Domino phase retrieval algorithm for structure determination using electron diffraction and high resolution transmission electron microscopy patterns

F.N. Chukhovskii¹, A.M. Poliakov², I.V. Prokopenko³

¹ Institute of Crystallography, Russian Academy of Sciences, 59 Leninsky prospect, 117333 Moscow, Russia
² Department of Materials Science, Moscow Institute of Steel and Alloys, 4 Leninsky prospect, 117279 Moscow, Russia
³ Institute of Semiconductor Physics, NAS of Ukraine, 45 prospect Nauky, 03028 Kyiv, Ukraine

Abstract. Direct method formalism to determine atomic structures using the electron diffraction data is here aimed at a general solution of the phase retrieval problem, consequently combining the electron diffraction (ED) and the high-resolution transmission electron microscopy (HRTEM) patterns in a “domino” fashion. While there are similarities to what there is in conventional (kinematical) direct methods, there remain major differences, in particular, owing to the dynamical effects in the data the ED structure factors prove to be complex and then, the positivity of the reconstructed electron density is no longer a valid constraint for ‘dynamical’ direct methods. Besides, due to the dynamical effects heavy atoms need not dominantly contribute the HRTEM images any more. The clear sense is that the domino transform algorithm proposed works well and represents a valuable method for phasing diffraction patterns in electron structural crystallography using an experiment, which is readily to perform within collecting the ED and HRTEM data.

Keywords: phase retrieval problem, electron diffraction, high-resolution transmission electron microscopy.

Paper received 05.06.03; accepted for publication 17.06.03.

1. Introduction

An application of conventional (kinematical) direct methods and refinements for purposes of electron structure determination raises some fundamental problems. There are some simple cases, for instance a surface, where such refinements are legitimate as a good first approximation, but even here correct results require the inclusion of dynamical effects. It is very well established theoretically that even one atom layer of a heavy element such as gold is a ‘dynamical’ scattering unit (see, e.g., [1,2]), so there are no cases with real samples where the kinematical approximation is rigorously valid. While exceedingly thin values such as 2 nm are sometimes reported for sample thickness used in interpreting experimental high-resolution transmission electron microscopy (HRTEM) images, the actual thickness is certainly larger; the discrepancy is often due to neglecting terms such as sample vibration or beam tilt. At the same time, under some optimal imaging conditions the HRTEM technique provides real-space information toward crystal structures though HRTEM images in general do not contain true atom positions. On the other hand, to obtain true real-space information concerning atom positions one may explore the direct methods relying on the diffraction information alone and extracting the unknown phase information among the structure factors [3–5]. The most part of earlier combinations of HRTEM and direct methods have tested disregarding dynamical effects in electron diffraction (ED) data, namely: the direct methods have been applied within a kinematical approach [6–10].

Noteworthy is the fact that the first quantitative relationships being capable to be used for phasing diffraction patterns via direct methods were Sayre’s equation
and statistical phase invariants [11, 12]. Switching to a set of structure factors, \( \{ \mathcal{U}(g) \} \), the phase set \( \{ \mathbf{\phi}(g) \} \) with \( \mathbf{\psi}(g) = \text{Im}[\ln(\mathcal{U}(g))] \) and \( g \) a reciprocal lattice vector, could be determined using the iterative transform algorithm based on the Sayre–Tangent formula as a convolution equation (namely, \( \mathbf{\psi}(g) = \text{Im}[\sum h \mathcal{U}(h) \mathbf{\psi}(h+g)] \)) with the known set of moduli \( \{ \mathcal{U}(g) \} \); (see, e.g., [3, 5, 13] for details). For this, the Sayre–Tangent formula is relatively exact for pointlike atomic structures as long as the atom-scatter positions are not overlapped that coincides with the basic assumption of the ED channeling approximation [14-16].

At the same time, the phase information is preserved in the HRTEM images and can be extracted sometimes using the Fourier transform [17, 18]. Ishizuka et al. (1982) have developed the resolution improvement and/or phase correction method first proposed by Hoppe and Gassmann (1968) [19, 20]. They have combined in the electron micrograph and electron diffraction within the weak-scattering approximation and have applied the procedure to the model structure of the crystal of copper perchlorophthalocyanine for determining the phases (signs) of the structure factors.

The prospects for least-squares refinement [21] and direct method technique [22] to decode structural data from ED experiment was recently reviewed in many cases, in the most part of which dynamical ED effects have been ignored. The work [23] has to be drawn out, in which the direct methods combining with the Fourier transform of the HRTEM image were attempted to solving the phase problem on an example of \((\text{Ga,In}_2\text{SnO}_3\)) ceramic.

And at once, after that, the two- and three-phase structure invariants, \( \Sigma_0 \) and \( \Sigma_2 \), within the scope of ‘dynamical’ direct methods has been analyzed in the works [24, 25]. It was shown how the relevant success of applying direct methods to dynamical ED data can be understood via an “effective kinematical approximation” since each of both the phase conditional probability \( \Sigma_0 \) distribution defined for a number of \((g, -g)\)-reflection pairs and the phase conditional probability \( \Sigma_2 \) distribution defined for a number of \((g, h, -g-h)\)-reflection triplets are proved to display a strong peak in many cases. The above assertion is obtained using the theoretical probability background [12] and confirmed by numerical multislice calculations. Particularly noteworthy is the fact that the recovered effective dynamical potential may be similar to the kinematical one but does not have to be and in general will not be.

In this paper we push the concept of direct methods one step further, and apply the lattermost using both the two-dimensional ED and HRTEM data sets as constraints for a phase restoration. To attribute this concept the iterative transform algorithm, which has at its core the Gerchberg-Saxton algorithm ([26]; see [27] as well) and makes the phase retrieval procedure in a dominio fashion, is proposed. We believe that no such kind of iterative transform algorithm for direct phasing of the ED complex structure factors has been earlier reported. A prerequisite to solving the problem herein is to elaborate the appropriate numerical algorithm that creates a unique phase restoration. For purposes of this study, some numerical simulations are given on a sample of \((\text{Ga,In}_2\text{SnO}_3\)) ceramic, which exhibits a convergence property of the domino iterative transform algorithm proposed. In particular, the new algorithm, which is described here, has been proven to be convergent in the general case of complex structure factors. Noteworthy is the fact that its implication can facilitate a true determination of nanometer-size crystal structures combining both the ED and HRTEM data in a sequent domino manner.

## 2. Problem foundation. Domino iterative transform algorithm combining the ED and HRTEM data

The problem of phase retrieval is related to

- the complex structure factors set \( \{ \mathcal{U}(g) \} \)
- the HRTEM image \( I(x) \).

They are formed as a result of the plane wave propagation through an electron optical system consisting of selected-area apertures and focusing lens taking into account all likely aberrations (e.g., defocus, spherical aberration and so on).

The focused electron wave is imaged onto the two-dimensional (2D-) arrays of detectors that specifically measure intensity either in the back focal or in the image plane, respectively.

- Intensities measured in the back focal plane are referred to the ED magnitudes \( \{ \mathcal{U}(g) \} \), with diverse reciprocal lattice vectors \( g \).
- Intensity distribution \( I(x) \) in the image plane is the modulus squared of the Fourier transform of the complex structure factors set \( \{ \mathcal{U}(g) \} \). The 2D-periodical function \( I(x) \) is the HRTEM image of the crystal structure.

From the mathematical viewpoint the data array of \( \{ \mathcal{U}(g) \} \) and the data array \( \{ A(x) \} \) as well, with \( A(x) = \sqrt{I(x)} \) is non-convex. The nonconvexity of both the underlying sets, \( \{ \mathcal{U}(g) \} \), \( \{ A(x) \} \) is a main obstacle to apply the Gerchberg-Saxton-type algorithm for phasing diffraction patterns directly.

In order to overcome the fundamental nonconvexity limitation and as a result, to avoid numerous redundant solutions, let us introduce the oblique-angled selected-area constraint sets of \( \{ \mathcal{U}(g) \}, \{ A(x) \} \) for the consecutive values of \( v = 2, 3, 4, \ldots \), \( N(\mathcal{U}(g)_{v=0} = 0 \) is assumed to be unity for any \( v \). The principal idea of our phase retrieval method is to obtain the phase fit for the complete structure factors set \( \{ \mathcal{U}(g) \} \) sequentially step by step, starting from the first subsets \( \{ \mathcal{U}(g) \} \), \( \{ A(x) \} \) up to \( \{ \mathcal{U}(g) \}, \{ A(x) \} \), and using the standard Gerchberg-Saxton iteration procedure at each step. Considering the oblique-angled selected-area array rank \( n \) as a fixed iteration number, the numerical algorithm scheme takes the form

\[
(\nu)_{M_k+1}(x) = (\nu)_{A(x) \exp[i(\nu)_{\psi}(x)]}
\]

\[
(\nu)_{U_k+1}(g) = \text{InverseFourier}_{(\nu)_{M_k+1}(x)}
\]

398

SQO, 6(3), 2003
\((\psi_{k+1}(g) = \text{Im}[\ln(\text{InverseFourier}(\psi_M(x)))]\)

\((U_{k+1}(g) = \|U(g)|\exp(\psi_{k+1}(g))\)

\((\mathbf{A}_{k+1}(x)) = \text{Fourier}(\psi_{k+1}(g))\)

\((\varphi_{k+1}(x) = \text{Im}[\text{Fourier}(U_{k+1}(g))]).\)

(1)

where \((\psi_{k+1}(g)\) is the given phase of the complex structure factor \((U_{k+1}(g))\) at the beginning of each cycle \(k\), and \((\psi_{k+1}(g)\) is the new phases calculated at the end of the same cycle. After the phase set of \((\psi_{k+1}(g))\) is calculated, then it feed back into the cycle iteration fashion according to the ‘flow’ equation (1). The \((k+1)\)th phase sets of \((\psi_{k+1}(g), \varphi_{k+1}(x))\) is calculated by use of the preceding \(k\)-th ones \((\psi_{k}(g), \varphi_{k}(x))\), keeping in mind that the ED and HRTEM sets of \((U(g), A(x))\) are a priori fixed and do not depend on the cycle value of \(k\) over all the iteration processes of \(n\) (e.g., they are equal to the ‘experimental’ values).

Our main focus is on the fact that some important restrictions are imposed onto the phase set of \((\psi_{k+1}(g))\), specifically, the phase set of \((\psi_{k+1}(g))\) fixedly contains the subset of \((\psi_{k+1}(g))\), with the array size of \(n\times n\), determined within the preceding iteration of \(n\). And clearly, the other elements of the phase set of \((\psi_{k+1}(g))\) are chosen to be random in the range of \((-\pi, \pi)\) for the first cycle, \(k = 1\), only.

A ‘domino’ idea of the iterative transform algorithm is lying on the restriction of the total number of the likely solutions for the ED and HRTEM input data set of \((U(g), A(x))\), where \(n = 2\), that launch the general iteration procedure. It is easy to show that in the case of the input iteration value of \(n = 2\) there are eight diverse solutions for the phase set of \((\psi_{k+1}(g))\), more precisely, there are the diverse four pairs of the complex conjugated structure factors, one of which is referred to as a true solution (recall that \((\psi_{k+1}(g))_{g = 0} = 0\)). Going on from the iteration of \(n = 2\) until the final iteration \(N\) via the consecutive oblique-angled selected-area iterations the true phase solution holds on up to some numerical uncertainties depending on the final iteration number \(N\), whereas other redundant solutions can be readily detected yielding inappropriate values of FOM. Presumably, numerical uncertainties of the ‘correct’ phases increase with increasing the iteration number (the reflection array rank) \(n\) that might be compensated by enlarging the cycle range value of \(K\).

The two figures of merit (FOM) related to the reciprocal space, \(R^{(\psi)}\), and real space, \(X^{(\psi)}\), are evaluated at each cycle over \(k\)

\(R^{(\psi)}_{k+1} = \frac{1}{2} \sum_{g} \|\psi_{k+1}(g) - \psi_{k}(g)\|^2,\)

\(X^{(\psi)}_{k+1} = \frac{1}{2} \sum_{x} \|\mathbf{A}_{k+1}(x) - \mathbf{A}_{k}(x)\|^2,\)

(2)

Complementarily, we use the phase consistency check factor (a ‘correctness’ factor)

\(C^{(\psi)}_{k+1} = \frac{1}{2} \sum_{g} \|\psi_{k+1}(g) - \psi_{k}(g)\|^2,\)

(3)

that is calculated in a process of numerical simulations for all the values of \(k\), and the sum, \(\Sigma\), taken over all the reflections except \(g = 0\). The \(C^{(\psi)}_{k}\)-factor being calculated at the each iteration step \(n\) as a function of \(k\) provides a means of monitoring the progress of the code during the entire iteration procedure. The numerical phase simulation procedure (see Eqs.1)-(3) from \(n = 2\) until the final iteration value of \(N\) is iterated guiding the phases as long as the values of \(R^{(\psi)}\) and \(X^{(\psi)}\) reduce to the consensus estimates. Note that Eqs.(2) play the role of penalty functions for constraining likely atomic structure solutions operating with the calculated and a priori known ED and HRTEM data.

3. Numerical simulations. Results and discussion

The test case we consider is a centrosymmetric structure of the ceramic \((\text{Ga,In})_2\text{SnO}_3\) (Fig. 1). The structure is monoclinic with \(a = 1.169\) nm, \(b = 0.317\) nm, \(c = 1.073\) nm, and \(\beta = 99.00^\circ\). The atom positions are taken from the work (Sinkler et al., 1998). The complex structure factors related to the proper dynamical ED pattern for \((\text{Ga,In})_2\text{SnO}_3\) were calculated using 1-state channeling approach (see, e.g., [24, 25] for details) for 4.121 nm thickness (Fig. 2) and 7.291 nm thickness (Fig. 3) along the \(b\)-axis. The standard images calculated using the above values of thickness and 0.01 nm resolution are shown in Fig. 2a and Fig. 3a, respectively. It was assumed, that the incident electron beam propagate along the \(b\)-axis, and the accelerating voltage is 300 kV.

![Fig. 1. Structure of (Ga,In)_2SnO_3 viewed along [010] direction.](image-url)
Fig. 2. Images of (Ga,In)\textsubscript{2}SnO\textsubscript{3}-structure (2D- and 3D-plots) showing the phase retrieval solution for the iteration range of \( N = 20 \) in (b) and the phase retrieval solution for the iteration range of \( N = 40 \) in (c) in comparison with the standard gauge image in (a) calculated for the iteration range of \( N = 100 \). The cycle range of \( K \) is 1000. The sample thickness along the \( b \)-axis is 4.121 nm.

To illustrate the convergent features of the new algorithm guiding the phase determination based upon the flow equation (1) the data sets of \( \{ |U(g)|, A(x) \} \) were explored in a sequent manner from the initial value of \( v = 2 \) until the two different final iteration values of \( N \), namely: \( N = 20 \) and \( N = 40 \), respectively. Fig. 2(b,c) and Fig. 3(b,c) show the calculated unit cell images, \((20)I(x)\) and \((40)I(x)\), obtained as the modulus squared of the Fourier transform of the unique phase-recovered structure factors sets of \((20)\{U(g)\}\) and \((40)\{U(g)\}\) using the domino iterative transform algorithm detailed above. It is interesting that in the case of the iteration number of
N = 20 the real-space unit cell images (see Figs 2b, 3b) do not keep the centrosymmetry property as a result of the insufficient spatial resolution of order of 0.05 nm (it should be noted that the atom positions need not be coincided with nodes of a numerical net) but not to the phase retrieval solution based on the present iteration transform algorithm application.

As an example, Fig. 4 shows the behavior of the calculated parametric plots of the R-FOM versus a 'correctness' C-factor (cf. Eqs (2)-(3)). It is found that in the case...
of the domino iterative transform algorithm application to the solution of (Ga,In)$_2$SnO$_3$ the calculated R-FOM and C-factor are achieving the appropriate values of order of $10^{-5}$ and $10^{-7}$ for the cycle range of $K = 100$ used for the calculated images shown in Fig. 3b,c. It is worth to notice that in the case of the cycle range value of $K = 1000$ taken for the numerical simulation of the real-space structure images in Fig. 2b,c the corresponding values of R-FOM and C-factor are practically equal to zero (less than $10^{-27}$).

Thus, what is referred to as trends of the R-FOM reduction observed as well as confirmed by the C-factor reduction behavior, the domino iteration transform algorithm by consequently combining the ED and HRTEM data indicates a good convergent skill in a process of the phase retrieval procedure. It should be noted that depending on the initial random phases of an input phase set of complex structure factors for the initial iteration value of $v = 2$ the numerical code proposed generates the redundant solutions too. Fortunately, they can be effectively discriminated by imposing the ‘break’-condition incorporated within complete iteration procedure, namely: for instance, by choosing inappropriate barrier values of R-FOM being equal to 0.01.

4. Concluding remarks

In this paper, the goal of our study is to justify the application of the domino iterative transform algorithm detailed above for phasing diffraction patterns. The main point of the new numerical code is that the latter provides the robust unambiguous iterative procedure, at least, for two-dimensional phasing problems by operating with the ED and HRTEM data as the physically measured constraints. Specifically, what is referred to as a general phase retrieval problem the present domino method can be considered as a synthesis of the diverse direct methods, which are widely explored for electron structure determination [6, 22]. Numerical simulations of the crystal model structure show that unlike the direct methods earlier utilized the routine algorithm by using the composite input ED and HRTEM data sets in a domino fashion is reasonable from the physical viewpoint and does not depend on any assumptions (e.g., the weak-scattering approximation) and primary models of the crystal structure in question to restore phases of the complex structure factors. Generally, the present algorithm code is proven to be convergent and successful for solving phase retrieval problems (at least in the case of the two-dimensional data sets). It seems likely that even complicated crystal structures, for which some part of reflections are not available to be measured, can be solvable in this way by extending the feasible complex structure factors to unmeasurable ones by means of the Sayre equation.

A few final comments are appropriate here about the implementation of the present algorithm code to the practical structural analysis using the ED and HRTEM data. As is pointed out, the ED dynamical effects were taken into consideration within the scope of a 1s electron channeling approach for the input ED data. It needs to be remembered that the validity and application prospect of the domino phase retrieval algorithm presented are closely related to the ED being dominated by 1s channeling states. That is only true for moderate values of sample thickness and atomic numbers. In addition to sample thickness, accelerating voltage is another adjustable parameter. As the voltage becomes higher, there are more strongly bound states (e.g., 2s, 3s), which complicate the ED [24]. Yet, there are many structures where 2p-states are important if the sample is tilted off the zone axis. The question of whether the domino phase retrieval algorithm has any validity in these cases remains topics for future research.

Besides, in view of testing with the present algorithm code we didn’t include the aberration phase distortions of electron optical system, the values of which are in general known a priori to form relevant HRTEM images (cf. [2]) and ignored a possible thickness variation within the sample area selected, albeit the latter might be minimized using the convergent-beam patterns.

With these confines of the input ED and HRTEM data sets, we do not claim anything except that the domino algorithm code for the phase retrieval of structure factors is really convergent and works well. How well it will work...
in practice, particularly using the input experimental ED and HRTEM data, remains to be seen and is then of a special interest for future work.

It should once more be stated in conclusion that the convergent feature and feasibility of the numerical solution of the crystal model structure tested in the general case of complex structure factors imply that the direct methods may be properly modified to facilitate further the practical electron structure analysis, particularly, by applying the domino iterative transform algorithm technique, a good topic of future work.

References