Sn–O structures investigated here. The ability of direct methods to locate light atoms is all the more valuable as these are the most difficult to locate by most other readily accessible techniques, in particular X-ray diffraction.

Most of the salient points have already been covered regarding the features of direct methods using strongly dynamical electron diffraction. A brief comment is due concerning the role of the present technique in a structure determination. It should be cautioned that at present the direct-methods results alone do not provide a basis to judge whether a model structure is correct or
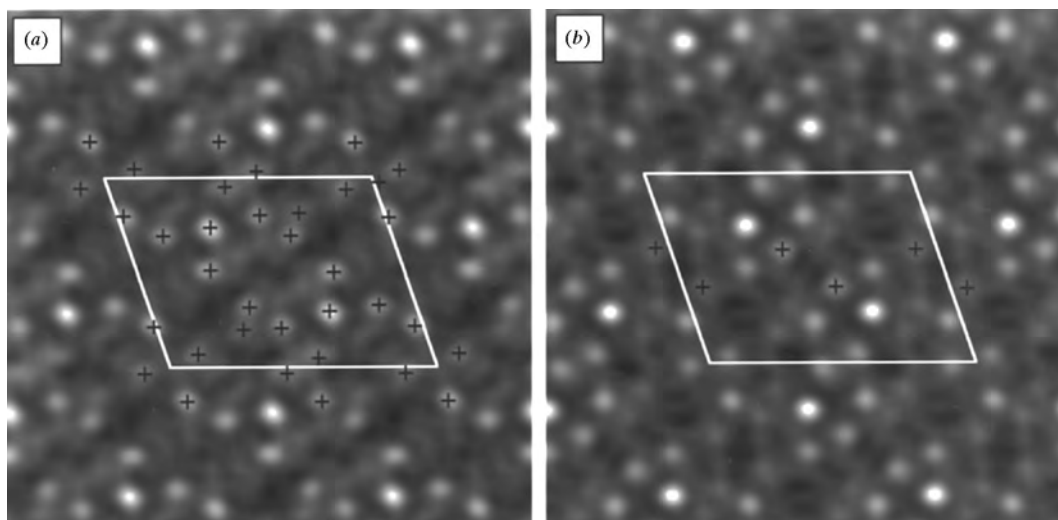

Fig. 10. Phasing solutions from experimental data for $(Ga, In)_4Sn_3O_{12}$. The peaks in the phasing solutions indicate the O-atom positions in the final structure shown in Fig. 9. (*a*) 3rd-ranked solution showing peaks at 18 O-atom positions (crosses). (*b*) 13th-ranked solution showing an additional four O-atom positions (crosses). Combining (*a*) and (*b*), 22 of a total of 24 O-atom positions are indicated.

not. Corroboration of any structure model derived by the present means must be sought using another technique. In cases in which the phase of interest is only obtainable as a minority phase, confirmation of a structure obtained with the aid of the present technique can be sought *via* multislice simulation of HRTEM images or quantitative HRTEM (Zhang *et al.*, 1995). Refinement of electron diffraction intensities (Zandbergen *et al.*, 1994; Weirich *et al.*, 1996) would also be a convenient approach as the same data employed for phasing could be used. However, as strongly indicated in this work, any such refinement must take dynamical effects into account. The strength of direct methods is that they provide precisely what most techniques available for quantitative refinement lack, namely the ability to suggest initial atom positions for light elements.

A final note is due on the necessity of setting $F(-\mathbf{h}) = F^*(\mathbf{h})$ and thus ignoring the complex nature of the electron exit wave. While the present work has applied direct methods in the context of a solution that is real-valued, this is in fact not necessary, and the approaches exemplified by (8) and (9) do not depend on it. Allowing complex solutions is equivalent to phasing double the number of spots, *i.e.* $\mathbf{h}$ and $-\mathbf{h}$ independently. This can be approached in a completely analogous way to normal direct methods and may in fact require relatively minor modification of code. It may be possible to devise algorithms that are specialized for phase extension in the complex case, permitting restoration of the complex exit wave $\psi(\mathbf{r})$ or the Babinet $1 - \psi(\mathbf{r})$. Preliminary attempts at this have been made by the present authors and are the subject of continuing research. To extend further to thicker crystals would require a reformulation to take account specifically of the unbound states, a topic for further work.

APPENDIX A
Minimum relative entropy/Kullback–Leibler distance

While 'maximum entropy' has been used in a crystallographic environment, we are not aware of the use of the relative entropy or Kullback–Leibler (1951) distance [also known as cross entropy, information divergence and information for discrimination, see Cover & Thomas (1991) for a much more detailed analysis] so a little clarification is appropriate. Considering the standard definition of entropy or self-information (in real space $\mathbf{r}$) as:

$$S = -\sum \rho(\mathbf{r}) \ln \rho(\mathbf{r}) \quad \text{with} \quad \sum \rho(\mathbf{r}) = 1 \quad (15)$$

$$S = -\sum \{\rho(\mathbf{r}) \ln[\rho(\mathbf{r})/e\langle\rho(\mathbf{r})\rangle] + \langle\rho(\mathbf{r})\rangle\}$$
$$- \sum \rho(\mathbf{r}) \ln[e\langle\rho(\mathbf{r})\rangle] + 1 \quad (16)$$

$$= -\sum \{\rho(\mathbf{r}) \ln[\rho(\mathbf{r})/e\langle\rho(\mathbf{r})\rangle] + \langle\rho(\mathbf{r})\rangle\}$$
$$- \ln[\langle\rho(\mathbf{r})\rangle]. \quad (17)$$

The relative entropy is the negative of the first term in (17) and maximizing the conventional entropy is thus equivalent to minimizing the relative entropy. More usefully, the relative entropy as defined above is a metric of the deviation of the distribution $\rho(\mathbf{r})$ from flat and can be used as such. The FOM used herein is the projection of the relative entropy onto the set of measured reflections in reciprocal space.

To strengthen the algorithm further, we modify the unitary structure factors in reciprocal space by a 'window function' $W(\mathbf{h})$ using $U'(\mathbf{h})$, where

$$U'(\mathbf{h}) = W(\mathbf{h})U(\mathbf{h}). \quad (18)$$

The window function (in real space) is defined to be an eigensolution of the 'relative entropy sharpening operator' $\hat{\mathbf{O}}$ *via* the equation

$$A(\mathbf{r}) * \hat{\mathbf{O}}w(\mathbf{r}) = A(\mathbf{r}) * \{w(\mathbf{r}) \ln[w(\mathbf{r})/\langle w(\mathbf{r})\rangle]\}$$
$$= \lambda A(\mathbf{r}) * w(\mathbf{r}), \quad (19)$$

in which $*$ is convolution and $A(\mathbf{r})$ is the Fourier transform of an aperture function surrounding the set of measured reflections. For a set of identical non-overlapping atoms, with kinematical diffraction, the relative entropy defined in this way is identically zero for the true solution – we have built in a pattern-recognition component. As such, the FOM is the log-likelihood (Cover & Thomas, 1991, p. 18) of the error in the hypothesis that a given solution corresponds to a set of atoms.

There are very strong similarities to the Gerchberg–Saxton (1972) and Fienup (1978, 1987) algorithms, as well as the Sayre equation (Sayre, 1952), with the relative entropy sharpening operator applying a self-consistent (for atoms) constraint on the solutions with a self-consistency FOM. The construction

$$u_{n+1}(\mathbf{r}) = u_n(\mathbf{r}) \ln[u_n(\mathbf{r})/\langle u_n(\mathbf{r})\rangle] \quad (20)$$

is a relatively primitive first-order iterative solution for zeros of the relative entropy used in a tangent-formula update scheme; since we only want approximate solutions, a primitive search is appropriate. (The radius of convergence of the iterative scheme is finite so a search over initial starting phases is required.)

The method has strong similarities to crystallographic maximum-entropy approaches, for instance strong nonlinearities with the ability to interpolate unmeasured reflections. Rather than a precise numerical optimization of the (absolute) entropy and weak constraints on the moduli of the structure factors, a loose optimization is used *via* the iteration (20), with tight constraints on the structure factors. The avoidance of any gradient search for entropy maximization both simplifies and increases the numerical speed.

References

Berry, M. V. (1971). *J. Phys. C*, **4**, 697–722.

Collazo-Davila, C., Marks, L. D., Nishii, K. & Tanishiro, Y. (1997). *Surf. Rev. Lett.* **4**, 65–70.

Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*. New York: John Wiley.

Dong, W., Baird, T., Fryer, J. R., Gilmore, C. J., MacNicol, D. D., Bricogne, G., Smith, D. J., O'Keefe, M. A. & Hovmöller, S. (1992). *Nature (London)*, **355**, 605–609.

Dorset, D. L. (1996). *Acta Cryst.* A**52**, 753–796.

Doyle, P. A. & Turner, P. S. (1968). *Acta Cryst.* A**24**, 390–397.

Edwards, D. D., Mason, T. O., Sinkler, W., Marks, L. D., Poeppelmeier, K. R., Hu, Z. & Jorgensen, J. D. (1998). In preparation.

Fan, H.-F., Xiang, S. B., Li, F. H., Pan, Q., Uyeda, N. & Fujiyoshi, Y. (1991). *Ultramicroscopy*, **36**, 361–365.

Fienup, J. R. (1978). *Optics Lett.* **3**, 27–29.

Fienup, J. R. (1987). *J. Opt. Soc. Am.* **4**, 118–123.

Fu, Z. Q., Huang, D. X., Li, F. H., Zhao, Z. X., Cheng, T. Z. & Fan, H. F. (1994). *Ultramicroscopy*, **54**, 229–236.

Fujimoto, F. (1978). *Phys. Status Solidi*, **45**, 99–106.

Gemmell, D. S. (1974). *Rev. Mod. Phys.* **46**, 129–227.

Gerchberg, R. W. & Saxton, W. O. (1972). *Optik (Stuttgart)*, **35**, 237–246.

Gilmore, C. J., Marks, L. D., Grozea, D., Collazo-Davila, C., Landree, E. & Twesten, R. D. (1997). *Surf. Sci.* **381**, 77–91.

Gilmore, C. J., Shankland, K. & Bricogne, G. (1993). *Proc. R. Soc. London Ser. A*, **442**, 97–111.

Hirsch, P. B., Howie, A., Nicholson, R. B., Pashley, D. W. & Whelan, M. J. (1977). *Electron Microscopy of Thin Crystals*. Malabar, Florida: Robert E. Krieger.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Hu, J. J., Li, F. H. & Fan, H. F. (1992). *Ultramicroscopy*, **41**, 387–397.

Kambe, K. (1982). *Ultramicroscopy*, **10**, 223–228.

Kambe, K., Lehmpfuhl, G. & Fujimoto, F. (1974). *Z. Naturforsch. Teil A*, **29**, 1034–1044.

Kullback, S. & Leibler, R. A. (1951). *Ann. Math. Stat.* **22**, 79–86.

Landree, E., Collazo-Davila, C. & Marks, L. D. (1997). *Acta Cryst.* B**53**, 916–922.

Landree, E., Marks, L. D., Zschack, P. & Gilmore, C. (1998). *Surf. Sci.* In the press.

Marks, L. D. (1985). *Ultramicroscopy*, **18**, 33–38.

Marks, L. D. (1988). *Ultramicroscopy*, **25**, 85–88.

Marks, L. D. & Landree, E. (1998). *Acta Cryst.* A**54**, 296–305.

Marks, L. D., Plass, R. & Dorset, D. L. (1997). *Surf. Rev. Lett.* **4**, 1–8.

Marks, L. D., Savage, T. S., Zhang, J. P. & Ai, R. (1991). *Ultramicroscopy*, **38**, 343–347.

Ohtsuki, Y. H. (1983). *Charged Beam Interaction with Solids*. New York: Taylor and Francis.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Sinkler, W., Marks, L. D., Edwards, D. D., Mason, T. O., Poeppelmeier, K. R., Hu, Z. & Jorgensen, J. D. (1998). *J. Solid State Chem.* **136**, 145–149.

Spence, J. C. H. (1988). *Experimental High-Resolution Electron Microscopy*. Oxford University Press.

Tamura, A. & Ohtsuki, Y. H. (1974). *Phys. Status Solidi*, **62**, 477–480.

Tanishiro, Y. & Takayanagi, K. (1989). *Ultramicroscopy*, **27**, 1–8.

Van Dyck, D. (1985). *Advances in Electronics and Electron Physics*, No. 65, edited by P. Hawkes, pp. 296–357. New York: Academic Press.

Van Dyck, D. & Op de Beeck, M. (1996). *Ultramicroscopy*, **64**, 99–107.

Weirich, T. E., Ramlau, R., Simon, A., Hovmöller, S. & Zou, X. (1996). *Nature (London)*, **382**, 144–146.

Xu, P., Jayaram, G. & Marks, L. D. (1994). *Ultramicroscopy*, **53**, 15–18.

Zandbergen, H. W., Jansen, J., Cava, R. J., Krajewski, J. J. & Peck, W. F. (1994). *Nature (London)*, **372**, 759–761.

Zhang, H., Marks, L. D., Wang, Y. Y., Zhang, H., Dravid, V. P., Han, P. & Payne, D. A. (1995). *Ultramicroscopy*, **57**, 103–111.

Ziman, J. M. (1972). *Principles of the Theory of Solid State Physics*. Cambridge University Press.