High-resolution structure of viruses from random diffraction snapshots

A. Hosseinzadeh, P. Schwander, A. Dashti, R. Fung, R. M. D'Souza and A. Ourmazd

*Phil. Trans. R. Soc. B* 2014 **369**, 20130326, published 9 June 2014

**Supplementary data**

"Data Supplement"
http://rstb.royalsocietypublishing.org/content/suppl/2014/06/03/rstb.2013.0326.DC1.html

"Audio Supplement"
http://rstb.royalsocietypublishing.org/content/suppl/2014/06/09/rstb.2013.0326.DC2.html

**References**

This article cites 27 articles, 5 of which can be accessed free
http://rstb.royalsocietypublishing.org/content/369/1647/20130326.full.html#ref-list-1

This article is free to access

**Subject collections**

Articles on similar topics can be found in the following collections

- structural biology (52 articles)
- theoretical biology (63 articles)

**Email alerting service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click here

To subscribe to *Phil. Trans. R. Soc. B* go to: http://rstb.royalsocietypublishing.org/subscriptions
High-resolution structure of viruses from random diffraction snapshots

A. Hosseinizadeh1,†, P. Schwander1,†, A. Dashti2, R. Fung1, R. M. D’Souza2 and A. Ourmazd1

1Department of Physics, and 2Department of Mechanical Engineering, University of Wisconsin Milwaukee, 1900 East Kenwood Boulevard, Milwaukee, WI 53211, USA

The advent of the X-ray free-electron laser (XFEL) has made it possible to record diffraction snapshots of biological entities injected into the X-ray beam before the onset of radiation damage. Algorithmic means must then be used to determine the snapshot orientations and thence the three-dimensional structure of the object. Existing Bayesian approaches are limited in reconstruction resolution typically to 1/10 of the object diameter, with the computational expense increasing as the eighth power of the ratio of diameter to resolution. We present an approach capable of exploiting object symmetries to recover three-dimensional structure to high resolution, and thus reconstruct the structure of the satellite tobacco necrosis virus to atomic level. Our approach offers the highest reconstruction resolution for XFEL snapshots to date and provides a potentially powerful alternative route for analysis of data from crystalline and nano-crystalline objects.

1. Introduction

Ultrashort pulses from X-ray free-electron lasers (XFELs) have recently made it possible to record snapshots before the object is damaged by the intense pulse [1,2]. This has, for example, resulted in de novo determination of protein structure from nano-crystals fabricated in vivo [3]. The ultimate goal, however, remains the determination of the three-dimensional structure of individual proteins and viruses [4] and their conformations [5]. This requires the ability to recover structure from an ensemble of ultralow-signal diffraction snapshots of unknown orientation. The three-dimensional diffracted intensity can then be reconstructed, from which the real-space structure is recovered by iterative phasing algorithms [6–9].

The algorithmic challenge of determining XFEL snapshot orientations was first solved by iterative Bayesian approaches [10,11], which assign an orientation to each snapshot based on maximum likelihood. A key measure of algorithmic performance is computational cost, which determines the range of amenable problems. Orientation recovery methods typically scales as \( R^n \) per iteration, with the magnitude and scaling of the number of iterations unknown [12,13]. For Bayesian algorithms, \( n = 8 \) [6,7,12–14], limiting the amenable resolution to approximately 1/10 of the object diameter [10–12]. At this level, biologically relevant study of almost all interesting objects such as proteins and viruses is out of reach. More recent methods offer improved performance, either by obviating the need for iteration [13], or by improved scaling per iteration, e.g. \( (R^5 \log R) \) [14], though the magnitude and scaling of the number of iterations, where needed, remain unknown.

Despite these developments, computational expense remains a primary challenge. The highest resolution reported to date by methods conforming to the Shannon–Nyquist sampling theorem is approximately 1/30 of the object diameter [13,15]. This is still inadequate for protein assemblies such as viruses. As viruses are expected to scatter more strongly than single molecules, they are under intense study by XFEL methods. High-resolution reconstruction of the three-dimensional structure of a virus from XFEL snapshots thus represents an important milestone in the campaign towards single molecules. For all structure
recovery techniques, the exploitation of symmetry offers an important and hitherto unused weapon in this endeavour.

The Shannon–Nyquist sampling theorem links the resolution \( r \) with which an object can be reconstructed to the number of available snapshots \( N_{\text{snap}} \), the object diameter \( D \) and the number of elements \( N_C \) in the point group of a symmetric object [12]

\[
r = \left( \frac{8 \pi^2}{N_C N_{\text{snap}}} \right)^{1/3} D.
\]  

This equation shows that the presence of symmetry can substantially increase the achievable resolution or reduce the number of snapshots needed to achieve a certain resolution. The current experimental concentration on strongly scattering giant viruses [16] (large \( D \)) and the scarcity of ‘useful’ single-particle snapshots [17] (small \( N_{\text{snap}} \)) make the exploitation of symmetry crucial for further progress. No reconstruction method capable of operating at the signal-to-noise ratios expected from single-molecule diffraction has, to date, exploited object symmetry.

Here, we present an approach capable of determining structure from diffraction snapshots of symmetric objects to 1/100 of the object diameter and demonstrate three-dimensional structure recovery to atomic resolution from simulated noisy snapshots of the satellite tobacco necrosis virus (STNV) at the signal level expected from viruses currently under study at the LCLS X-ray free-electron laser (figure 1). This approach can be applied to symmetric objects of any kind, opening the way to the high-resolution study of a wide variety of crystalline and non-crystalline biological and non-biological entities without radiation damage.

Owing to the superior computational efficiency and hence reconstruction capability of non-iterative manifold approaches [13], we focus on incorporating this capability into these powerful algorithms [13,15,18,19]. In brief, these approaches recognize that scattering ‘maps’ a given object orientation to a diffraction snapshot. The collection of all possible orientations in three-dimensional space spans an SO(3) manifold. Scattering maps this manifold to a topologically equivalent compact manifold in the space spanned by the snapshots. We have shown that, to a good approximation, the manifold formed by the snapshots is endowed with the same metric as that of a ‘symmetric top’, loosely speaking a sphere squashed in the direction of the incident beam due to the effect of projection [13]. Such a manifold is naturally described by the Wigner \( D \)-functions [20], which are intimately related to the elements of the \((3 \times 3)\) rotation matrix [13]. Via so-called empirical orthogonal functions, powerful graph-based algorithms [18,21,22] provide access to the Wigner \( D \)-functions describing manifolds produced by scattering [13], from which the snapshot orientations can be extracted [13,15].

It is the object of this paper to incorporate object symmetry into manifold-based approaches, and thus enable high-resolution structure recovery by XFEL methods. We show that the diffusion map algorithm [21], a theoretically sound (in the sense of guaranteed convergence to eigenfunctions of known operators) and algorithmically powerful (in the sense of intrinsic sparsity) manifold-based approach can be used to recover structure from random snapshots of a symmetric object to high resolution. For concreteness, the discussion is restricted to icosahedral objects, but the approach can be applied to any crystalline or non-crystalline object with symmetry.

The paper is organized as follows. Section 2 outlines our theoretical approach. Specifically, it addresses the construction of eigenfunctions suitable for manifolds produced by scattering from symmetric objects and describes how symmetry-related ambiguities in orientation recovery may be resolved. Section 3 demonstrates structure recovery from simulated diffraction snapshots of a symmetric object to 1/100 of its diameter. For the STNV used as example, this corresponds to atomic resolution. Section 4 places our work in the context of ongoing efforts to determine structure by scattering from single particles and nano-crystals. Section 5 concludes the paper with a brief summary of the implications of our work for structure determination by XFEL techniques. Theoretical and algorithmic details, including pseudocode are presented as the electronic supplementary material.

2. Theoretical approach

We begin by constructing the eigenfunctions needed to describe manifolds produced by scattering from symmetric objects. Diffusion map describes a manifold in terms of the eigenfunctions of the Laplace–Beltrami operator with respect to an unknown metric [13,18]. In the absence of object symmetry, manifolds produced by scattering are well
approximated by a homogeneous metric, with the eigenfunctions of the Laplace–Beltrami operator corresponding to the Wigner D-functions [13,15]. In the presence of object symmetry, appropriately symmetrized eigenfunctions are needed. As shown in the electronic supplementary material, sections A and B, these can be obtained by summing over the Wigner D-functions after operation by the elements of the object point-group, viz.

\[
\tilde{D}_{m,c}^j(\alpha) = \frac{1}{N_G} \sum_{g \in G} O_g D_{m,c}^j(\alpha),
\]

where \(\alpha\) denotes the three numbers collectively representing any rotation, \(N_G\) the number of operations \(O_g\) in the point-group \(G\) and \(D_{m,c}^j(\alpha)\) the (real) Wigner D-functions. This approach is applicable to all point groups. For the icosahedral group, the lowest allowed eigenfunctions consist of 39 non-zero \(D_{m,c}^j(\alpha)\), whose orthogonalization leads to 13 independent icosahedral functions \(\tilde{D}_{m,c}^j(\alpha), (6 \leq m \leq 6)\) (see the electronic supplementary material, section B). These comprise one non-degenerate \((m = 0)\) and six degenerate pairs of eigenfunctions, with the \(m\) in each pair differing only in sign (figure 2). A similar set of eigenvalues results from diffusion map analysis of diffraction snapshots. (The differences between the two sets of eigenvalues are most likely due to the homogeneous metric approximation [13].)

Direct comparison of the eigenvalues of icosahedral Wigner D-functions with those obtained from diffusion map analysis (designated here by \(\tilde{\psi}\)) is not a sufficiently reliable means of identifying each \(\psi\) with its correct partner among the Wigner D-functions. This can be achieved by reference to plots of all snapshot coordinates for different pairs of \(\psi\). These display characteristic patterns, from which each of the 13 \(\psi\) can be reliably associated with one of the symmetrized Wigner D-functions \(\tilde{D}_{m,c}^j(\alpha)\) (figure 3). (The plots corresponding to \(m = \pm 3\) and \(\pm 5\) are similar. However, snapshots with fivefold symmetry occur at the centre of the \(m = \pm 3\) plot and along a circle in \(m = \pm 5\), allowing unambiguous distinction.)

Next, we describe how orientations can be extracted from analysis of diffraction snapshots. In principle, once each of the first 13 \(\psi\) has been identified with a symmetrized eigenfunction \(\tilde{D}_{m,c}^j(\alpha)\), the orientation of each snapshot can be extracted from its coordinates in the space spanned by the 13 \(\psi\). This is complicated, however, by the presence of symmetry, which introduces degeneracies in the symmetrized eigenfunctions, as outlined above. Clearly, all orthogonal and normalized degenerate \((\psi_i, \psi_j)\) pairs are equally acceptable. More precisely, any orthogonal operation on such a pair

---

**Figure 2.** Eigenvalue spectra of the Laplace–Beltrami operator. (a) Spectrum for icosahedral Wigner D-functions. (b) Spectrum obtained from the manifold produced by noise-free simulated diffraction snapshots of the STNV. Note the close agreement between the two spectra.

**Figure 3.** Scatter plots to identify eigenfunctions. These, together with the distribution of fivefold symmetric snapshots allow unambiguous association of the diffusion map eigenvectors \(\tilde{\psi}\) with their counterparts among the icosahedral Wigner D-functions \(\psi\). (a) \(\tilde{\psi}\) versus \(\psi\) plots of snapshot coordinates obtained from diffusion map. (b) \(\tilde{D}_m^j\) versus \(\tilde{D}_{-m}^j\) plots for randomly sampled points in the space of orientations.
of eigenfunctions leads to another equivalent pair. Each degenerate pair \((\psi_i, \psi_j)\) is thus related to its counterpart \((D^m_i, D^m_j)\) via an unknown mixing angle \(\theta_m\), and a scaling factor, viz.

\[
\begin{pmatrix}
D^m_i \\
D^m_j
\end{pmatrix} = 1/\sqrt{13} \begin{pmatrix}
\cos \theta_m & (m+1) \sin \theta_m \\
(1-m) \sin \theta_m & \cos \theta_m
\end{pmatrix} \begin{pmatrix}
\psi_i \\
\psi_j
\end{pmatrix}
\]

(2.2)

Additionally, it must be established whether an inversion operation should be inserted on the right side of equation (2.2).

The mixing angle for each of the six degenerate pairs can be thought of as the position of the hour hand on a clock. Ideally, one would like all six clocks to display Greenwich Mean Time (have the same mixing angle \(\theta_m\)). However, the arbitrary orthogonal operations allowed by the presence of degeneracy mean that each clock could show a different ‘local’ time. As orthogonal operations also include inversion, the sense of rotation of each clock could also be reversed. One must therefore determine the local time and the sense of clock rotation in order to relate the diffusion map eigenfunctions \(\psi_i\) to the symmetrized eigenfunctions \(D^m_i\). As described in more detail in the electronic supplementary material, section C, this can be accomplished as follows.

First, we describe how the orthogonal transformation of degenerate pairs or, equivalently, the mixing angle \(\theta_m\) can be determined. It can be easily shown that the rotation of an object through \(\pi\) about the \(y\)-axis changes the position of a snapshot in a plot of real Wigner \(-D^m\)-functions about the line \(\theta_m=0\), viz.: \((D^m_i, D^m_j) \rightarrow (D^m_i, -D^m_j)\). The line corresponding to \(\theta_m\) can be determined by rotating the object by \(\pi\) about the \(y\)-axis (snapshots \(a\) and \(\tilde{a}\) in figure 4). As shown in figure 5, in the presence of Friedel symmetry, the conjugate snapshot \(\tilde{a}\) can be simply produced by mirroring \(a\) about the detector \(x\)-axis. The mixing angle can thus be determined to within \(\pi\) by adding the appropriate mirror images of a subset of the snapshots to the dataset before diffusion map embedding (figure 4).

3. Results

We now demonstrate our approach by reconstructing the structure of an icosahedral virus to 1/100 of its diameter both with and without noise. For the STNV (PDB designation: 2BUK) used here, this corresponds to atomic resolution (0.2 nm).

To estimate the signal expected from viruses, we calculated the number of elastically scattered photons per Shannon pixel for the STNV [23], one of the smallest viruses, and for the Para- mecum bursaria chlorella virus [24], one of the larger viruses known (see table 1 and the electronic supplementary material, section E). The exact number of scattered photons depends on a number of parameters, but for each virus, it can be estimated from the number and energy of incident photons, the beam diameter and the maximum scattering angle. At photon energies typically used for XFEL studies of viruses, the number of photons scattered to a Shannon pixel at 30° collection angle ranges from approximately 1 to 3000. The signal level producing one scattered photon per Shannon pixel at 30° was therefore used to simulate Poisson noise. The resulting signal-to-noise ratio is well above those amenable to our approach without a denoising step [14].

Figures 1 and 6 demonstrate the performance of our approach with reference to simulated snapshots of STNV to 0.2 nm (crystallographic) resolution (corresponding to 1/100 of the object diameter) at an incident photon energy of 12.4 keV. The demonstration includes two sets of simulated
Figure 6. Comparison of the exact and recovered diffraction volumes. (a) A slice through the exact three-dimensional diffraction volume. Same slice through the recovered diffraction volumes (b) without noise and (c) with Poisson noise corresponding to a mean signal of one photon per Shannon pixel at a resolution corresponding to 1/100 of the object diameter.

Table 1. Number of photons scattered to a Shannon pixel at 30° by a small and a large virus. The beam diameter is matched to the object size (STNV: 20 nm; chlorella: 190 nm).

<table>
<thead>
<tr>
<th>Virus</th>
<th>Photon energy (keV)</th>
<th>Photons/pulse (mm$^{-2}$)</th>
<th>Obj. diameter/resolution</th>
<th>Scattered photons/Shannon pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>STNV</td>
<td>0.5</td>
<td>$10^{13}$</td>
<td>4</td>
<td>3770</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$4 \times 10^{12}$</td>
<td>17</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$10^{12}$</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$10^{12}$</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>Chlorella</td>
<td>0.5</td>
<td>$10^{13}$</td>
<td>40</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$4 \times 10^{12}$</td>
<td>158</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>$10^{12}$</td>
<td>198</td>
<td>1</td>
</tr>
</tbody>
</table>

data: one noise-free; the second including shot-noise corresponding to a mean of one photon per Shannon pixel at 0.2 nm. (For details, see the electronic supplementary material, section E.) The conditions are chosen to highlight the noise robustness and resolution of our approach. (The Chlorella virus at 2.5 keV would have served equally well.) The appropriate experimental conditions, of course, depend on a number of additional parameters, such as differential scattering with respect to the solvent, etc.

4. Discussion

We now outline the primary implications of our work. Incorporation of object symmetry has proved a powerful tool for recovering structure by established single-particle techniques, such as cryo electron microscopy (cryo-EM) [25]. By enhancing the effective number of snapshots and improving resolution, our approach promises to play a similarly important role in three-dimensional structure recovery by XFEL methods. In terms of resolution expressed as a fraction of the object diameter, our approach is comparable with the best achieved by cryo-EM approaches [26], but without phase information. Combined with its superior noise robustness [15,17], our approach offers a vital route to determining high-resolution structure at signal levels expected even from single macromolecules in XFEL experiments [10,15,27]. For biological entities in particular, this is essential for obtaining ‘biologically relevant’ information.

**Figure 6.** Comparison of the exact and recovered diffraction volumes. (a) A slice through the exact three-dimensional diffraction volume. (b) Same slice through the recovered diffraction volumes without noise and (c) with Poisson noise corresponding to a mean signal of one photon per Shannon pixel at a resolution corresponding to 1/100 of the object diameter.

Table 1. Number of photons scattered to a Shannon pixel at 30° by a small and a large virus. The beam diameter is matched to the object size (STNV: 20 nm; chlorella: 190 nm).

<table>
<thead>
<tr>
<th>Virus</th>
<th>Photon energy (keV)</th>
<th>Photons/pulse (mm$^{-2}$)</th>
<th>Obj. diameter/resolution</th>
<th>Scattered photons/Shannon pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>STNV</td>
<td>0.5</td>
<td>$10^{13}$</td>
<td>4</td>
<td>3770</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$4 \times 10^{12}$</td>
<td>17</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$10^{12}$</td>
<td>42</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>$10^{12}$</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>Chlorella</td>
<td>0.5</td>
<td>$10^{13}$</td>
<td>40</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$4 \times 10^{12}$</td>
<td>158</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>$10^{12}$</td>
<td>198</td>
<td>1</td>
</tr>
</tbody>
</table>

We now outline the primary implications of our work. Incorporation of object symmetry has proved a powerful tool for recovering structure by established single-particle techniques, such as cryo electron microscopy (cryo-EM) [25]. By enhancing the effective number of snapshots and improving resolution, our approach promises to play a similarly important role in three-dimensional structure recovery by XFEL methods. In terms of resolution expressed as a fraction of the object diameter, our approach is comparable with the best achieved by cryo-EM approaches [26], but without phase information. Combined with its superior noise robustness [15,17], our approach offers a vital route to determining high-resolution structure at signal levels expected even from single macromolecules in XFEL experiments [10,15,27]. For biological entities in particular, this is essential for obtaining ‘biologically relevant’ information.

**5. Summary and conclusion**

We have demonstrated the first approach capable of extracting high-resolution three-dimensional structure from diffraction snapshots of symmetric objects and presented structure recovery to 1/100 of the object diameter at signal-to-noise ratios expected from currents XFELs. This opens the way to the study of individual biological entities before the onset of significant radiation damage. Our approach also offers the possibility to apply powerful graph-theoretic techniques to the study of crystalline objects, with the potential to extract more information from the rich and rapidly growing body of nanocrystallographic data.

**Acknowledgements.** We are grateful to Dimitrios Giannakis and Dilano Saldin for valuable discussions.

**Funding statement.** This research was supported by: the US Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-FC02-09ER16114 (overall design); the US National Science Foundation under award no. MCB-1240590 (algorithm development); CCF-1013278 and CNS-0968519 (GPU algorithms) and the UWM Research Growth Initiative (theory). The publication of this work was supported by the US National Science Foundation under award no. STC 1231306.