



Inter-Bragg crystallographic phase retrieval from shape transforms, stacking faults and substitutional disorder

J.P.J. Chen ^{a,b,*}, K. Pande ^c, J.J. Donatelli ^c, A.V. Martin ^d, K. Ayer ^e, H.N. Chapman ^f, R. Bean ^g, K.E. Schmidt ^a, R.A. Kirian ^a

^a Department of Physics, Arizona State University, Tempe, AZ 85287, USA

^b Department of Electrical and Computer Engineering, University of Canterbury, Christchurch 8041, New Zealand

^c Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

^d School of Science, RMIT University, Melbourne, VIC 3000, Australia

^e Max Planck Institute for the Structure and Dynamics of Matter, 22761 Hamburg, Germany

^f Center for Free-Electron Laser Science, 22761 Hamburg, Germany

^g European XFEL, 22869 Schenefeld, Germany

ARTICLE INFO

Dedicated to the memory of John C.H. Spence

Keywords:

Shape transform phasing

Lattice occupancy model

Crystal disorder

Phase retrieval

Iterative projection algorithms

ABSTRACT

One of the brilliant ideas of John Spence when he saw the first diffraction patterns from the Linac Coherent Light Source was that one could solve the crystallographic phase problem by utilising the intensities between Bragg peaks. Because these intensities are due to the Fourier transform of the shape of the crystal, the approach came to be known as “shape-transform phasing.”

Shape-transform phasing was developed over the next ten years and formed the basis for many other interesting ideas and pursuits. Here we describe the current best implementation of the original idea using a lattice occupancy formalism and show that certain types of crystal defects can also be modelled via this approach, allowing the molecular structure to be recovered from the additional information offered by the inter-Bragg intensities from these crystal defects.

1. Introduction

Once upon a time; December 2009 to be exact, a small crystal was dropped into the very first operational X-rays generated by the Linac Coherent Light Source (LCLS). The X-rays diffracted from that crystal and created the diffraction pattern shown in Fig. 1. The pattern caused a stir in the control room that prompted John Spence to leap up from his slumber. Waving his arms in the air and jumping up from his chair, John proclaimed: “We can phase that!”

This accurate historical account triggered the work by Spence et al. [2] that outlined a method to reconstruct the electron density of a molecule directly from the diffraction data. The central idea is that the diffracted intensities between Bragg peaks provide sufficient sampling of the molecular transform to solve the crystallographic phase problem [3]. In the case of the experiment from 2009, these inter-Bragg intensities are due to the Fourier transform of the shape of the crystal, which resulted in the approach being known as “shape transform phasing”.

The first part of this manuscript introduces shape transform phasing and recounts some of the history and the relevant theory. The second part describes the lattice occupancy approach that provided a viable way of performing shape transform phasing. Applications of the lattice occupancy model to describe various crystal defect scenarios are described in the third part, with simulations presented in 2D. Finally we conclude by outlining the current difficulties of the technique and future directions of the shape transform phasing method.

2. Origins

Shape transforms from protein crystals have been observed by Boutet and Robinson [4] using synchrotrons. These signals have since become prominent in X-ray free-electron laser (XFEL) experiments termed “serial femtosecond crystallography” because smaller crystals can be used [5]. In serial femtosecond crystallography, diffraction

* Corresponding author at: Department of Electrical and Computer Engineering, University of Canterbury, Christchurch 8041, New Zealand.

E-mail address: joe.chen@canterbury.ac.nz (J.P.J. Chen).

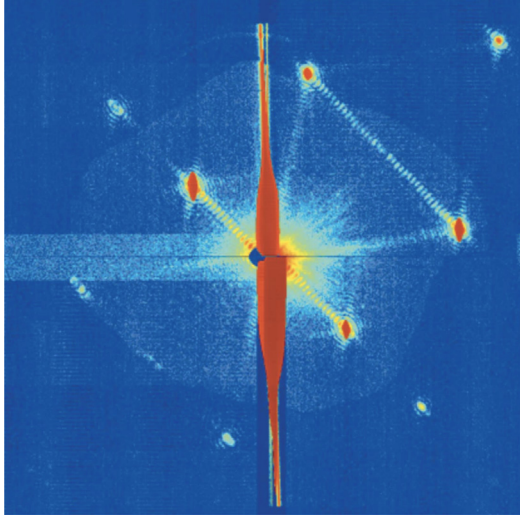


Fig. 1. A single diffraction pattern from a sub-micron-sized crystal of Photosystem I obtained at the LCLS [1].

patterns from crystals of different sizes and shapes are injected into the X-ray pulse in a serial fashion. The real-space scattering density of the n th crystal, $g_n(\mathbf{r})$, can be written in terms of the scattering density of the unit molecule, $f(\mathbf{r})$, in the crystal basis as

$$g_n(\mathbf{r}) = \sum_{\alpha} f(\mathbf{r} - \mathbf{r}_{\alpha}) \quad (1)$$

where the integer tuple $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ indexes the unit cells and the position of those unit cells are given by the lattice vectors $\mathbf{r}_{\alpha} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3$, a linear combination of the crystal basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. The sum is taken over the set of α tuples that describes the crystal, which we denote here by S_n .

Illuminating the entire crystal coherently with the X-ray beam, the diffraction pattern in the far-field is given by the Fourier transform of the electron density of the unit cell. The Fourier transform of the crystal density is equal to

$$\begin{aligned} G_n(\mathbf{q}) &= \sum_{\alpha} \int_{-\infty}^{\infty} d^3r e^{i\mathbf{q} \cdot \mathbf{r}} f(\mathbf{r} - \mathbf{r}_{\alpha}) \\ &= \sum_{\alpha} F(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}_{\alpha}} \\ &= F(\mathbf{q}) \sum_{\alpha} e^{i\mathbf{q} \cdot \mathbf{r}_{\alpha}} \end{aligned} \quad (2)$$

The intensity is then

$$I_n(\mathbf{q}) = |G_n(\mathbf{q})|^2 = |F(\mathbf{q})|^2 \left| \sum_{\alpha} e^{i\mathbf{q} \cdot \mathbf{r}_{\alpha}} \right|^2 \quad (3)$$

which for a cuboidal crystal, of sides (N_x, N_y, N_z) , the summation in Eq. (3) can be evaluated to reduced the equation to

$$I_n(\mathbf{q}) = |F(\mathbf{q})|^2 \frac{\sin^2(N_x q_x)}{\sin^2(q_x)} \frac{\sin^2(N_y q_y)}{\sin^2(q_y)} \frac{\sin^2(N_z q_z)}{\sin^2(q_z)} \quad (4)$$

The magnitude squared of the Fourier transform of the molecule $|F(\mathbf{q})|^2$ modulated by the $\sin^2 / \sin^2$ function is visualised on the left hand side of Fig. 2 in 1D, for different values of N and explains the diffraction pattern observed in Fig. 1 where it is possible to count the number of unit cells!

In serial femtosecond crystallography, the diffraction patterns from crystals of different sizes and shapes in the same orientation are averaged to increase the signal-to-noise ratio [7,8]. The average intensity over many crystals is

$$I(\mathbf{q}) = \left\langle |G_n(\mathbf{q})|^2 \right\rangle_n = |F(\mathbf{q})|^2 \left\langle \left| \sum_{\alpha} e^{i\mathbf{q} \cdot \mathbf{r}_{\alpha}} \right|^2 \right\rangle_n \quad (5)$$

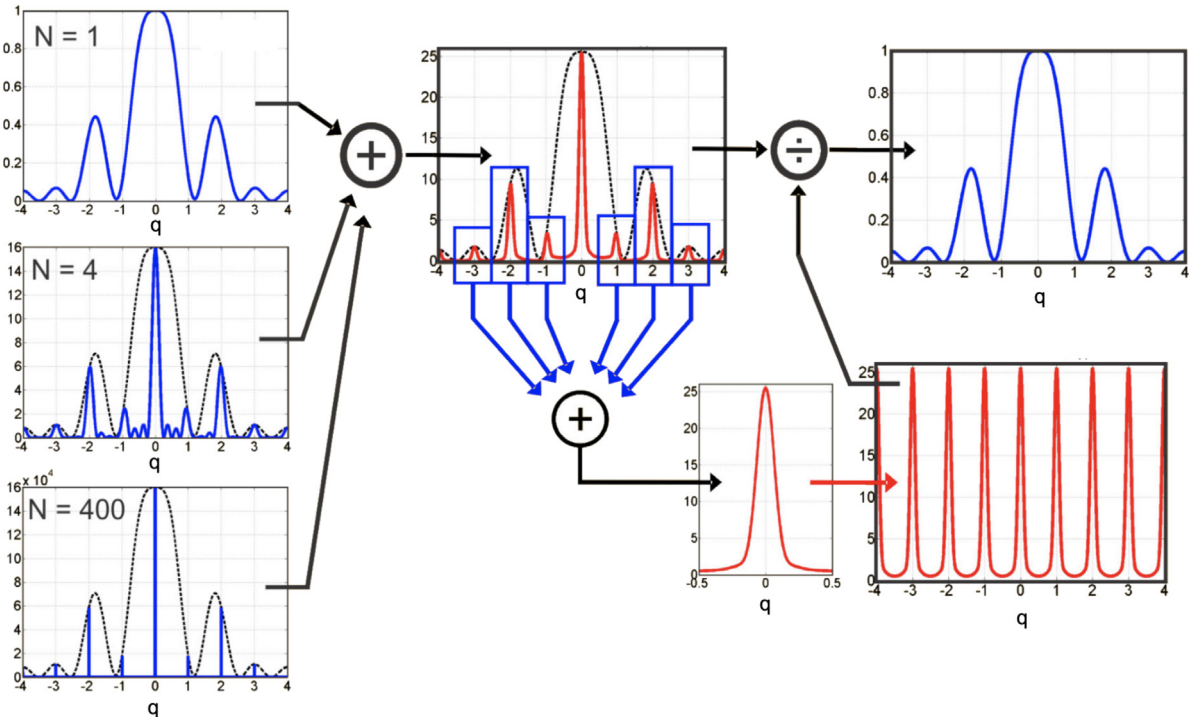


Fig. 2. The process of dividing out the shape transform first proposed by Spence et al. [2] illustrated in 1D. The N on the left-hand-side denotes the number of unit cells in each 1D crystal. The blue rectangular boxes indicate the Wigner-Seitz cells.

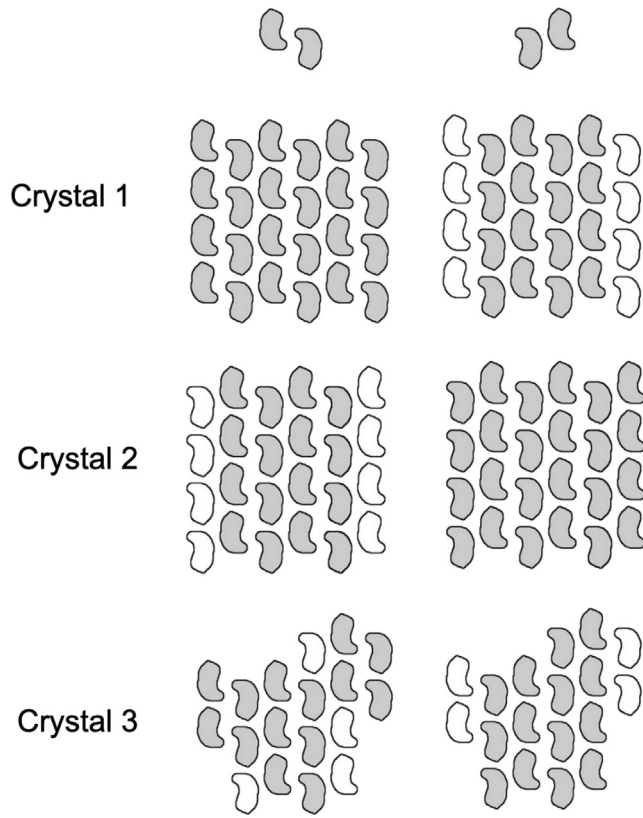


Fig. 3. Two types of unit cells (top row) and three different crystals composed of the same molecular building blocks but terminated in different ways. The shaded molecules indicate the asymmetric units that can be described by the unit cell at the top of each column. The un-shaded molecules are what are known as “partial” or “incomplete” unit cells.

which for the cuboidal case simplifies to

$$I(q) = |F(q)|^2 \left\langle \frac{\sin^2(N_{xn}q_x)}{\sin^2(q_x)} \frac{\sin^2(N_{yn}q_y)}{\sin^2(q_y)} \frac{\sin^2(N_{zn}q_z)}{\sin^2(q_z)} \right\rangle_n \quad (6)$$

The expectation value for an individual $\sin^2 / \sin^2$ function was studied by [9] where they showed that the minimum value of this mean is 1/2 for any well-behaved probability density function describing the distribution of the number of unit cells in the crystal ensemble.

Spence et al. [2] proceed to show that you can estimate the averaged shape transform by summing over many Wigner-Seitz cells. Upon dividing out the estimated shape transform one can recover the Fourier magnitude of one molecule which is then able to be phased with phase retrieval methods from single particle imaging [10]. The whole process is depicted in Fig. 2. Chen et al. [11] studied the phasing process when dividing out the shape transform in the presence of noise.

3. Non-uniqueness of unit cell

A problem with the shape transform phasing method as proposed by Spence et al. [2] gradually emerged. The problem was that when a *finite* crystal is composed of molecules in more than one orientations, i.e., when the crystal has a space group other than P1, the unit cell does not have a unique definition. Different crystal truncations will give rise to crystals described by different unit cells, as depicted in Fig. 3.

Elser [12] first realised this problem and proposes the idea of unwrapping the electron density, along side a phasing method utilising the gradients of the molecular transform.

Richard Bean (unpublished) looked at estimating the intensities mid-way between the Bragg reflections by interpolation and then sorting the diffraction patterns to disentangle the different molecular transforms.

Chen and Millane [13,14], Kirian et al. [15], Liu et al. [16], Millane and Chen [17], Williams et al. [18] all analysed the effects of partial unit cells and made contributions towards a deeper understanding of the problem.

Kirian et al. [19] carried out the first experimental demonstration of shape transform phasing, utilising artificial 2D crystals on silicon nitride membranes. They showed that shape transform phasing could work even in the presence of unknown and jittery wavefront, but also demonstrated the non-uniqueness of the unit cell problem by experimentally fabricating crystals with four different kinds of unit cells, and showing that their molecular transforms differed greatly away from the Bragg locations.

Chen et al. [20] provided a partial solution to the unit cell non-uniqueness problem by coming up with an algorithm that could phase the intensities of a weighted sum of different molecular transforms. However that algorithm could not deal with the multitude of unit cells that would be generated from arbitrary crystal truncations.

It was not until 2019 that the problem of non-uniqueness of unit cell in shape transform phasing was fully resolved, leading to the lattice occupancy model of a crystal.

4. The occupancy model

The lattice occupancy model proposed by Chen et al. [6] describes a crystal by laying down infinite lattices and multiplying the electron density of the molecule by a 0 when the molecule does not exist on a particular lattice point and a 1 when the molecule does exist on that lattice point. We outline the mathematics associated with this model in this section, and write down the averaged diffracted intensity of a crystal in this formalism.

The real-space scattering density of the n th crystal, $g_n(\mathbf{r})$, can be written in terms of the scattering density of the k th asymmetric unit, $f_k(\mathbf{r})$, as

$$g_n(\mathbf{r}) = \sum_{\alpha} \sum_{k=1}^K w_{\alpha kn} f_k(\mathbf{r} - \mathbf{r}_{\alpha} - \mathbf{s}_k) \quad (7)$$

where $w_{\alpha kn}$ is the occupancy of the k th asymmetric unit in the unit cell indexed by the 3-tuple of integers $\alpha = (\alpha_1, \alpha_2, \alpha_3)$, $\mathbf{s}_k$ is the constant shift of the k th asymmetric unit as determined by the space group of the crystal.

Again, the lattice vectors $\mathbf{r}_{\alpha}$ are given by a linear combination of the three crystal basis vectors as $\mathbf{r}_{\alpha} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \alpha_3 \mathbf{a}_3$. The weights $w_{\alpha kn}$ take on values of either 0 or 1 when interpreted as a molecular occupancy but can more generally be any complex number to model, say, the partial coherence of the incident X-ray wavefront. For the partial coherence modelling to work, we assume the molecule is much smaller than the fluctuations of the wavefront of the beam and so when each molecule gets a different complex weight, this can potentially be interpreted as piece-wise approximations to a partially-coherent incident wavefront illuminating a volume of molecules that are all weighted equally.

The Fourier transform of the crystal density is equal to

$$\begin{aligned} G_n(q) &= \sum_{\alpha} \sum_{k=1}^K w_{\alpha kn} \int_{-\infty}^{\infty} d^3r e^{-iq \cdot r} f_k(\mathbf{r} - \mathbf{r}_{\alpha} - \mathbf{s}_k) \\ &= \sum_{\alpha} \sum_{k=1}^K w_{\alpha kn} F_k(q) e^{-iq \cdot \mathbf{r}_{\alpha}} e^{-iq \cdot \mathbf{s}_k}. \end{aligned} \quad (8)$$

The intensity for the n th crystal is then

$$I_n(q) = |G_n(q)|^2 = \sum_{\alpha\beta} \sum_{kl} w_{\alpha kn} w_{\beta ln} F_l^*(q) F_k(q) e^{-iq \cdot (\mathbf{r}_{\alpha} - \mathbf{r}_{\beta})} e^{-iq \cdot (\mathbf{s}_k - \mathbf{s}_l)}, \quad (9)$$

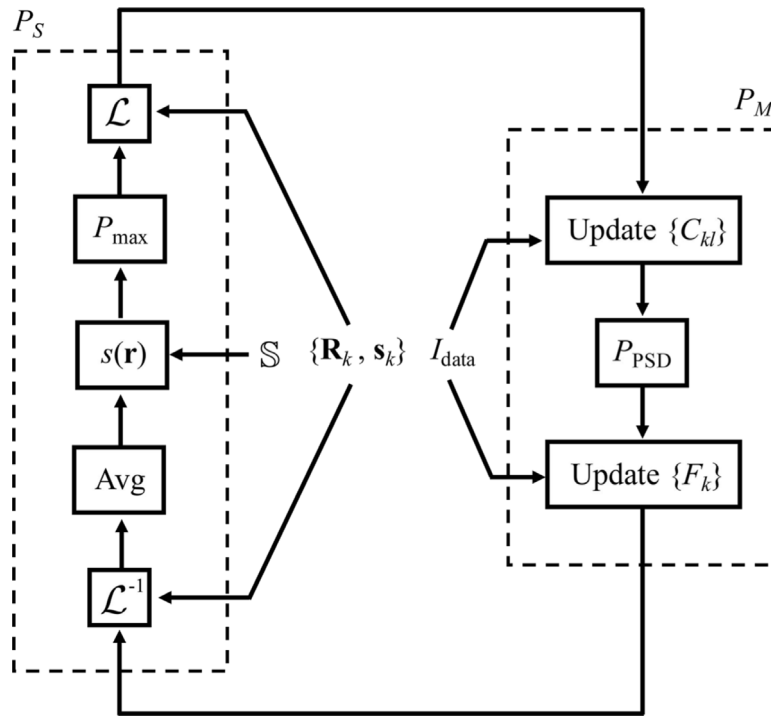


Fig. 4. Flow diagram of the error-reduction form of the algorithm, adapted from Chen et al. [6]. $\mathcal{L}$ denotes symmetry operations on the molecule due to the space group (rotations $\mathbf{R}_k$ and translations $\mathbf{s}_k$), $s(\mathbf{r})$ is the support of the molecule in the position basis, $\mathbb{S}$ contains the information about the support, I_{data} is the averaged diffracted intensity from the edgy crystal ensemble, P_{max} and P_{PSD} are described in the main text.

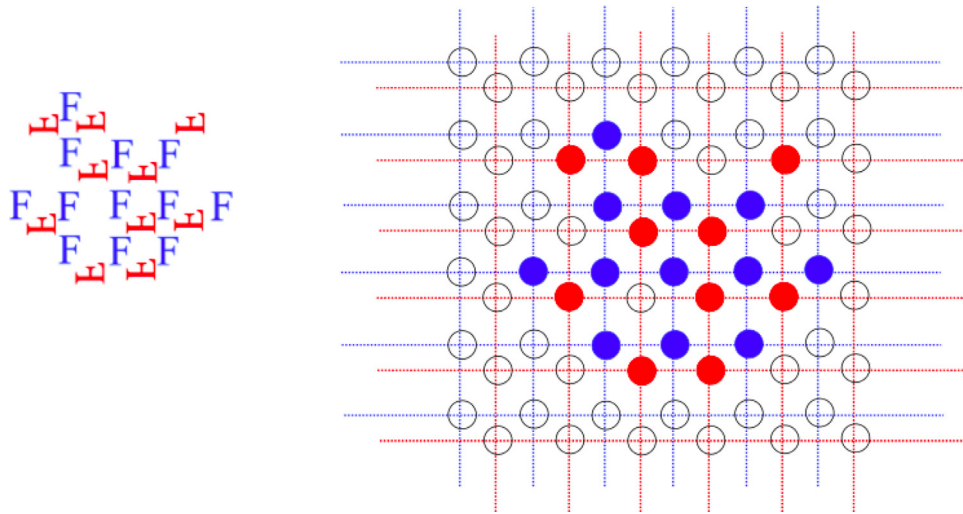


Fig. 5. Edgy finite crystal and its occupancy model. Filled circle denote $w_{akn} = 1$ and empty circles $w_{akn} = 0$.

and the average intensity over many crystals is

$$I(q) = \langle |G_n(q)|^2 \rangle_n = \sum_{\alpha\beta} e^{-iq \cdot (r_\alpha - r_\beta)} \sum_{kl} \langle w_{akn} w_{\beta ln} \rangle_n F_l^*(q) F_k(q) e^{-iq \cdot (s_k - s_l)} \\ = \sum_{\alpha\beta} e^{-iq \cdot (r_\alpha - r_\beta)} \sum_{kl} W_{\alpha k \beta l} F_l^*(q) F_k(q) e^{-iq \cdot (s_k - s_l)}, \quad (10)$$

where the term

$$W_{\alpha k \beta l} \equiv \langle w_{akn} w_{\beta ln} \rangle_n \quad (11)$$

is referred to as the “occupancy matrix” or the “crystal covariance matrix”.

Now define the averaged lattice-transform functions as

$$C_{kl}(q) \equiv \sum_{\alpha\beta} W_{\alpha k \beta l} e^{-iq \cdot (r_\alpha - r_\beta)} \quad (12)$$

then the averaged intensity can be written as

$$I(q) = \langle I_n(q) \rangle_n = \sum_{kl} C_{kl}(q) F_k(q) F_l^*(q) e^{-iq \cdot (s_k - s_l)}. \quad (13)$$

Eq. (13) can be written in matrix notation as

$$I(q) = \mathbf{F}^\dagger(q) \mathbf{C}(q) \mathbf{F}(q), \quad (14)$$

where $\mathbf{F}(q)$ is a $K \times 1$ array and $\mathbf{C}(q)$ is a $K \times K$ array containing the complex amplitudes of the molecular transform and averaged lattice-transform, respectively. The $\dagger$ denotes complex conjugation transpose.

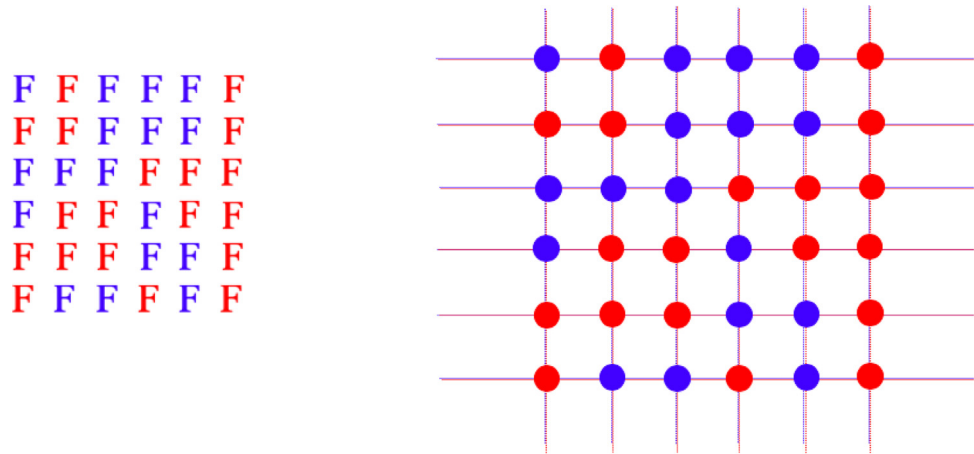


Fig. 6. Crystal with substitutional disorder and its occupancy model. Filled circle denote $w_{akn} = 1$.

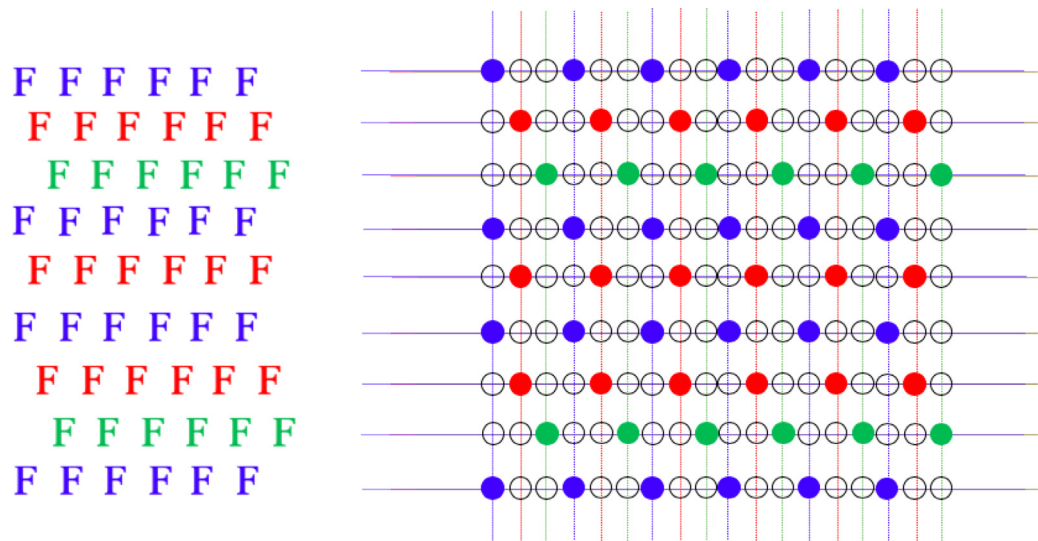


Fig. 7. Crystal with ABC stacking fault and its occupancy model. Filled circle denote $w_{akn} = 1$ and empty circles $w_{akn} = 0$.

		Lattice	
		Same	Different
Molecular density	Same	Perfect crystals	Stacking fault
	Different	Substitutional disorder	Edgy crystals

Fig. 8. A summary table of four different crystal scenarios.

The $C(q)$ matrix is positive-semidefinite and periodic, with a period given by the Wigner–Seitz cells, as shown by Chen et al. [6]. These properties will be utilised in the reconstruction algorithm to recover the averaged lattice transform, along with the molecular density, as described in the next section.

5. Reconstruction algorithm

Once we have an equation describing the forward problem, we can proceed to find a method to back-out the quantities inputted into

our model, to solve the inverse problem. Specifically, we would like to recover the molecular density $f(r)$. The reconstruction algorithm proposed by Chen et al. [6] takes the averaged intensity $I(q)$ and recovers both $C_{kl}(q)$ and $f(r)$ for all k and l . Briefly, the algorithm uses the framework of iterative projection algorithms, detailed by Fienup [21],Bauschke et al. [22],Elser [23,24],Luke [25],Marchesini [26],Elser et al. [27],Millane and Lo [28], where the problem is posed in terms of the satisfaction of constraints, and projection operators are utilised to move an initial guess towards the intersection of the constraints. In the case of the algorithm proposed by Chen et al. [6], the iterate

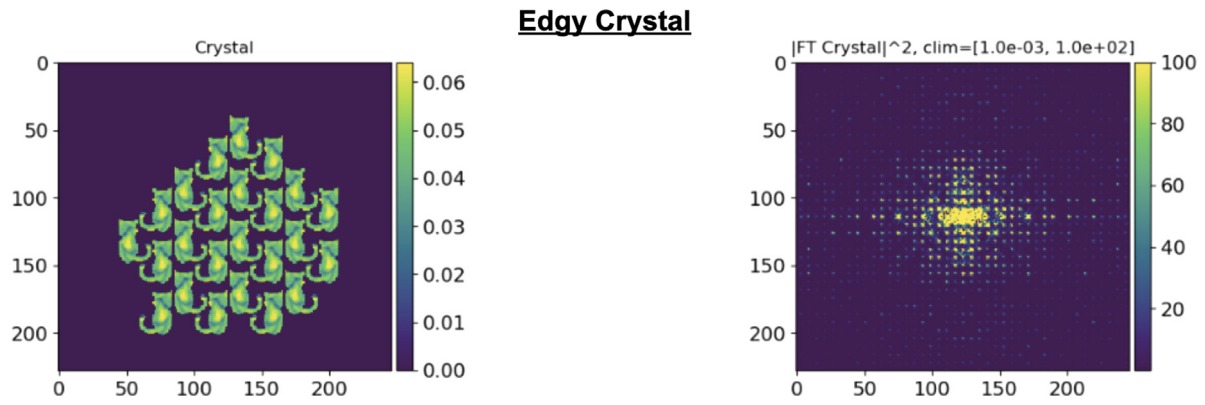


Fig. 9. Diffracted intensity data, and the crystal that it is generated from for the edgy crystal reconstruction.

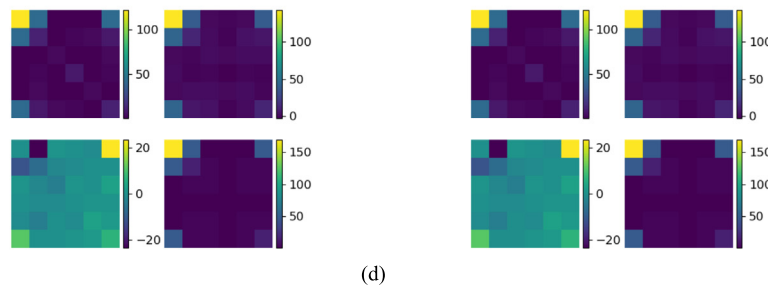
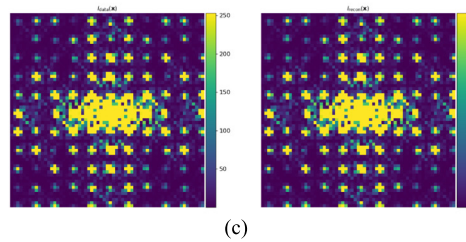
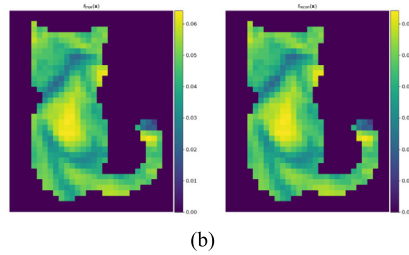
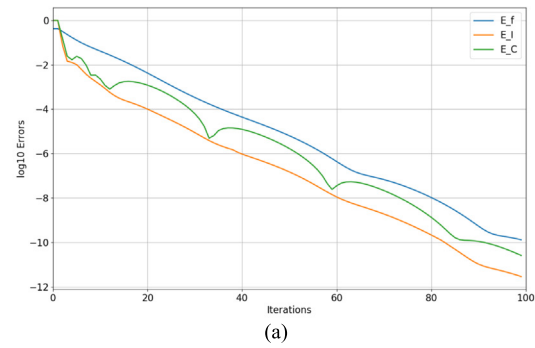


Fig. 10. Edgy crystal reconstruction. Ground truths are on the left column and reconstructions on the right. (a) The root-mean-squared errors for the real-space molecular density, E_f , diffracted intensity data, E_I , and averaged lattice transform matrix, E_C , (b) Real-space density. (c) Zoomed-in diffracted intensity. (d) A single period of the averaged lattice transform functions, arranged on a K by K array such that the on-diagonal images are $C_{kk}(q)$, upper triangular images are $\text{Re}(C_{kl}(q))$ and bottom triangular images are $\text{Im}(C_{kl}(q))$.

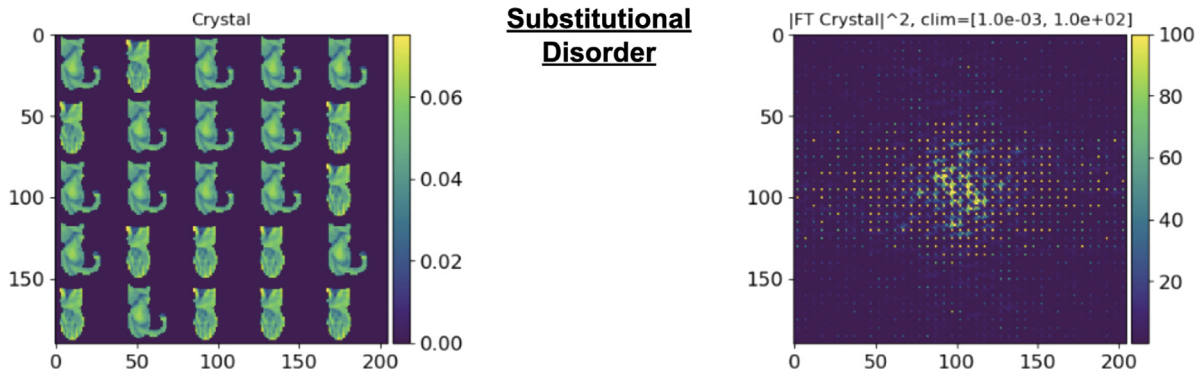


Fig. 11. Diffracted intensity data, and the crystal that it is generated from, for the substitutional disorder reconstruction.

consists of the set of all symmetry partners of the molecule $\{f_k(\mathbf{r})\}$ and the averaged lattice transform matrices $\{C_{kl}(\mathbf{q})\}$, and the two projection operators, P_S and P_M , are formulated as follows:

P_S consists of first (1) undoing the symmetry operators on each of the $f_k(\mathbf{r})$ to map the density to a common molecular orientation, (2) average the result, (3) apply the support constraint, (4) normalise the densities ($P_{\max}$) so that its root-mean-squared value is 1 for numerical stability, and (5) applying the symmetry mapping to the updated common molecule to redistribute copies of the molecule to form a set of new $\{f_k(\mathbf{r})\}$.

P_M consists of (1) updating the $\{C_{kl}(\mathbf{q})\}$ through a least-squares optimisation on the common periods of the matrix while holding $\{f_k(\mathbf{r})\}$ constant, (2) changing each $\{C_{kl}(\mathbf{q})\}$ matrices so that they are positive semi-definite with the projection operator P_{PSD} , and then (3) updating the Fourier transform of the $\{f_k(\mathbf{r})\}$ while holding $\{C_{kl}(\mathbf{q})\}$ constant through a projection onto an ellipsoid [29].

A flow diagram of the error-reduction implementation of the algorithm utilising these projections is depicted in Fig. 4. Details of the operators and the entire process can be found in Chen et al. [6].

6. Applications to other situations

The lattice occupancy formalism can be applied to model a number of crystal defects, which we summarise in this section.

6.1. Edgy finite crystals

To describe finite edgy crystals, as is needed for shape transform phasing, every lattice point where the molecule does not exist, we set the weight at that lattice point, w_{akn} , to 0, otherwise we set the weight to 1, indicating the presence of a molecule. An illustration of this description is shown in Fig. 5.

6.2. Substitutional disorder

Substitutional disorders can be modelled by using overlapping lattices for all molecule types, which is equivalent to saying $s_k = 0$ for all k . For example, if there are three different molecular types, $K = 3$, and at every lattice point, either w_{a1n} or w_{a2n} or w_{a3n} will be 1, the rest will be 0. An illustration of this description is shown in Fig. 6.

6.3. Stacking faults

ABC stacking faults can be modelled by three separate lattices, each of which determines the occurrence of the same molecular density, i.e., $f_k(\mathbf{x}) = f(\mathbf{x})$ for all k . An illustration of this description is shown in Fig. 7.

6.4. Summary of different defects

The cases described in the previous section can be summarised by the table in Fig. 8. When all molecules are the same, but the lattices are different, we have stacking faults. When all lattices are the same, but the molecules are different, we have substitutional disorder. When both molecule densities and lattices are different we get edgy crystals.

6.5. Uniqueness

As was pointed out in Section 3.3.4 in the paper by Chen et al. [6], a degeneracy exists that prevents the unique recovery of $\{C_{kl}(\mathbf{q})\}$. This degeneracy can occur when at least two of the molecular transforms are equal. This is because we can write the intensity from Eq. (13) as

$$\begin{aligned} I(\mathbf{q}) &= \sum_{k=1}^K \sum_{l=1}^K F_l^*(\mathbf{q}) C_{kl}(\mathbf{q}) F_k(\mathbf{q}) \\ &= \sum_{k=1}^K |F_k(\mathbf{q})|^2 C_{kk}(\mathbf{q}) + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K F_l^*(\mathbf{q}) C_{kl}(\mathbf{q}) F_k(\mathbf{q}). \end{aligned} \quad (15)$$

Assume all of the molecular transforms are the same, i.e., $|F_k(\mathbf{q})|^2 = |F(\mathbf{q})|^2$ for all k , then we can factor out the molecular transform from the first term, leaving

$$I(\mathbf{q}) = |F(\mathbf{q})|^2 \sum_{k=1}^K C_{kk}(\mathbf{q}) + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K F_l^*(\mathbf{q}) C_{kl}(\mathbf{q}) F_k(\mathbf{q}). \quad (16)$$

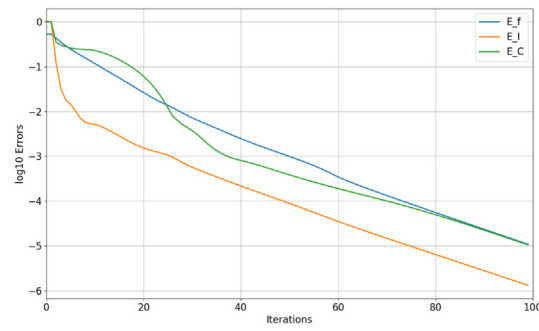
Thus we cannot distinguish between the different $C_{kk}(\mathbf{q})$ and can only recover up to their sum. An even more significant situation is the case for our occupancy model of the ABC stacking fault where $f_k(\mathbf{x}) = f(\mathbf{x})$ for all k , giving

$$\begin{aligned} I(\mathbf{q}) &= |F(\mathbf{q})|^2 \sum_{k=1}^K \sum_{l=1}^K C_{kl}(\mathbf{q}) \\ &= |F(\mathbf{q})|^2 \left(\sum_{k=1}^K C_{kk}(\mathbf{q}) + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq k}}^K C_{kl}(\mathbf{q}) \right) \\ &= |F(\mathbf{q})|^2 \left(\sum_{k=1}^K C_{kk}(\mathbf{q}) + 2 \sum_{k=1}^K \sum_{\substack{l=1 \\ l > k}}^K \text{Re}(C_{kl}(\mathbf{q})) \right) \end{aligned} \quad (17)$$

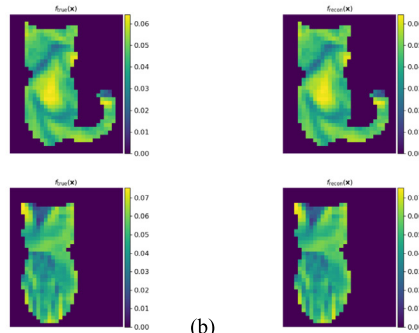
since $C_{kl}(\mathbf{q}) = C_{lk}^*(\mathbf{q})$. Thus, for the case of the ABC stacking fault, we can only recover up to the sum of $C_{kk}(\mathbf{q})$ and the sum of the real parts of $C_{kl}(\mathbf{q})$, with the imaginary parts of $C_{kl}(\mathbf{q})$ left unconstrained.

7. Simulations

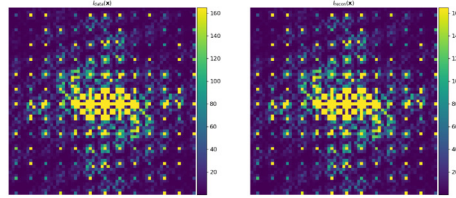
In this section we show some phasing results for diffraction patterns from single 2D crystals for each of the three aforementioned defect



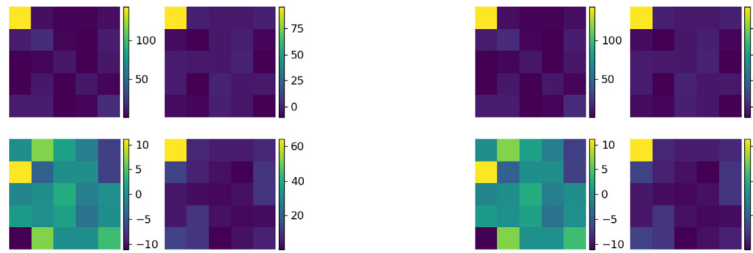
(a)



(b)



(c)



(d)

Fig. 12. Substitutional disorder reconstruction. Ground truths are on the left column and reconstructions on the right. (a) The root-mean-squared errors for the real-space molecular density, E_f , diffracted intensity data, E_I , and averaged lattice transform matrix, E_C , (b) Real-space density. (c) Zoomed-in diffracted intensity. (d) A single period of the averaged lattice transform functions, arranged on a K by K array such that the on-diagonal images are $C_{kk}(q)$, upper triangular images are $\text{Re}(C_{kl}(q))$ and bottom triangular images are $\text{Im}(C_{kl}(q))$.

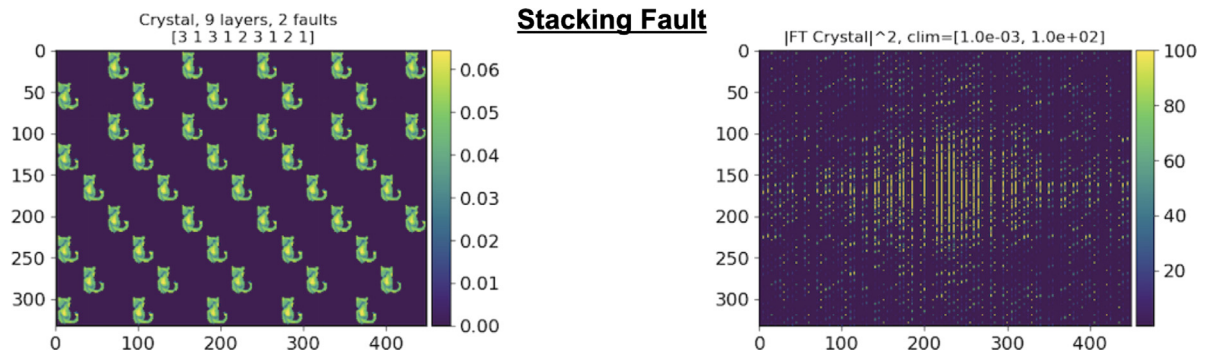


Fig. 13. Diffracted intensity data, and the crystal that it is generated from for the stacking fault reconstruction.

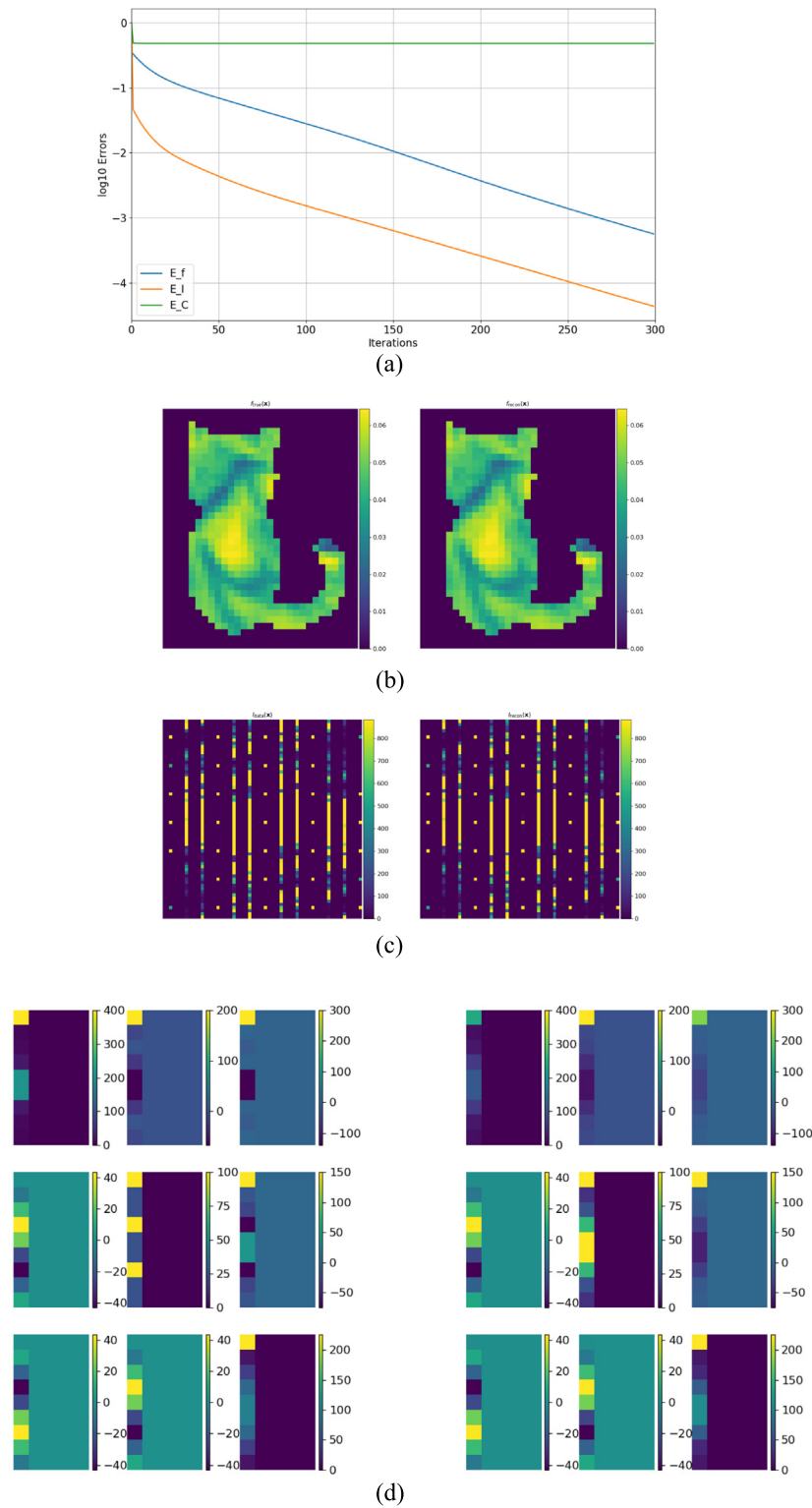


Fig. 14. Stacking fault reconstruction. Ground truths are on the left column and reconstructions on the right. (a) The root-mean-squared errors for the real-space molecular density, E_f , diffracted intensity data, E_I , and averaged lattice transform matrix, E_C . (b) Real-space density. (c) Zoomed-in diffracted intensity. (d) A single period of the averaged lattice transform functions, arranged on a K by K array such that the on-diagonal images are $C_{kk}(q)$, upper triangular images are $\text{Re}(C_{kl}(q))$ and bottom triangular images are $\text{Im}(C_{kl}(q))$.

cases, i.e., there is only a single crystal in the ensemble for each of the cases (following the quick-and-dirty, emphasis-on-clear-illustrations-and-ideas spirit of John!). The three cases are: (1) edge crystal (Figs. 9, 10), (2) substitutional disorder (Figs. 11, 12), (3) ABC stacking fault (Figs. 13, 14). The result figures for the three cases are displayed in

pairs, the first figure for the input quantities and the second figure for the reconstruction results.

The Relaxed Averaged Alternating Reflections (RAAR) algorithm [25] was used, with $\beta = 0.7$ for the edge crystal and substitutional disorder reconstructions. For the ABC stacking fault case, the error-reduction

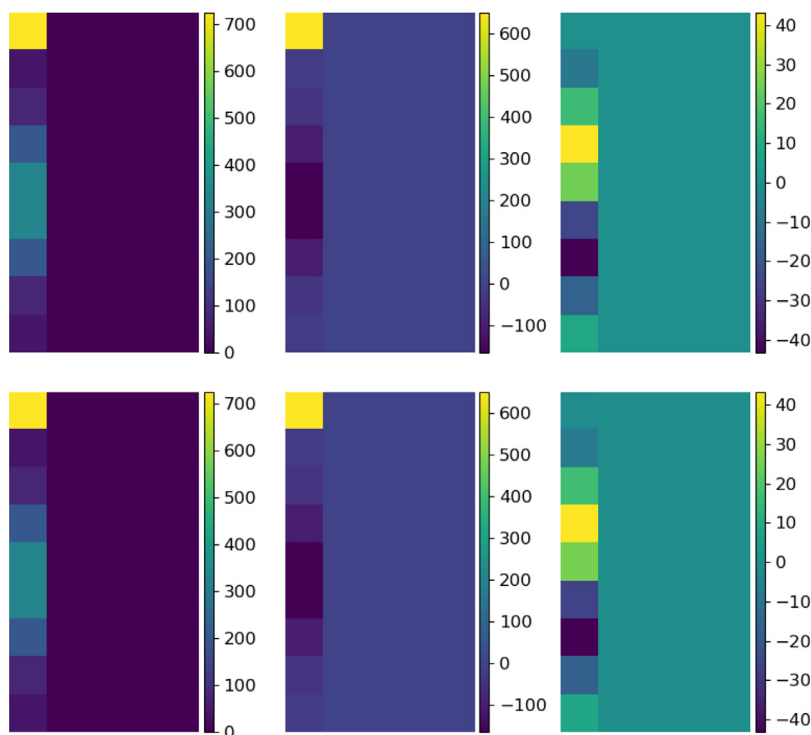


Fig. 15. The summed quantities (Left) $\sum_k C_{kk}(q)$, (Middle) $\sum_{kl} \text{Re}(C_{kl}(q))$, (Right) $\sum_{kl} \text{Im}(C_{kl}(q))$ for the ABC stacking fault case. (Top row) Ground truth, (Bottom row) reconstruction.

algorithm was applied as that algorithm was sufficient to produce the ambiguity outlined in Section 6.5 that prevents the reconstruction from converging to a unique solution. An intensity mask was also applied as there are many zeros in the intensity data of the single diffraction pattern. The mask allowed the pixel value in Fourier space to float when the intensity of that pixel is below a chosen threshold of 10^{-5} .

The root-mean-squared (RMS) reconstruction errors shown in Figs. 10(a), 12(a) and 14(a) are calculated for the real space molecular density, E_f , diffracted intensity data, E_I , and averaged lattice transform matrix, E_C .

Note that E_C in Fig. 14(a) stays very high. This is due to the imaginary parts of $C_{kl}(q)$ being unconstrained as shown in Section 6.5. From Fig. 15, we can see that the sum, $\sum_k C_{kk}(q)$, and the sum of the real parts of the lattice transform functions $\sum_{kl} \text{Re}(C_{kl}(q))$, are indeed the same as the ground truth. The sum of the imaginary parts, $\sum_{kl} \text{Im}(C_{kl}(q))$ also looks visually the same as the ground truth, but the actual numerical error for that is in fact a bit higher: The RMS errors for the case $\sum_k C_{kk}(q)$, $\sum_{kl} \text{Re}(C_{kl}(q))$, and $\sum_{kl} \text{Im}(C_{kl}(q))$ are 3.86×10^{-5} , 5.82×10^{-5} , and 3.35×10^{-2} , respectively, as compared to the ground truth, illustrating the fact that the imaginary parts are indeed unconstrained, as shown in Section 6.5.

8. Conclusion and future outlook

Shape transform phasing is a phasing method that has in principle no resolution restrictions or the need for sample modifications, and thus is an ab initio phasing method that is able to give unbiased reconstructions of molecular structures.

One way to formulate the forward problem in shape transform phasing is through the lattice occupancy model. Appropriate molecules can be turned on or off at the desired lattice points by assigning a weight value to each lattice points — this is the basic idea of the lattice occupancy model and this formalism can be used to describe various disorder. We have considered three types of disorder in this manuscript: (1) random crystal truncations (origin of shape transforms), (2) substitutional disorder, and (3) ABC stacking faults, and successfully tested

the reconstruction algorithm for all of these cases under noiseless conditions in 2D. The reader is referred to Chen et al. [6] for results from 3D simulations under noisy conditions in the random crystal truncation case.

Despite its utility, the lattice occupancy model is still quite rigid and is unable to efficiently describe continuous movements of molecules such as those present in rotational disorder. The future of shape transform phasing, and phasing using disordered structures in general, looks likely to move towards more “fluid” and “dynamic” models incorporating methods such as molecular dynamics (see for example Mazumder and Ayyer [30]) and machine learning approaches. Nevertheless, the shape transform phasing paradigm initiated by John lives on and continues to flourish — just like the memory, the ideas and the inspirational spirit of John himself.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

JPJC acknowledges support from the Marsden Fast-Start fund. RAK acknowledges support from NSF Award DBI-1943448. JPJC, KES and RAK acknowledge support from NSF Award DBI-1231306. We heartily acknowledge the support, guidance, mentorship, and friendship of John C. H. Spence.

References

- [1] H.N. Chapman, et al., Femtosecond X-ray protein nanocrystallography, *Nature* 470 (2011) 73–77.

- [2] J.C.H. Spence, R.A. Kirian, X.Y. Wang, U. Weierstall, K.E. Schmidt, T.A. White, A. Barty, H.N. Chapman, S. Marchesini, J. Holton, Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals, *Opt. Express* 19 (2011) 2866–2873.
- [3] D. Sayre, Some implications of a theorem due to Shannon, *Acta Crystallogr.* 5 (1952) 843.
- [4] S. Boutet, I.K. Robinson, Coherent X-Ray diffractive imaging of protein crystals, *J. Synchrotron Radiat.* 15 (2008) 576–583.
- [5] J.C.H. Spence, XFELs for structure and dynamics in biology, *IUCrJ* 4 (2017) 322–339.
- [6] J.P.J. Chen, J.J. Donatelli, K.E. Schmidt, R.A. Kirian, Shape transform phasing of edgy nanocrystals, *Acta Crystallogr. Sect. A* 75 (2) (2019) 239–259.
- [7] R. Grosse-Kunstleve, N. Sauter, N. Moriarty, P. Adams, The computational crystallography toolbox: crystallographic algorithms in a reusable software framework, *J. Appl. Crystallogr.* 35 (1) (2002) 126–136.
- [8] T.A. White, R.A. Kirian, A.V. Martin, A. Aquila, K. Nass, A. Barty, H.N. Chapman, CrystFEL: a software suite for snapshot serial crystallography, *J. Appl. Crystallogr.* 45 (2) (2012) 335–341.
- [9] J.P.J. Chen, J.C.H. Spence, R.P. Millane, Direct phasing in femtosecond nanocrystallography. I. Diffraction characteristics, *Acta Cryst. A* 70 (2014) 143–153.
- [10] J.C.H. Spence, Diffractive imaging of single particles, in: *Springer Handbook of Microscopy*, Springer, 2019, pp. 1009–1036.
- [11] J.P.J. Chen, J.C.H. Spence, R.P. Millane, Direct phasing in femtosecond nanocrystallography. II. Phaserecovery, *Acta Cryst. A* 70 (2014) 154–161.
- [12] V. Elser, Direct phasing of nanocrystal diffraction, *Acta Cryst. A* 69 (2013) 559–569.
- [13] J.P.J. Chen, R.P. Millane, Diffraction by nanocrystals, *J. Opt. Soc. Amer. A* 30 (2013) 2627–2634.
- [14] J.P.J. Chen, R.P. Millane, Diffraction by nanocrystals II, *J. Opt. Soc. Amer. A* 31 (2014) 1730–1737.
- [15] R.A. Kirian, R.J. Bean, K.R. Beyerlein, O.M. Yefanov, T.A. White, A. Barty, H.N. Chapman, Phasing coherently illuminated nanocrystals bounded by partial unit cells, *Philos. Trans. R. Soc. B* 369 (2014) 20130331.
- [16] H. Liu, N.A. Zatsepin, J.C.H. Spence, Ab-initio phasing using nanocrystal shape transforms with incomplete unit cells, *IUCrJ* 1 (2014) 19–27.
- [17] R.P. Millane, J.P.J. Chen, Aspects of direct phasing in femtosecond nanocrystallography, *Philos. Trans. R. Soc. B* 369 (2014) 20130498.
- [18] S.R. Williams, R.A. Dilanian, H.M. Quiney, A.V. Martin, Analysis of diffracted intensities from finite protein crystals with incomplete unit cells, *Crystals* 7 (2017) 220.
- [19] R.A. Kirian, R.J. Bean, K.R. Beyerlein, M. Barthelmess, C.H. Yoon, F. Wang, F. Capotondi, E. Pedersoli, A. Barty, H.N. Chapman, Direct phasing of finite crystals illuminated with a free-electron laser, *Phys. Rev. X* 5 (2015) 011015.
- [20] J.P.J. Chen, R.D. Arnal, A.J. Morgan, R.J. Bean, K.R. Beyerlein, H.N. Chapman, P.J. Bones, R.P. Millane, R.A. Kirian, Reconstruction of an object from diffraction intensities averaged over multiple object clusters, *J. Opt.* 18 (11) (2016) 114003.
- [21] J.R. Fienup, Phase retrieval algorithms: a comparison, *Appl. Opt.* 21 (1982) 2758–2769.
- [22] H.H. Bauschke, P.L. Combettes, D.R. Luke, Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization, *J. Opt. Soc. Amer. A* 19 (7) (2002) 1334–1345.
- [23] V. Elser, Phase retrieval by iterated projections, *J. Opt. Soc. Amer. A* 20 (2003) 40–55.
- [24] V. Elser, Random projections and the optimization of an algorithm for phase retrieval, *J. Phys. A: Math. Gen.* 36 (2003) 2995–3007.
- [25] D.R. Luke, Relaxed averaged alternating reflections for diffraction imaging, *Inverse Problems* 21 (2005) 37–50.
- [26] S. Marchesini, A unified evaluation of iterative projection algorithms for phase retrieval, *Rev. Sci. Instrum.* 78 (2007) 001301.
- [27] V. Elser, I. Rankenburg, P. Thibault, Searching with iterated maps, *Proc. Natl. Acad. Sci. USA* 104 (2) (2007) 418–423.
- [28] R.P. Millane, V.L. Lo, Iterative projection algorithms in protein crystallography. I. theory, *Acta Cryst. A* 69 (2013) 517–527.
- [29] A.J. Morgan, K. Ayyer, A. Barty, J.P.J. Chen, T. Ekeberg, D. Oberthuer, T.A. White, O. Yefanov, H.N. Chapman, Ab initio phasing of the diffraction of crystals with translational disorder, *Acta Cryst. A* 75 (2019) 25–40.
- [30] P. Mazumder, K. Ayyer, Understanding conformational dynamics from macromolecular crystal diffuse scattering, 2021, *BioRxiv*.