Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 30 June 2003
Accepted 24 September 2003

Phase extension in crystallography using the iterative Fienup–Gerchberg–Saxton algorithm and Hilbert transforms

J. S. Wu* and J. C. H. Spence

Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1504, USA. Correspondence e-mail: jinsong.wu@asu.edu

A procedure for phase extension in electron crystallography is proposed based on the iterative Fienup–Gerchberg–Saxton algorithm in combination with the use of discrete Hilbert transforms. This transform is used to provide oversampling in reciprocal space, thus satisfying the Shannon sampling requirement and introducing reflections with fractional indices. When the procedure is combined with the knowledge of a small set of strong phased Bragg reflections from electron-microscope images (or direct methods), the magnitudes of many non-Bragg reflections can be calculated with useful accuracy, thus enhancing the performance of the iterative algorithm for phase extension. The effects of various constraints used in the iterative algorithm are discussed. In this way, it is shown that the iterative algorithm conventionally used for phasing diffuse scattering from non-periodic objects can also be applied to problems in conventional crystallography to find the phases of high-order (resolution) beams from a known set of low-order (resolution) phases.

1. Introduction

It was pointed out long ago (Sayre, 1952) that Bragg sampling undersamples the intensity scattered by a molecule in a crystal. By this it is meant that the Shannon or Nyquist sampling interval needed to synthesize the scattered intensity as a continuous function of angle would require diffraction from a crystal at half the Bragg angle (in one dimension). The intensity scattered by the molecule is here treated as a band-limited function, with the 'band-limit' applied in real space by the autocorrelation function of the molecule, which has twice the size of the molecule. It has also been found that, if the scattering from an isolated charge density can be sampled at the Shannon rate, then the phase problem can usually be solved using the iterative Fienup–Gerchberg–Saxton algorithm, which we here refer to as *HiO* (hybrid input–output) (Fienup, 1982). [Ambiguities can be expected, for example, from the rare homomorphic crystals (Buerger, 1954).] For recent experimental examples of the inversion to images of soft X-ray and electron diffraction patterns from isolated objects, see He *et al.* (2003) and Zuo *et al.* (2003). The algorithm is based on three constraints – the known sign of the (real) charge density, the known Fourier moduli, and the known boundary of the object.

The success of the *HiO* algorithm in solving the phase problem for non-crystallographic isolated objects whose boundary (or 'support') is approximately known has encouraged many efforts to apply this method to crystals, using, for example 'non-crystallographic' symmetries, fiber diffraction (Millane, 1990; Millane & Stroud, 1997; De Caro *et al.* 2002) and finite-crystal effects. Since the number of unknown phases is equal to the number of Fourier equations linking real and reciprocal space if the Shannon sampling interval is respected, application of the method to crystals reduces to the problem of finding the 'half-order' Bragg reflections that would be produced by doubling the unit-cell dimensions but not the size of the molecule within each cell ('oversampling'). (The method of solvent flattening, applied to proteins surrounded by a water jacket of known density, is closely related.)

Recently, it has been pointed out that a Hilbert transform relationship exists between the real and imaginary parts of the transform of the density of a single molecule (the scattering factor for the molecule), since the density for an isolated molecule is a causal function (zero for negative arguments) for some choice of origin (Mishnev, 1996). Application of this condition is thus equivalent to the application of a known support condition in the *HiO* algorithm, which is required for its convergence. However the Hilbert transform also provides an expression for the 'half-order' reflections mentioned above in terms of the known complex Bragg reflections. The Hilbert transform has been proposed previously as a means for phase extension and improvement (Zanotti *et al.*, 1996), although later it was shown by De Caro *et al.* (2002) that a non-realistic weighting scheme could then be involved. By phase extension, we refer to the improvement of resolution by phasing high-order Bragg beams based on certain known phases for low-order reflections. (Phase improvement refers to a similar process without resolution improvement.)

Table 1
Structure factors for some non-Bragg reflections with fractional indices for α-copper phthalocyanine structure.

The structure factors calculated using known atomic coordinates [$|\mathbf{F}_g| = \sum f \exp(-2\pi i \mathbf{r}\cdot\mathbf{g})/V$] are listed in column 3, based on the Fourier transform of one molecule. These are compared with those calculated using the Hilbert transform [equations (2) and (3)] applied to: (i) the magnitudes of the 3481 Bragg reflections with $|h| < 12$ and $|k| < 29$ ($|\mathbf{F}_g|\_1$ in column 4); (ii) 137 strong phased Bragg reflections with $|h| < 12$ and $|k| < 12$ ($|\mathbf{F}_g|\_2$ in column 5); (iii) 81 strong phased reflections with $|h| < 12$ and $|k| < 12$ ($|\mathbf{F}_g|\_3$ in column 6); and (iv) 23 strong phased reflections with $|h| < 12$ and $|k| < 12$ ($|\mathbf{F}_g|\_4$ in column 7). The R_m factor [equation (4)] is also given for each set of data.

h'	k'	$\lvert\mathbf{F}_g\rvert$ true	$\lvert\mathbf{F}_g\rvert\_1$ intensity only	$\lvert\mathbf{F}_g\rvert\_2$ 137 phases	$\lvert\mathbf{F}_g\rvert\_3$ 81 phases	$\lvert\mathbf{F}_g\rvert\_4$ 23 phases
0	0.5	5.7714	5.7160	5.7740	5.7765	5.7753
0.5	0	5.7500	5.7098	5.7563	5.7523	5.7404
0.5	0.5	3.6113	3.5449	3.6196	3.6275	3.6332
0.5	−0.5	3.6000	3.5633	3.6043	3.5904	3.5685
0	1.5	1.3265	1.1386	1.3170	1.3245	1.3210
0.5	1.5	0.9378	0.7519	0.9273	0.9425	0.9346
0.5	2	0.5565	0.7109	0.5671	0.5563	0.6114
1.5	0	1.2317	1.0969	1.2237	1.2114	1.1694
1.5	0.5	0.7096	0.6645	0.7130	0.7026	0.7265
1.5	−0.5	0.8681	0.7580	0.8500	0.8198	0.7470
1.5	−1.5	0.2318	0.3164	0.2312	0.1167	0.0801
1.5	1.5	0.2391	0.2491	0.2509	0.1954	0.2285
1.5	2.5	0.0799	0.1232	0.0664	0.0072	0.1587
2	0.5	1.1192	0.9630	1.1008	1.1283	1.0352
2	1.5	0.4007	0.1711	0.3814	0.5224	0.4030
2	2.5	0.6134	0.2301	0.5978	0.4520	0.2659
2.5	1.5	0.7483	0.4603	0.7265	0.8458	0.7126
2.5	−2	0.7176	0.1856	0.7151	0.5368	0.4911
2.5	2.5	0.4310	0.1410	0.4294	0.3269	0.2345
R_m			0.49	0.15	0.30	0.68

In this paper, our aim is phase extension for increased resolution in electron crystallography. Thus, we seek to use these ideas to solve the phase problem for high orders for a thin organic crystal (perhaps a protein monolayer), using electron diffraction intensity data, together with some electron-microscope images to provide structure-factor phases for some low orders. In a previous paper, we have shown how the finite thickness of such a thin crystal measured along the beam direction can be used to provide a compact support condition and so assist in solving the phase problem (Spence *et al.*, 2003). The aim here is to use the Hilbert transform and *HiO* iterations to find the phases of the higher-order Bragg beams, and so improve the resolution of the images. Experimentally, the collection of high-resolution images of organic thin crystals is extremely difficult because of radiation damage effects, especially if data are required in three dimensions. Diffraction data are much less demanding to collect. Since the lowest resolution at which amino acids can be identified is about 3 Å, a useful goal would be to extend image data that provides phased Bragg beams to 5 Å resolution out to, say, 2 Å resolution using the *HiO* algorithm and high-order Bragg-beam intensities. Our overall strategy is to use the Hilbert transform and the fewest possible image phases to find some low-order half-order reflections. Then the *HiO* algorithm uses all this information to phase the full set of high-resolution Bragg beams. In the process, it finds all the remaining fractional order beams. The resulting charge-density map shows the molecule within a doubled cell, surrounded by a border of zero charge density. This has been referred to elsewhere as the so-called 'confined structures' defined by De Caro *et al.* (2002). For continuously bonded solids (such as silicon), this leads to truncated bonds which must be re-assembled.

2. Fractional-indexed reflections and the Hilbert transform

By application of the Shannon theorem and the Hilbert transform, Mishnev (1996) obtains the following expression for complex half-order structure factors in terms of Bragg structure factors:

$$F(h'+0.5,k'+0.5,l'+0.5) = \frac{1}{i\pi^3}\sum_{h,k,l}\frac{F(h,k,l)}{(h'-h+0.5)(k'-k+0.5)(l'-l+0.5)}. \quad (1)$$

These fractional order structure factors are samples of the molecular scattering factor at points midway between Bragg beams. We have confirmed using artificial superlattice calculations that (1) predicts the same half-order structure factors as those obtained by doubling the cell constants (but not the size of the molecule within the cell), on the condition that they have the same cell origin. This can be referred to as the 'confined structure' of De Caro *et al.* (2002).

We have used the structure of α-copper phthalocyanine (α-CuPc) with a monoclinic cell [a = 2.592, b = 0.379, c = 2.392 nm and β = 90.4° (space group: $C2/c$) (Ashida *et al.*, 1966; Brown, 1968)] for these two-dimensional simulations. Since only Bragg intensities are known initially, we first attempt to estimate the fractional order structure factors from these intensities alone, then add in increasing numbers of known phases from images. We therefore use the following formulas to compute two-dimensional structure factors:

$$|F(h'+0.5,k'+0.5)| = \frac{1}{\pi^2}\left|\sum_{h,k}\frac{F(h,k)}{(h'-h+0.5)(k'-k+0.5)}\right| \quad (2)$$

and

$$|F(h'+0.5,k)| = \left|\frac{1}{i\pi}\sum_{h}\frac{F(h,k)}{(h'-h+0.5)}\right|. \quad (3)$$

The results are shown in Table 1 for some low-order non-Bragg reflections with indices smaller than 3. In the table, the symbols h' and k' represent the indices of reflections that may have either integer or fractional values.

The third column in Table 1 lists the true values of the fractional-index structure factors, as obtained from the Fourier transform of the molecule. In column 4, for comparison, we list values predicted using slightly modified versions of (2) and (3), so as to use only the magnitudes on the left of the equations. [The term $F(h,k)$ is replaced by $|F(h,k)|$ in (2) and (3)]. Thus the sum is taken over only the magnitude

of 3481 simulated Bragg reflections with a maximum resolution of 0.9 Å. These predictions are seen to show poor agreement with the true values in column 3, since the real coefficients in the sum do not represent a causal density. Mishnev (1996) has discussed the Hilbert transform for intensity, and it has been shown that, in the general case, 'intensity from intensities' is not feasible. This problem was also addressed by means of a probabilistic approach (Giacovazzo *et al.*, 1999) and the Patterson property of confined structures (De Caro *et al.*, 2002). The situation improves if some known phases are used.

In columns 5–7, a steadily decreasing number of phases are added to the sum on the right of (2) and (3) in order to determine the lowest-resolution images that will accurately predict the half-orders needed by the *HiO* algorithm for convergence. At the bottom of each column is given an R factor to indicate agreement with the true fractional orders:

$$R_m = \frac{\sum ||\mathbf{F}_g^{\rm known}| - |\mathbf{F}_g^{\rm cal}||}{\sum |\mathbf{F}_g^{\rm known}|}. \quad (4)$$

Here, $|\mathbf{F}_g^{\rm known}|$ are the true magnitudes, while $|\mathbf{F}_g^{\rm cal}|$ are the calculated magnitudes using (2) and (3). We see that the set of 3481 intensities can be used together with only 137 known phases to give a rather accurate prediction (R = 0.15) of the fractional-order complex structure factors. We note that, for these calculations, the value of $\mathbf{F}(000)$ is needed, which has a magnitude much larger than that of other reflections. Empirical tests also show that, when the resolution of the calculated fractional-indexed reflections is higher than that of known Bragg reflections, a large error occurs. This may result from the fact that (2) and (3) cannot be used owing to the limited number reflections with known phases. We note also that, if the number of known Bragg reflections is too small, the error due to truncation becomes very large (column 7 in Table 1).

3. Phase-extension results

Mathematical details of the *HiO* algorithm are given in Weierstall *et al.* (2001). In this application, we modify the algorithm as follows:

1. Upsampling the magnitudes of the given Bragg beams $|F(u,v)|$ of dimension $n \times n$ by a factor of 2 generates $|G(u,v)|$ with dimension $2n \times 2n$ by assigning arbitrary magnitudes to fractional-indexed reflections in between the Bragg reflections. This is equivalent to surrounding each molecule in the crystal by an envelope (similar to a water jacket around a protein) of zero density, and doubling every cell dimension to accommodate the jacket.

2. We assign each reflection in $G(u,v)$ a random phase to obtain the starting structure-factor matrix denoted $G_k(u,v) = |G_k(u,v)| \exp[i\Phi_k(u,v)]$, where k represents the iteration number.

3. Apply the known data to $G_k(u,v)$ to get $G'_k(u,v)$:

$$G'_k(u,v) = \begin{cases} |F(u,v)| \exp[i\Phi_k(u,v)] & \text{if } (u,v) \in L(u,v) \\ G_k(u,v)[F(0,0)/G_k(0,0)] & \text{if } (u,v) \notin L(u,v). \end{cases} \quad (5)$$

Here, $L(u,v)$ represents the set of points in reciprocal space at which the magnitudes of Fourier coefficients $|F(u,v)|$ are known. For the in between reflections, whose magnitudes are unknown, their values are taken from the current estimate $G_k(u,v)$ with a normalization factor $F(0,0)/G_k(0,0)$. The normalization factor can also be calculated using

$$\sum_{L(u,v)} F(u,v) \Big/ \sum_{L(u,v)} G_k(u,v).$$

In this step, any additional valid constraints can also be applied: $G'_k(u,v) \overset{\text{constraints}}{\longrightarrow} G'_k(u,v)$ (details of constraints will be discussed in §4).

4. Inverse Fourier transform $G'_k(u,v)$ to obtain the potential in real space: $G'_k(u,v) \overset{\text{iFFT}}{\longrightarrow} g'_k(x,y)$.

5. Apply the support $S(x,y)$ in real space as for the hybrid input–output (*HiO*) method:

$$g_{k+1}(x,y) = \begin{cases} g'_k(x,y) & \text{if } (x,y) \in S(x,y) \\ g_k(x,y) - \beta g'_k(x,y) & \text{if } (x,y) \notin S(x,y). \end{cases} \quad (6)$$

Here $S(x,y)$ is a two-dimensional support defined as the region outside of which the potential is known to be zero, and β is a feedback parameter between 0.5 and 1. For the error-reduction (ER) method, the current potential is defined as:

$$g_{k+1}(x,y) = \begin{cases} g'_k(x,y) & \text{if } (x,y) \in S(x,y) \\ 0 & \text{if } (x,y) \notin S(x,y). \end{cases} \quad (7)$$

Here we can also apply other real-space constraints, such as the known sign of the potential or the histogram constraint: $g_{k+1}(x,y) \overset{\text{constraint}}{\longrightarrow} g_{k+1}(x,y)$.

6. Fourier transform $g_{k+1}(x,y)$ to obtain $G_{k+1}(u,v)$.

7. Go to step 3 with k replaced by $(k+1)$.

We have used the Hilbert transform equations to obtain many reliable magnitudes of the non-Bragg reflections. We then use these estimates in the *HiO* algorithm as additional known constraints. Thus, in addition to the magnitudes of the Bragg beams, the calculated magnitudes of some non-Bragg reflections from the Hilbert transform were also used in step 3 in the *HiO* algorithm. The progress of the iterations can be followed using a normalized root-mean-square (RMS) error (Fienup, 1982), which shows the amount by which the potential violates real-space constraints outside the support (where it must be zero):

$$\text{RMS} = \left[\sum_{(x,y) \notin S} |g(x,y)|^2 \Big/ \sum_{x,y} |g(x,y)|^2 \right]^{1/2}. \quad (8)$$

The cross-correlation coefficient (CC) between the true and estimated structures was also used to monitor the convergence of the algorithm (Read, 1986):

$$\text{CC} = \frac{\sum_h |\mathbf{F}_h^k||\mathbf{F}_h^c| \cos(\Phi_h^k - \Phi_h^c)}{[\sum_h |\mathbf{F}_n^k|^2 \sum_h |\mathbf{F}_h^c|^2]^{1/2}}, \quad (9)$$

where $\mathbf{F}_h^k = |\mathbf{F}_h^k| \exp(\Phi_h^k)$ is the structure factor of the known structure and $\mathbf{F}_h^c = |F_h^c| \exp(\Phi_h^c)$ is that of the current calculated structure. The value of CC ranges from 1 representing perfect agreement to -1 for an anticorrelated result. We note that CC cannot be evaluated for an unknown structure, hence it is important to show that the dependence of RMS (which *is* known for an unknown structure) is monotonic with CC.

We used the data with 81 phased reflections (column 6 in Table 1) as a test of phase extension. As listed in Table 1, R_m is 0.30 when all fractional reflections with resolution out to 0.9 Å were calculated. For the potential map, we did not include so many high-resolution fractional-order reflections. Only 712 magnitudes of non-Bragg beams with resolution down to 2.0 Å were calculated using (2) and (3), based on the known reflections, giving $R_m = 0.1889$ [equation (4)]. Compared to the results listed in Table 1, we have reduced the resolution of the fractional-indexed reflections, which leads to an improvement in the accuracy of the calculated magnitudes. The potential obtained by Fourier synthesis using the 81 known reflections is shown in Fig. 1 and its cross-correlation coefficient, calculated using (9), is CC = 0.81. The calculated magnitudes and partially known phases were taken as known constraints in the iterative Fienup–Gerchberg–Saxton algorithm. We used 15 *HiO* iterations followed by 5 ER iterations as one cycle with a feedback factor of 0.78. The program runs 5 cycles altogether. Fig. 2 shows the reconstructed potential after 100 iterations, and the RMS and CC values as a function of iteration number. The final CC value has been improved to 0.96 by this method, showing that the phase extension is successful. If the number of Bragg beams used with known phases is much smaller (*e.g.* only 27 as listed in column 7 in Table 1), then phase extension using the iterative algorithm does not improve the cross-correlation coefficient.

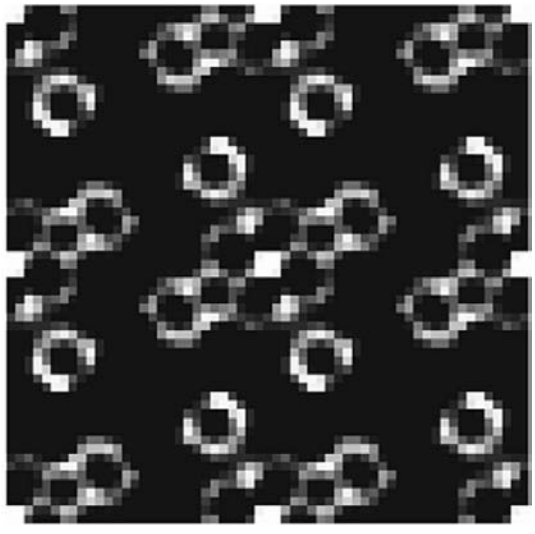

Figure 1
Calculated potential of the α-copper phthalocyanine structure by the Fourier synthesis using 81 strong Bragg reflections with correct phases. Its CC value is calculated to be 0.81 compared to the known structure at 0.9 Å resolution.

4. Discussion

Constraints play a very important role in the *HiO* iterative algorithm. Constraints in reciprocal and real space are the key points that drive the iterations toward a unique solution. A constraint is effective if many equations result from its application. Non-negativity is a well known constraint used in X-ray and electron crystallography when kinematic diffraction conditions apply. In recent work, two approaches have commonly been used: one is to invert the sign of all negative points and the other is to set all the negative points to zero or a constant, determined by some property of the crystal, for example, the average potential. The kinematic diffraction condition also ensures that the structure factors obey Friedel's law ($|F| = |\bar{F}|$ and $\Phi_F = -\Phi_{\bar{F}}$).

Symmetry can also be a powerful constraint if it is known. We must distinguish two kinds of symmetries: one is the non-crystallographic symmetry and the other is the conventional crystallographic symmetry (space group). Non-crystallographic symmetry can relate the fractional-order reflections,

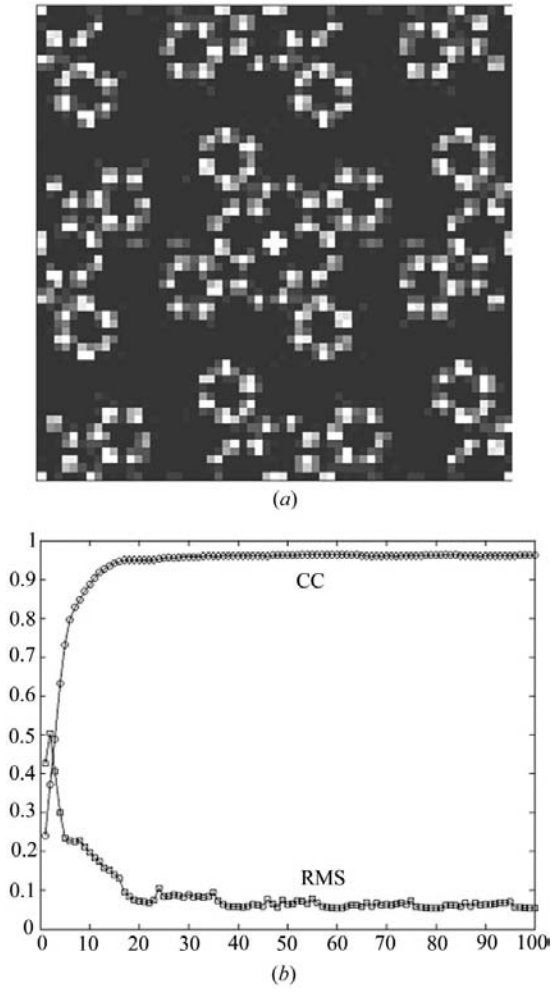

Figure 2
(*a*) Reconstructed (010) potential of α-copper phthalocyanine using the iterative Fienup–Gerchberg–Saxton phase-extension algorithm after 100 iterations. The phases are improved and the cross-correlation coefficient (CC) is 0.96 for the final iteration. (*b*) The normalized root-mean-square error (RMS) and CC as a function of iteration.

and this has been a critical factor in some successful applications (Millane & Stroud, 1997). However, obtaining non-crystallographic symmetries for a crystal requires prior knowledge of the local symmetry of molecules in the unit cell (point group), *e.g.* the conformation of the molecule, and this must be a point-group symmetry which is incompatible with the periodic reciprocal lattice, such as a fivefold rotation axis in a quasicrystal or virus, in order to derive information relating points off the reciprocal lattice. Crystallographic symmetry as revealed by the space group of the crystal is also useful. Although the distribution of measured structure-factor magnitudes obeys the requirements imposed by crystallographic symmetry, it can also be used as a constraint for current estimates in the *HiO*.

Some previously used density-modification techniques, *e.g.* real-space histogram matching (Zhang & Main, 1990) and identification of the molecular envelope (Wang, 1985) can also be effective. Those constraints take effect only in the region where the support $S(x,y)$ is not zero. It would be helpful to apply histogram constraints in reciprocal space to establish more equations, since many fractionally indexed reflections with unknown magnitudes have been added. At present, this does not seem possible. Unlike a general Wilson statistical property of structure factors with integral indices, the density function of the magnitudes of structure factors with fractional indices $P(|F|)$ do not have universal properties and depend on the structure complexity (Giacovazzo & Siliqi, 1998).

We have found a useful constraint that can be used for all reflections (Bragg and non-Bragg) in reciprocal space and that is also applicable in real space. In a Fourier transform calculation, the potential (or charge density) of a crystal is firstly sampled into grid points in real space, and the diffraction pattern after transformation is also a discrete function, represented by grid points in reciprocal space. Some distances are preserved by Fourier transform, *i.e.*

$$\sum_{x,y} |g(x,y)|^2 = \sum_{u,v} |G(u,v)|^2 \big/ p \qquad (10)$$

for a two-dimensional Fourier transform $g(x,y) \overset{\text{Fourier transform}}{\longleftrightarrow} G(u,v)$, where p is the number of grid points. Letting $\langle |G(u,v)|^2 \rangle = \sum_{u,v} |G(u,v)|^2/p$ represent the mean square of the structure-factor magnitudes, we have:

$$\sum_{x,y} |g(x,y)|^2 = \langle |G(u,v)|^2 \rangle. \qquad (11)$$

Similarly, we have

$$\sum_{x,y} |f(x,y)|^2 = \langle |F(u,v)|^2 \rangle \quad \text{and} \quad \sum_{x,y} |f(x,y)|^2 = \sum_{x,y} |g(x,y)|^2$$

since $g(x,y)$ is an expansion of $f(x,y)$ by adding zeros. Finally, it can be shown that

$$\langle |F(u,v)|^2 \rangle = \langle |G(u,v)|^2 \rangle = \sum_{x,y} |g(x,y)|^2. \qquad (12)$$

Equation (12) can thus be used as a constraint in both real and reciprocal space.

For phase retrieval in crystallography, atomicity constraints are used in direct methods, and these can also be a useful constraint. The Sayre equation (Sayre, 1952) expresses atomicity by using scattering factors for the normal and squared electron density. The Sayre equation is normally expressed in reciprocal space:

$$\mathbf{F}(\mathbf{h}) = [\theta(\mathbf{h})/V] \sum_{k} \mathbf{F}(\mathbf{k})\mathbf{F}(\mathbf{h}-\mathbf{k}), \qquad (13)$$

where the scale factor $\theta(\mathbf{h})$ is the ratio of the scattering factors of the real and squared atoms, $\mathbf{h}$ and $\mathbf{k}$ are reciprocal-lattice vectors and V is the volume of the unit cell. Taking the inverse Fourier transform of (13), Sayre's equation is expressed in real space as (Woolfson & Fan, 1995)

$$\rho(\mathbf{n}) = (V/N) \sum_{m} \rho^2(\mathbf{m})\Psi(\mathbf{n}-\mathbf{m}), \qquad (14)$$

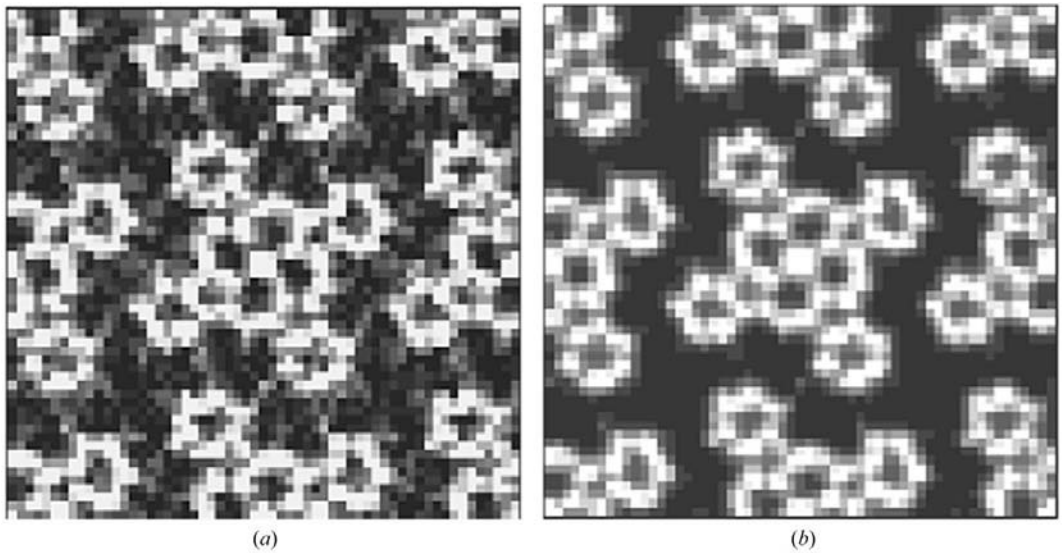

Figure 3
Application of the Sayre constraint in real space [equation (14)] of an estimated (010) projected potential of the α-copper phthalocyanine structure (*a*) gives rise to (*b*).

where $\rho(\mathbf{n})$ is the electron density (potential for electron diffraction) expressed as a discrete function evaluated at N grid points, and $\Psi(\mathbf{n})$ is the Fourier transform of $\theta(\mathbf{h})$. We have used (14) as a real-space constraint to save computing time [compared to (13)]. Fig. 3 shows an example of the application of Sayre's equation, which has an effect similar to contrast enhancement in image processing. In our tests, however, it was not found useful to greatly enhance the contrast of the current reconstructed potential since this would drive the iteration into a local minimum and lead to stagnation.

Non-negativity and the support constraint are the two fundamental constraints in *HiO*, and the iterations do not converge without them. We have used the (010) projection of the α-CuPc structure to test the effects of the constraints. Fig. 4(*a*) shows the results of phase extension when all the other constraints (*i.e.* constraints of symmetry, equal mean and real-space histogram) other than non-negativity and the support constraint were excluded. The cross correlation was 0.86 after 100 iterations. This shows that, although those constraints cannot be used alone to solve by *ab initio* phase identification, they help to find the solution more quickly. The magnitudes of the non-Bragg beams calculated using (2) and (3) are also important. Fig. 4(*b*) shows the results when those magnitudes were not included – we find that CC drops to 0.82 after 100 iterations.

Although there still appears to be no way to use the iterative algorithm to solve the phase problem *ab initio* in crystals using only the magnitudes of the Bragg reflections, we have shown by simulation that such an iterative process can be used for phase extension and improvement. This is not completely unexpected since it has already been noted that the iterative algorithm works with incomplete data sets (Weierstall *et al.*, 2001). This process can be greatly improved by calculating the fractional-index reflections using the Shannon sampling theorem and Hilbert transform. We then find that having the phases of a small set of strong reflections is sufficient for phase extension. The small number of known phases can be obtained using either direct methods or from a high-resolution electron-microscope image. The combination with direct methods may be especially beneficial.

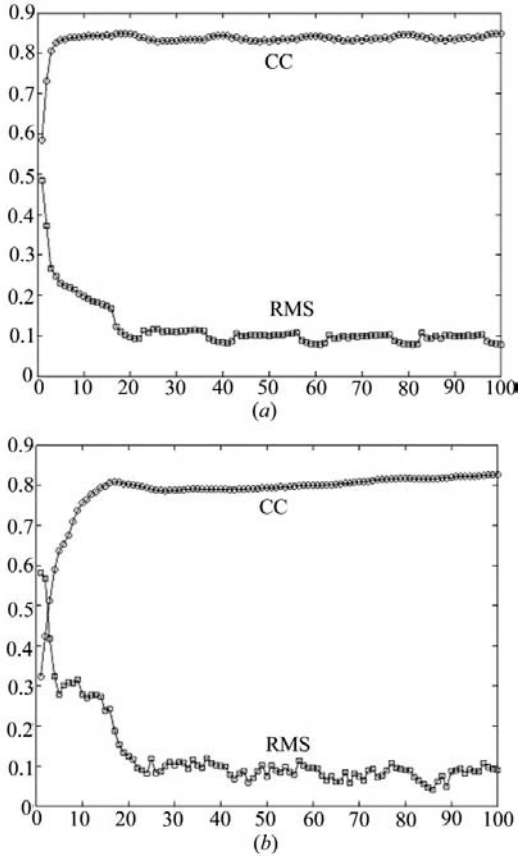

Figure 4
The normalized root-mean-square error (RMS) and the cross-correlation coefficient (CC) as functions of iteration for phase extension without several reciprocal- and real-space constraints except non-negativity and support constraints (*a*), and for phase extension with all the list constraints but not the magnitudes of non-Bragg reflections calculated using equation (2) and (3) (*b*).

5. Conclusions

The iterative Fienup–Gerchberg–Saxton *HiO* algorithm has been shown to be effective for phase extension and improvement. Based on an incomplete set of phases, the magnitudes of many non-Bragg reflections with fractional indices can be calculated with high accuracy using equations derived from the Shannon sampling theorem and the Hilbert transform. This greatly improves the performance of the iterative algorithm and the results of our tests, using a cross-correlation coefficient, improved from 0.81 to 0.96. In addition, the resolution improved from 2.4 to 0.9 Å. Although previously the *HiO* algorithm has been commonly used for retrieving phases from 'oversampled' intensity in reciprocal space from non-periodic objects, we show here that it can also be used for undersampled (Bragg beam) data for phase extension in crystals.

This work is supported by ARO award DAAD190010500.

References

Ashida, M., Uyeda, N. & Suito, E. (1966). *Bull. Chem. Soc. Jpn*, **39**, 2616–2624.
Brown, C. J. (1968). *J. Chem. Soc. A*, pp. 2494–2498.
Buerger, M. (1954). *Vector Space.* New York: McGraw-Hill.
De Caro, L., Giacovazzo, C. & Siliqi, D. (2002). *Acta Cryst.* A**58**, 415–423.
Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
Giacovazzo, C. & Siliqi, D. (1998). *Acta Cryst.* A**54**, 957–970.
Giacovazzo, C., Siliqi, D., Fernandez-Castano, C. & Comunale, G. (1999). *Acta Cryst.* A**55**, 525–532.
He, H., Marchesini, S., Howells, M., Weierstall, U., Hembree, G. & Spence, J. C. H. (2003). *Acta Cryst.* A**59**, 143–152.
Millane, R. P. (1990). *J. Opt. Soc. Am.* A**7**, 394–411.
Millane, R. P. & Stroud, W. J. (1997). *J. Opt. Soc. Am.* A**14**, 568–579.
Mishnev, A. F. (1996). *Acta Cryst.* A**52**, 629–633.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
Spence, J. C. H., Weierstall, U., Downing, K. & Glaeser, R .M. (2003). *J. Struct. Biol.* In the press.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.
Weierstall, U., Chen, Q., Spence, J., Howells, M., Isaacson, M. & Panepucci, R. (2001). *Ultramicroscopy*, **90**, 171–195.
Woolfson M. & Fan H. F. (1995). *Physical and Non-physical Methods of Solving Crystal Structures*. Cambridge University Press.
Zanotti, G., Fogale, F. & Capitani, G. (1996). *Acta Cryst.*, A**52**, 757–765.
Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.*, A**46**, 41–46.
Zuo, J. M., Vartanyants, I., Gao, M., Zhang, R. & Nagahara, L. A. (2003). *Science*, **300**, 1419–1421.