

Simulations on time-resolved structure determination of uncrystallized biomolecules in the presence of shot noise

K. Pande,^{a)} M. Schmidt, P. Schwander, and D. K. Saldin
*Department of Physics, University of Wisconsin-Milwaukee, Milwaukee,
Wisconsin 53211, USA*

(Received 24 November 2014; accepted 26 March 2015; published online 10 April 2015)

Determination of fast structural changes of biomolecules is usually performed on crystalline samples in a time-resolved pump-probe experiment. Changes in the structure are found by the difference Fourier method using phases of a known reference structure. As we showed recently, such changes can also be determined from diffraction of uncrystallized molecules in random orientations. In this case, the difference in the angular correlations of the diffraction patterns is used to find structural changes. Similar to the difference Fourier method, there is no need for iterative phasing. We validated this approach previously with simulations in the absence of noise. In this paper, we show that the effects of noise can be adequately suppressed by averaging over a sufficiently large ensemble as they can be obtained using an X-ray free electron laser. © 2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License. [<http://dx.doi.org/10.1063/1.4916980>]

I. INTRODUCTION

Time-resolved crystallography (TRX)¹ can be used to determine the structure of short-lived (transient) reaction intermediates in biological macromolecules.² TRX is traditionally performed using synchrotron-based Laue diffraction experiments that are currently restricted by X-ray beam brilliance, uniform reaction initiation, and the time resolution of the probe x-ray pulse. The advent of X-ray free electron lasers (XFELs) such as the Linac Coherent Light Source (LCLS)^{3,4} and the SPring-8 Angstrom Compact Free-Electron Laser (SACLA)⁵ have opened the avenue for ultrafast time-resolved studies.^{6,7} Recently, high-quality difference electron density maps were generated from serial femtosecond crystallography (SFX)⁸ on nanocrystals of photoactive yellow protein (PYP),⁹ which allowed determination of reaction intermediates up to 1.6 Å. The critical bottleneck of synchrotron and XFEL-based crystallography lies in growing crystals of sufficient size and quality,¹⁰ which is particularly difficult for membrane proteins. At the same time, crystal lattice constraints might influence both structure and kinetics extracted from the time-resolved X-ray data.¹¹ Other successful methods for structure determination, like nuclear magnetic resonance (NMR) and cryogenic electron microscopy (cryo-EM), pose restrictions on the kind of systems that can be studied—NMR methods are limited to proteins of about 40 kDa in molecular weight,¹² while cryo-EM is limited to intermediate resolutions of 6–10 Å (Refs. 13 and 14) and cannot be used for time-resolved studies. It is therefore desirable to develop non-crystallographic techniques for structure determination and to perform time-resolved experiment on biomolecules in solution and at physiologically relevant temperatures, so that unconstrained molecular relaxations can occur.

It was pointed out by Kam¹⁵ that diffraction patterns from an ensemble of molecules in random orientations and frozen in time by using intense, brief X-ray pulses contain structural

^{a)} Author to whom correspondence should be addressed. Electronic mail: kpande@uwm.edu

information that tend to get washed out in the case of conventional small-angle X-ray scattering (SAXS),¹⁶ which is obtained by exposure of molecules to X-ray pulses on time-scales longer than their rotational diffusion times. This structural information can be extracted by measuring intensity angular correlations, which yield a quantity that can be related to the three-dimensional diffraction volume of a single molecule. Several proof-of-principle experiments have validated Kam's basic theory. Saldin and coworkers have employed particle symmetry to reconstruct the 3D structure of icosahedral¹⁷ and cylindrical virus¹⁸ from angular correlations. Liu *et al.*^{19,20} have demonstrated the feasibility of *ab initio* reconstruction of low-resolution structures by using angular correlations in a fluctuation scattering experiment. In a previous paper,²¹ we proposed and demonstrated by computer simulations the possibility of performing a pump-probe experiment to determine the difference electron density of a biomolecule on photoexcitation of an ensemble of uncrystallized randomly oriented biomolecules. We showed that the difference in angular correlations of two closely related sets of molecules, such as the "dark" and photoexcited molecules of PYP, is linearly related to the change in the electron density provided the structural change is small.

The method of correlated X-ray scattering has tremendous appeal because it is applicable to *solution-phase studies at room temperature* and does not suffer from algorithmic issues associated with "hit-finding" and orientation determination. Despite various successful theoretical studies on the basic methodology of correlated X-ray scattering, there is still a significant lack of understanding of the underlying factors that limit the signal-to-noise ratio (SNR) for intensity correlation measurements, which makes the analysis of experimental data a challenge. An important source of error is the detector shot noise, which we have modeled by Poisson statistics due to the discrete nature of photon arrival at the detector pixels. When the expected number of photons/pixel is large (~ 10), Poisson distribution can be approximated by Gaussian distribution. While it is widely accepted that scattering from solution is the largest contributor towards shot noise, we have only considered scattering from biomolecules in the present study. A detailed treatment of scattering from solvent and the resultant contrast loss will be the focus of a future paper. Another source of error comes from XFEL beam jitter due to the stochastic nature of the emitted pulses,²² resulting in shot-to-shot intensity variations.

The main goal of this paper is to assess the behavior of the computational approach suggested in Ref. 21 for the extraction of difference electron density from angular correlations in the presence of substantial noise. We performed simulations employing a large number of "noisy" diffraction patterns calculated from an ensemble of molecules in random orientations and show that meaningful difference electron density can be extracted even in the presence of substantial noise and from an ensemble of particles that contribute to individual diffraction patterns.

II. METHODS

We begin by defining an angular pair correlation function between the intensities at points in the detector plane specified by the polar coordinates (q, ϕ) and $(q', \phi + \Delta\phi)$, averaged over a set of measured XFEL diffraction patterns corresponding to several random particle orientations

$$J(q, q', \Delta\phi) = \langle I(q, \phi)I(q', \phi + \Delta\phi) \rangle_{\phi, DP}, \quad (1)$$

where $\langle \dots \rangle_{\phi, DP}$ represents an average over all azimuthal angles and all diffraction patterns. If N particles contribute incoherently to a single diffraction pattern, the total intensity may be written as

$$I_N(q, \phi) = \sum_{i=1}^N I_i(q, \phi, \omega_i), \quad (2)$$

where ω_i is the particle's orientation, and the sum over intensities is justified even in the coherent case if the inter-particle separations are random.^{21,23} Thus, Eq. (1) has a contribution from a term where the particle orientations are the same, namely,

$$\left\langle \sum_i I(q, \phi, \omega_i) I(q', \phi + \Delta\phi, \omega_i) \right\rangle_{\phi, DP}, \quad (3)$$

and one where the particle orientations are different, namely,

$$\left\langle \sum_i I(q, \phi, \omega_i) \sum_{j \neq i} I(q', \phi + \Delta\phi, \omega_j) \right\rangle_{\phi, DP} \rightarrow \left\langle \sum_i I(q, \phi, \omega_i) \right\rangle_{\phi, DP} \left\langle \sum_j I(q', \phi, \omega_j) \right\rangle_{\phi, DP}. \quad (4)$$

In Eq. (4), the arrow indicates the mathematical limit when the number of diffraction patterns $\rightarrow \infty$. Note that the second term does not exist for scattering from single particles. However, since typical illumination areas in current applications of XFEL radiation tend to be much larger than a small biological molecule like PYP, often the subject of time-resolved studies in crystals, it is quite likely that a typical sample will contain multiple particles in a typical experiment, so we keep the discussion general for now. For multiple scattering particles, we can remove the second term from Eq. (1) by redefining the pair correlations as

$$C_2(q, q', \Delta\phi) = \langle I(q, \phi) I(q', \phi + \Delta\phi) \rangle_{\phi, DP} - \langle I(q, \phi) \rangle_{\phi, DP} \langle I(q', \phi) \rangle_{\phi, DP}. \quad (5)$$

Intensities $I(q, \phi)$ on a polar grid are computed by cubic interpolation of simulated intensities $I(q_x, q_y)$ from an oversampled Cartesian grid (since XFEL detectors are usually arranged on a Cartesian grid in the detector plane). The angular pair correlation function $C_2(q, q', \Delta\phi)$ is related to a quantity $B_l(q, q')$,¹⁵ which depends on the angular momentum quantum number l . The quantity B_l is obtained from the measured C_2 quantities by performing the integral

$$B_l(q, q') = \int C_2(q, q', \Delta\phi) P_l(\cos[\theta(q)] \cos[\theta(q')] + \sin[\theta(q)] \sin[\theta(q')] \cos[\Delta\phi]) d(\Delta\phi), \quad (6)$$

where

$$\theta(q) = \frac{\pi}{2} - \sin^{-1}\left(\frac{q}{2\kappa}\right), \quad (7)$$

κ is the wavenumber of the incident radiation (defined by $2\pi/\lambda$, where λ is the x-ray wavelength), and P_l is a Legendre function of l th degree. As has been shown previously,¹⁷ $B_l(q, q')$ can be expressed in terms of the spherical harmonic components $I_{lm}(q)$ of the 3D diffraction volume $I(\vec{q})$ of a single particle by

$$B_l(q, q') = \sum_m I_{lm}(q) I_{lm}^*(q'). \quad (8)$$

Consider small structural changes in the biomolecules upon photoexcitation. Under this assumption, we may take a variation on both sides of Eq. (8) to give

$$\delta B_l(q, q') = \sum_m [\delta I_{lm}(q) I_{lm}^*(q') + I_{lm}(q) \delta I_{lm}^*(q')]. \quad (9)$$

In terms of the complex structure factor $A(\vec{q})$ and the difference electron density $\delta\rho(\vec{r}_j)$ between the dark and photoexcited structures, we can write $\delta I_{lm}(q)$ as

$$\begin{aligned} \delta I_{lm}(q) &= \int d\Omega_{\hat{q}} [\delta A(\vec{q}) A^*(\vec{q}) + A(\vec{q}) \delta A^*(\vec{q})] Y_{lm}(\hat{q}), \\ &= \sum_{\vec{r}_j} \delta\rho(\vec{r}_j) \int d\Omega_{\hat{q}} [e^{i\vec{q}\cdot\vec{r}_j} A^*(\vec{q}) + A(\vec{q}) e^{-i\vec{q}\cdot\vec{r}_j}] Y_{lm}(\hat{q}), \\ &\equiv \sum_{\vec{r}_j} \delta\rho(\vec{r}_j) C_{lm}(q; \vec{r}_j), \end{aligned} \quad (10)$$

where we have used the relation

$$\delta I(\vec{q}) = \delta A(\vec{q})A^*(\vec{q}) + A(\vec{q})\delta A^*(\vec{q}), \quad (11)$$

and

$$\delta A(\vec{q}) = \sum_{\vec{r}_j} \delta \rho(\vec{r}_j) e^{i\vec{q}\cdot\vec{r}_j}. \quad (12)$$

Using Eq. (10) in Eq. (9), we get

$$\delta B_l(q, q') = \sum_{\vec{r}_j} \delta \rho(\vec{r}_j) \sum_m [C_{lm}(q; \vec{r}_j) I_{lm}^*(q') + I_{lm}(q) C_{lm}^*(q'; \vec{r}_j)], \quad (13)$$

and

$$\delta \rho(\vec{r}) = \sum_{lq q'} (M^{-1})_{\vec{r}; lq q'} \delta B_l(q, q'), \quad (14)$$

where

$$M_{lq q'; \vec{r}} = \sum_m [C_{lm}(q; \vec{r}) I_{lm}^*(q') + I_{lm}(q) C_{lm}^*(q', \vec{r})]. \quad (15)$$

In Eq. (15), matrix M depends only on the complex structure factor $A(\vec{q})$ of the dark structure, which is usually completely known in time-resolved experiments, and the quantity $\delta B_l(q, q')$ is determined directly from the experimentally measured intensities of the dark and excited structures. Since the correlations $C_2(q, q', \Delta\phi)$ are symmetric in q , $\delta B_l(q, q')$ is also symmetric in q , i.e.,

$$\delta B_l(q, q') = \delta B_l(q', q), \quad (16)$$

$$\Rightarrow M_{lq q'; \vec{r}_j} = M_{lq' q; \vec{r}_j}. \quad (17)$$

Looking at the form of M in Eq. (15), we see that

$$M_{lq' q; \vec{r}_j} = M_{lq q'; \vec{r}_j}^*, \quad (18)$$

hence the elements of M are real. Similarly, the elements of the column vector $B_l(q, q')$ are also real.

For a particle in an orientation ω , the scattered intensity collected on a detector pixel of solid angle $\Delta\Omega$ at a scattering vector \vec{q} is given as

$$I(\vec{q}, \omega) = F \left(\frac{d\sigma}{d\Omega} \right)_{\vec{q}, \omega} \Delta\Omega, \quad (19)$$

where F is the incident flux (photons/area), the differential cross-section of the particle is

$$\frac{d\sigma}{d\Omega} = r_e^2 |A(\vec{q}, \omega)|^2, \quad (20)$$

with r_e being the classical electron radius and the solid angle subtended at the sample by a detector pixel at an oversampling ratio $o \geq 2$ in the reciprocal space is given as

$$\Delta\Omega \approx \left(\frac{\lambda}{oD} \right)^2, \quad (21)$$

where λ is the wavelength of the incoming radiation and D is the diameter of the particle. Combining the above equations, the expression for the intensity detected by a single pixel is

$$I(\vec{q}, \omega) = F r_e^2 |A(\vec{q}, \omega)|^2 \left(\frac{\lambda}{oD} \right)^2. \quad (22)$$

To study the effect of noise on the reconstruction of the difference density $\delta\rho(\vec{r})$, we introduce shot noise at each detector pixel. We define $P(n, I)$ to be the probability of measuring a photon count n with an expected intensity of I . The probability $P(n, I)$ can be written as

$$P(n, I) = P(n|I)P(I), \quad (23)$$

where $P(n|I)$ is the conditional probability of measuring n photons when the expected intensity is known to be I and is given by the Poisson distribution

$$P(n|I) = \frac{I^n e^{-I}}{n!}. \quad (24)$$

III. RESULTS AND DISCUSSION

The simulations reported in this paper were done for the PYP (protein data bank (PDB) entry 2PHY). Since we are working with single isolated particles, we did not assume any periodicity of the reconstructed real-space image and took the size of the real space volume that we reconstructed as a 50 \AA cube, a little more than the size of a PYP molecule in a crystal. The simulations were done for a real space resolution $\text{res} = 5 \text{ \AA}$, giving rise to a $(21 \times 21 \times 21)$ grid in which the difference density was represented.

As per the following relations:

$$q_{max} = \frac{2\pi}{res}, \quad (25)$$

$$l_{max} = \frac{q_{max}D}{2}, \quad (26)$$

the values for q_{max} and l_{max} at 5 \AA resolution and a diameter of 50 \AA were found to be $\sim 1.25 \text{ \AA}^{-1}$ and 30, respectively. Since the intensities follow Friedel's law, i.e., $I(\vec{q}) = I(-\vec{q})$, only even values of l have non-zero $I_{lm}(q)$. Note that the width of the reciprocal space grid is taken in accordance with the Shannon spacing

$$dq = \frac{2\pi}{2D}, \quad (27)$$

to ensure that the rows of matrix $M_{lqq';\vec{r}}$ are linearly independent. In order to have an almost square matrix M , the number of resolution rings in the reciprocal space on a polar grid is given by

$$N_q^{sph} \sim \sqrt{\left(\frac{2N_{pixels}}{l_{max}} \right)}. \quad (28)$$

Hence, the intensities have to be calculated for q'_{max} higher than the q_{max} , where

$$q'_{max} = dq * N_q^{sph}. \quad (29)$$

The DPs were simulated on a Cartesian grid, oversampled eight times in a linear dimension, and then interpolated to a polar grid using cubic interpolation. We simulated DPs in

1 000 000 random orientations²⁴ and then selected a random subset of these DPs for our calculations. For simulations with noise, since the Poisson shot noise applies to each detector pixel, it was applied to intensities calculated on the Cartesian grid before cubic interpolation. Using $F = 8 \times 10^{12}$ photons/ μm^2 per pulse, $r_e = 2.8179 \times 10^{-15}$ m, and $\lambda = 2 \text{ \AA}$ ($\sim 6 \text{ keV}$) in Eq. (22), and Wilson statistics for structure factor, we find that PYP scatters ~ 3 photons/Shannon pixel at 5 \AA resolution, as shown in Fig. 1. Simulations have also been done for higher photon counts of 15 and 30 photons/Shannon pixel. Typical 2D DPs for these photon counts are shown in Fig. 2.

A. Matrix inversion: Singular value decomposition (SVD)

The SVD of a matrix M (where M is a $m \times n$ matrix, with $m > n$) is the factorization of M into the product of three matrices

$$M = U\Sigma V^T, \quad (30)$$

where U is an $m \times m$ matrix, Σ is an $m \times n$ matrix, and V is an $n \times n$ matrix. U and V have orthogonal columns so that

$$U^T U = I, \quad (31)$$

and

$$V^T V = I, \quad (32)$$

and Σ has entries only along the diagonal. The diagonal elements of Σ are non-negative and appear in decreasing order such that

$$\sigma_1 \geq \dots \geq \sigma_n \geq 0. \quad (33)$$

We find that the matrix M is highly ill-conditioned (the condition number of M is the ratio σ_1/σ_n , which for the present case is $\sim 10^{22}$), hence we use truncated SVD (TSVD) to compute M^{-1} to solve for $\delta\rho$. The solution $\delta\rho$ to the system of equations

$$\delta B = M\delta\rho \quad (34)$$

is given as

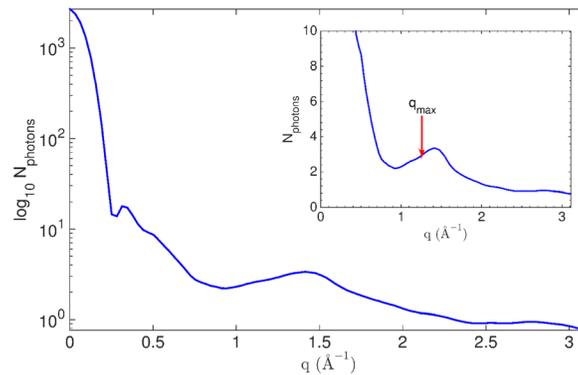


FIG. 1. Plot for log of number of photons scattered per Shannon pixel as a function of the magnitude of the scattering vector, for $\lambda = 2 \text{ \AA}$, incident flux $= 8 \times 10^{12}$ photons/ μm^2 per pulse, and a beam diameter of 200 \AA . Inset shows the region around $q_{\text{max}} = 1.25 \text{ \AA}^{-1}$ at which PYP scatters ~ 3 photons/Shannon pixel at 5 \AA resolution with the above beam parameters.

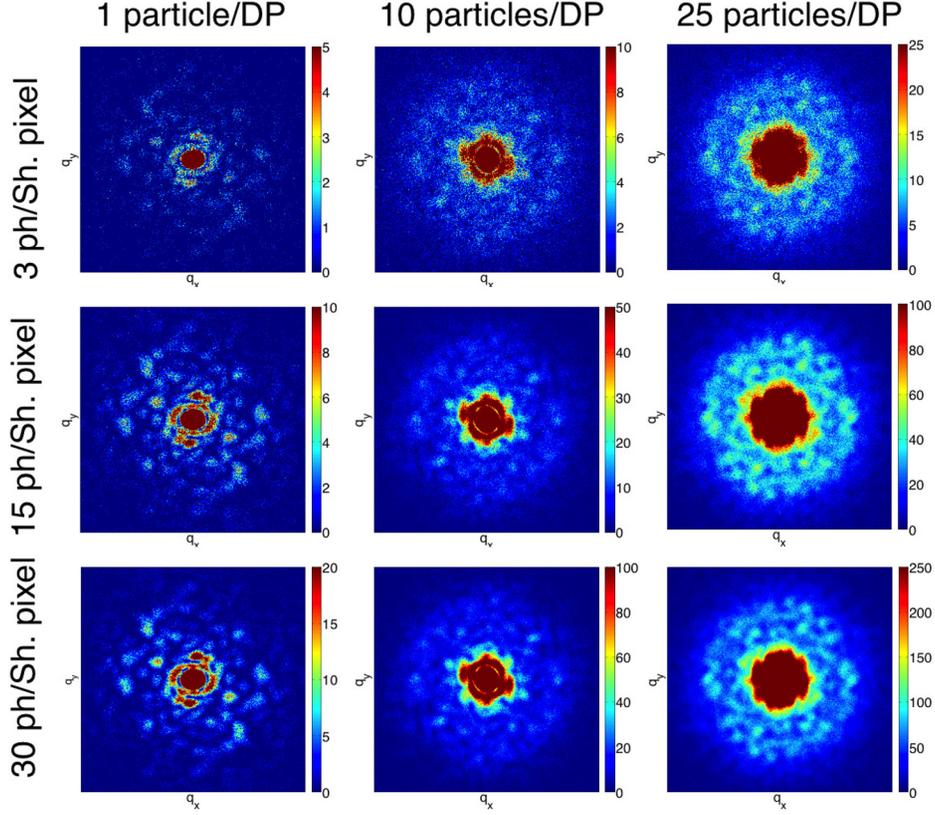


FIG. 2. Simulated 2D diffraction patterns at 5 Å resolution for 1 particle/DP, 10 particles/DP, and 25 particles/DP for photon count of 3, 15, and 30 photons/Shannon pixel.

$$\delta\rho = \sum_{i=1}^n f_i \frac{\mathbf{u}_i^T \delta\mathbf{B}}{\sigma_i} \mathbf{v}_i, \quad (35)$$

where f_i are filter factors, and for the case of TSVD $f_i = 0, 1$. Here, \mathbf{u}_i and \mathbf{v}_i are the columns of matrices U and V , respectively. For a discrete ill-posed problem, such as the one under consideration in this paper, the coefficients $|\mathbf{u}_i^T \delta\mathbf{B}|$ corresponding to the smaller singular values do not decay as fast as the singular values, but rather tend to level off as shown in Fig. 3. The solution is therefore dominated by the terms in the sum corresponding to the smallest σ_i and appears completely random.²⁵

B. Reconstruction of difference density

To quantify the quality of the difference electron density recovered with our algorithm, we use the Pearson correlation coefficient²⁶ CC_r , given by

$$CC_r = \frac{\langle \delta\rho_1 \cdot \delta\rho_2 \rangle - \langle \delta\rho_1 \rangle \cdot \langle \delta\rho_2 \rangle}{\sqrt{[\langle \delta\rho_1^2 \rangle - \langle \delta\rho_1 \rangle^2] \cdot [\langle \delta\rho_2^2 \rangle - \langle \delta\rho_2 \rangle^2]}}. \quad (36)$$

Here $\Delta\rho_1$ and $\Delta\rho_2$ are the two difference densities being compared and are limited to the pixels at which $|\Delta\rho| < 2\sigma$, where σ is the variance of $\Delta\rho$.

In the rest of the paper, we will look at the reconstruction of difference density for (I) the cis-trans isomerization of the chromophore (HC4) and arginine (ARG52) and (II) the movement of widely separated residues HC4 and PHE121. Since the initial dark structure is assumed

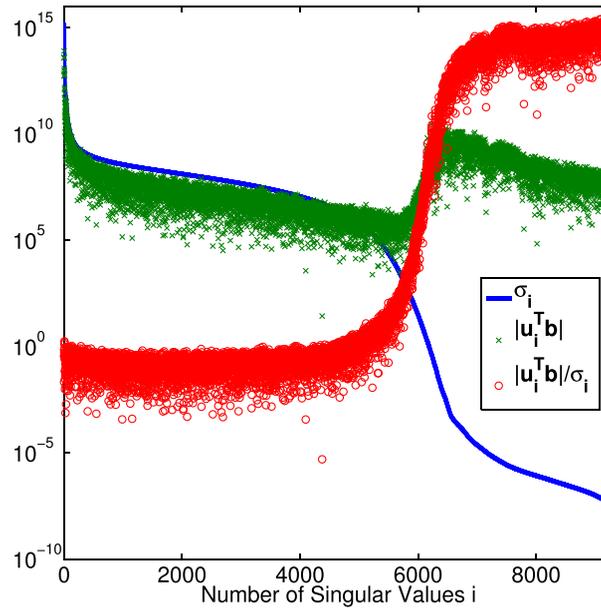


FIG. 3. Variation of $|\mathbf{u}_i^T \delta \mathbf{B}|$ (green), singular values σ_i (blue), and the ratio $|\mathbf{u}_i^T \delta \mathbf{B}|/\sigma_i$ (red) as a function of the number of singular values i . \mathbf{u}_i are columns of U and σ_i are the singular values of M . The vector $\delta \mathbf{B}$ is from correlations of noise free simulations with one particle per DP [Eq. (6)]. Note that the singular values decay faster than $|\mathbf{u}_i^T \delta \mathbf{B}|$ for $i \geq 5000$.

known, the matrix M and its SVD need to be computed only once. We first calculate the real-space correlation coefficient between difference density calculated directly from the PDB files of the two structures ($\delta\rho_1 - \delta\rho_2 = \delta\rho_{pdb}$), and that using our suggested inversion method [Eq. (14)]. The $\delta B_l(q, q')$ are computed using the spherical harmonic decomposition of intensities calculated from the PDB files ($\delta\rho_2 = \delta\rho_{I_m}$). Fig. 4 shows difference densities $\delta\rho_{pdb}$ and

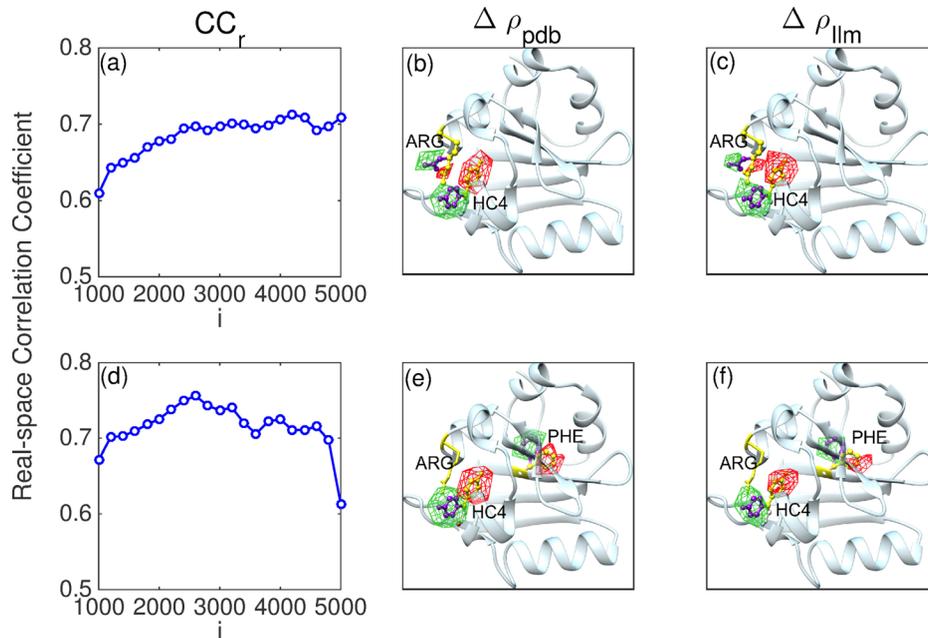


FIG. 4. Real-space correlation coefficient CC_r between $\delta\rho_{pdb}(\vec{r})$ calculated from PDB files and $\delta\rho_{I_m}(\vec{r})$ calculated using our algorithm for structural changes in (a) I and (d) II. $B_l(q, q')$ is obtained using the relation $B_l(q, q') = \sum_m I_m(q) I_m(q')$, wherein the spherical harmonic coefficients I_m of intensities are directly calculated from PDB files. CC_r is plotted as a function of the number i of singular values used for calculating M^{-1} . The corresponding $\delta\rho_{pdb}$ and $\delta\rho_{I_m}$ for I and II are shown in (b), (c) and (e), (f), respectively.

$\delta\rho_{l_m}$, and the real-space correlation coefficient between them as a function of the number of singular values i with filter factors $f_i = 1$ in Eq. (35). Top and bottom panels show the structural changes in I and II, respectively. The $\delta\rho$ features in Figs. 4(b) and 4(c) and those in Figs. 4(e) and 4(f) look very similar, and result in highest real-space correlation coefficient CC_r between $\delta\rho_{pdb}$ and $\delta\rho_{l_m}$ is ~ 0.7 for I and 0.75 for II. It must be pointed out that while the truncation of singular values helps to regularize the solution to $\delta\rho$, it results in a loss of resolution. For low values of truncation parameter i , the density on ARG residues is not clearly resolved. It should also be noted that the difference features are clearer when the residues are widely separated as in structure II (lower panel), than when they are very close, as in structure I (top panel).

Next, we consider the effect that the number of particles exposed on each diffraction pattern has on the resulting density reconstructions ($\delta\rho_{corr}$). In Fig. 5, we show results for CC_r and $\delta\rho_{corr}$ for quantities obtained from simulated noise-free DPs for structural change I, which involves movement of residues HC4 and ARG52. Figs. 5(a), 5(b), and 5(c) show CC_r plotted as a function of the number of singular values and the number of DPs over which the correlations are averaged for calculation of $B_l(q, q')$ for DPs with a single particle, 10 particles, and 25 particles per snapshot, respectively. The corresponding $\delta\rho(\vec{r})$ are shown in panels (d), (e), and (f). In all cases, we find that CC_r improves with increase in the number of DPs, and in all cases CC_r is highest when the singular values are truncated at between $i = 3000$ and 4000, dropping considerably for 5000 singular values. For DPs with a single particle per snapshot, a good convergence to the expected structure is obtained with as few as 30 000 DPs. On the other hand, for the DPs with 10 and 25 particles per snapshot, the CC_r for 100 000 DPs is lower than that for 10 000 DPs when there is only one particle per snapshot.

We now consider the case when the 2D snapshots contain shot noise given by the Poisson distribution in Eq. (24). Noisy correlations $C_2(q, q', \Delta\phi)$ are calculated for snapshots according to our prescription that each detector pixel detects only a whole number of photons. Calculations have been done for scattering from one particle, 10 particles, and 25 particles per snapshot each for three different values of measured photon counts per molecule: (i) 3 photons/Shannon pixel, which is the number of photons scattered by PYP at 5 Å resolution when the incoming flux $F = 8 \times 10^{12}$ photons/ μm^2 per pulse (expected at the European XFEL²⁷ due to come online in 2016) and beam diameter is 200 Å, (ii) 15 photons/Shannon pixel, and (iii) 30 photons/Shannon pixel.

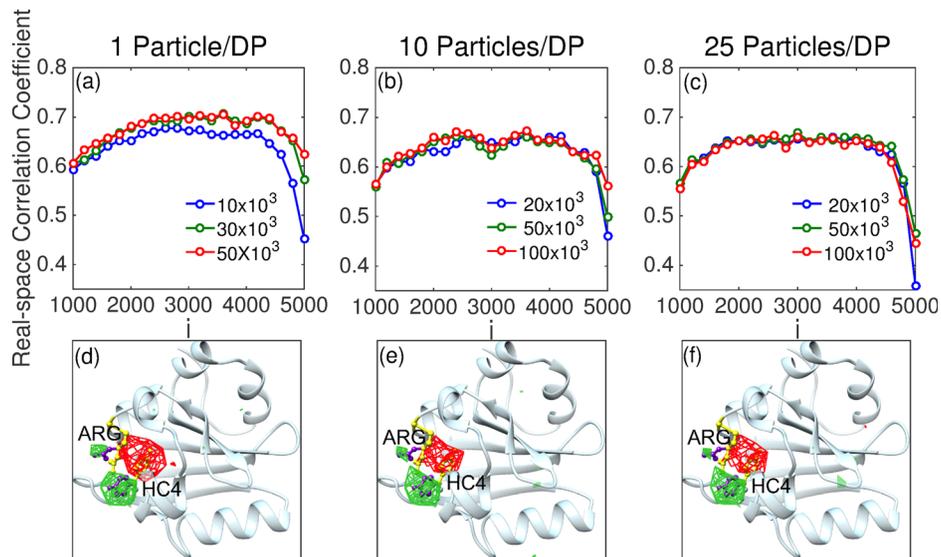


FIG. 5. CC_r between $\delta\rho_{pdb}$ and $\delta\rho_{corr}$ calculated from matrix inversion in Eq. (35), wherein vector $B_l(q, q')$ is obtained from the method of correlations on DPs with (a) one particle per snapshot, (b) 10 particles per snapshot, and (c) 25 particles per snapshot. CC_r is plotted as a function of the number i of singular values. The corresponding difference electron density $\delta\rho_{corr}$ from angular correlations averaged over (d) 50 000 DPs, and (e) and (f) 100 000 DPs.

The correlation coefficients CC_r between ideal difference density and those recovered by our technique are shown in Fig. 6, for expected movements of the chromophore and nearby ARG 52 residue in a time-resolved pump-probe photo excitation experiment. The CC_r is shown as a function of both the number of particles illuminated and the detected scattering signal expressed in terms of photons/Shannon pixel. The corresponding difference density reconstructed for the point with the highest CC_r is shown in Fig. 7. For a signal of 3 photons/Shannon pixel and scattering from one particle per snapshot, the signal is too weak to reproduce any difference features when the correlations are averaged over 400 000 DPs, as seen in Fig. 7(a). When the scattering is from 10 and 25 particles per snapshot, CC_r shows some improvement, which is manifested in the appearance of difference features on the chromophore (HC4 residue) of the molecule in Figs. 7(b) and 7(c), respectively. For a signal of 15 photons/Shannon pixel, only the difference density on the chromophore is recovered from simulations for both the single and multiple particle snapshots, as shown in Figs. 7(d)–7(f). A small signal on the ARG is seen for a signal of 30 photons/Shannon pixel when the scattering is from only one particle/DP. However, when there are multiple particles/DP, the CC_r is lower and the difference density features are weaker, and show similar trend as the CC_r for the noise free case. In Ref. 23, Kirian *et al.* have also observed that the SNR for correlated fluctuation SAXS (CFSAXS) in the high flux regime rapidly approaches a limit independent of the number of particles contributing to the diffraction pattern.

Fig. 8 shows the variation of CC_r as a function of the scattering signal and also the number of particles in the 2D snapshots. It is clear that at low flux (signal = 3 and 15 photons/Shannon pixel), the correlation coefficient improves when the scattering is from multiple particles per snapshot. While there is a slight increase in CC_r for 15 photons/pixel when going from 10 to 25 particles/DP, it decreases for signal of 3 photons/Shannon pixel. This is because both these data points come from correlations averaged over 400 000 snapshots. When the signal is lower and the number of particle is higher, a larger number of diffraction patterns are required to recover the difference features. It is important to note that for a signal of 30 photons/Shannon pixel as well as for the noise free case, increasing the number of particles in the snapshot does not improve the CC_r and the difference density. Similar conclusions have been reported in Ref. 23.

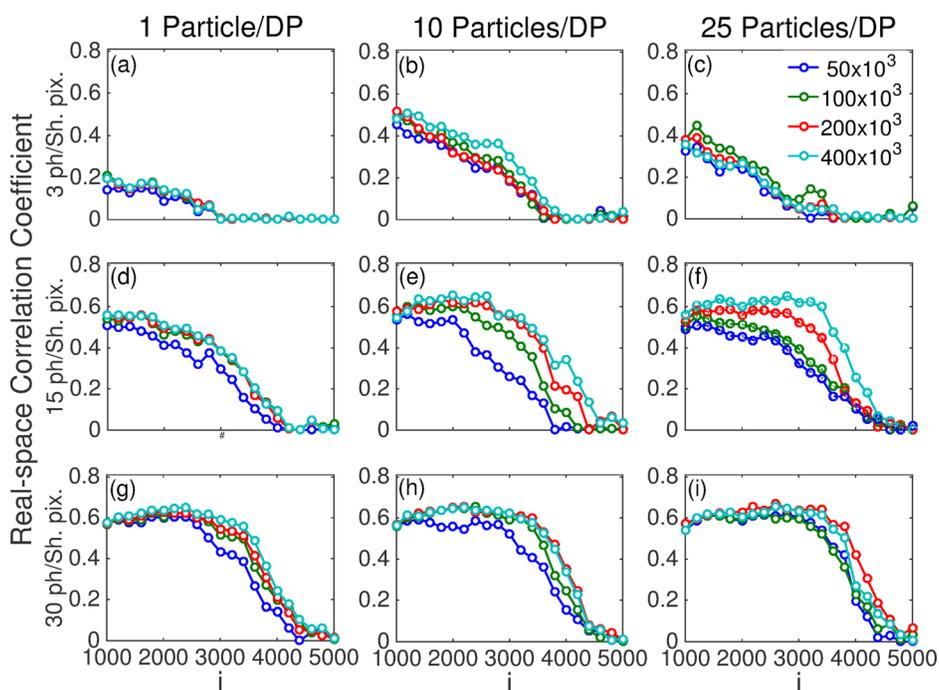


FIG. 6. Correlation coefficient CC_r as a function of the number i of singular values used for calculating M^{-1} . Results are from diffraction patterns with one, 10, and 25 particles/snapshot and for a signal of 3, 15, and 30 photons/Shannon pixel.

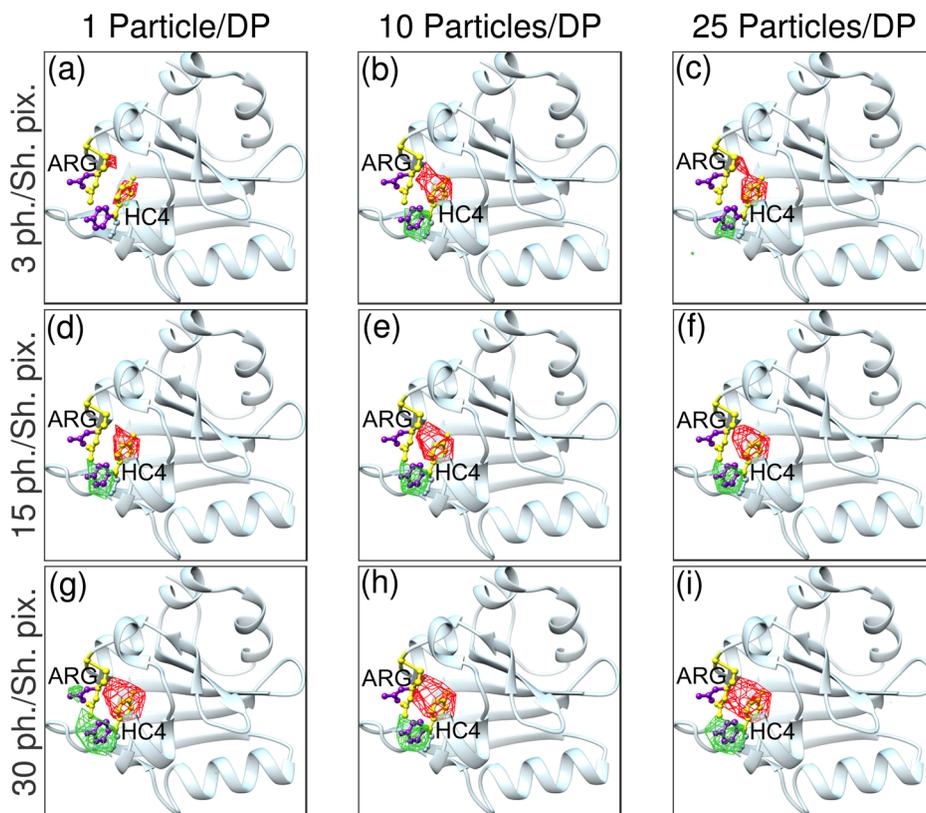


FIG. 7. Difference density for movements of the chromophore and arginine residue in a simulated time-resolved experiment on PYP. Results are for scattering signal of 3, 15, and 30 photons/Shannon pixel and from 2D snapshots with one, 10, and 25 particles/DP. For all the calculations, angular correlations are averaged over 400 000 DPs.

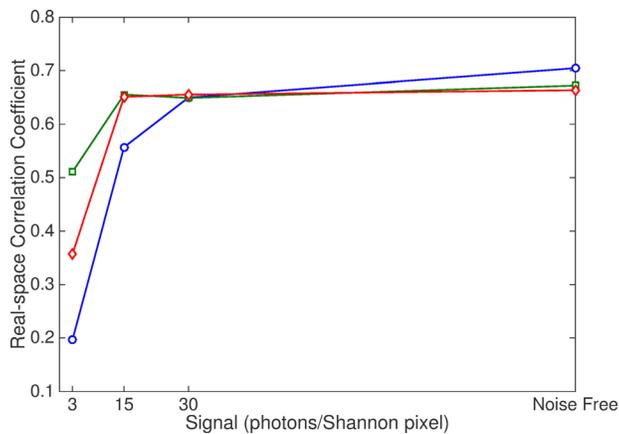


FIG. 8. Real-space correlation coefficient CC_r for cis-trans isomerization of HC4 and movement of ARG 52. CC_r is shown as a function of detected signal and for scattering from one (blue, circle), 10 (green, square), and 25 (red, diamond) particles per snapshot.

IV. CONCLUSIONS

The method of analyzing XFEL data of uncrystallized molecules using angular correlations can be exploited to reveal fast structural changes of biomolecules in a pump-probe experiment without the need for a phasing algorithm. Unlike algorithms for *ab initio* structure determination which reconstruct only a 3D diffraction volume from which the real-space structure needs to be deduced by means of an iterative phasing algorithm, it is possible²¹ to derive an algorithm

that can reveal directly the real-space difference structure in analogy to the difference Fourier method of time-resolved crystallography.

The present paper examines the feasibility of this method in the presence of noise. Kirian *et al.*²³ have pointed out that due to the fact that the method of angular correlations gives the signal required as a small perturbation of a much larger SAXS/wide-angle x-ray scattering (WAXS) signal from uncorrelated scattering from differently oriented particles, the noise on the SAXS/WAXS signal is actually comparable to the signal sought.

The same paper concluded that the signal-to-noise ratio actually improved with $N^{1/2}$, where N is the number of diffraction patterns used in the analysis. The point is that the distorting effect is noise, while the sought effect is signal. Noise can be reduced by averaging over a large number of diffraction patterns, whereas such reductions do not occur in the signal. However, the signal we seek in such experiments is very small since it is associated with a difference structure. Consequently, even a small amount of noise can significantly distort it, and averaging may need to be performed over a large number of diffraction patterns. Exactly how many are needed can only be found by detailed simulations as performed in this paper, or even better by experiment. The results of this paper suggest that such experiments may be feasible.

It should be noted that at the pulse repetition rate of a current generation XFEL, like the LCLS, it is not unusual to measure a few million diffraction patterns during the course of an experiment. Figs. 4–6 all suggest that the results improve with the number of diffraction patterns measured. More importantly, Figs. 6 and 7 suggest that the results improve quite significantly with greater incident flux. Both of these improvements become possible at future XFEL sources, so prospects of this method for the future are promising.

In experiments on uncrystallized biomolecules, scattering from the solution containing the biomolecules will likely be the dominant source of background noise, especially at low photon counts. At low resolution, the effect of the solvent can be modeled using Babinet's principle,²⁸ which states that the scattering from a uniform object is the same as the scattering from a hole of the same shape. However, for subnanometer resolutions, correlated scattering from closely packed solvent must be taken into consideration. Kirian *et al.* have outlined an approximate treatment of solvent scatter in Ref. 23 for the case of incoherent scattering, where they have shown that the SNR in the presence of solvent varies as $\frac{N}{N_s}$ of the SNR for scattering from biomolecules only. Here, $\frac{N}{N_s}$ is the ratio of protein to solvent molecules. Although the water background was not considered in the computation of the diffraction patterns here, the methods outlined here can already be applied to experiments where the molecular ensemble is injected with as little water as possible, for example, by using an aerodynamic lens particle injector.²⁹ Work is in progress on a detailed study of the effect of scattering from solution on the use of angular correlations for recovery of static as well as dynamic structures and will be the subject of a future report.

It must be noted that the simulations reported in this work have been done for sufficiently dilute particle concentration under which condition it is reasonable to neglect inter-particle interference fringes. For the case where N particles scatter coherently, the total diffracted intensity on a snapshot can be written as

$$I(\vec{q}) = \sum_{n=1}^N |F_n(\vec{q})|^2 + \sum_{n \neq m}^N F_n(\vec{q}) F_m^*(\vec{q}) e^{i\vec{q} \cdot \vec{r}_{nm}}, \quad (37)$$

where $F_n(\vec{q})$ is the complex structure factor of n^{th} particle, and \vec{r}_{nm} is the relative displacement between the particles. Since the particle positions and orientations are assumed random, it is expected that the interference term averages out when the correlations are calculated over a sufficiently large number of diffraction patterns.³⁰

Since PYP is a relatively small protein, only a few X-ray photons are scattered per molecule which is challenging for this method. With larger macromolecules, the number of scattered photons increases.³¹ In this case, meaningful difference densities may be already calculated from fewer diffraction pattern. Archetypal photoreactive molecules are those large molecules involved in photosynthesis. In a recent study, structural changes were obtained from a bacterial

photosynthetic reaction center using molecular dynamics simulations constrained by time-resolved difference SAXS/WAXS data.³² With an adaptation of our method, we may be able to directly invert the pump-probe diffraction patterns to difference electron densities

ACKNOWLEDGMENTS

The authors acknowledge funding support for this work from the National Science Foundation (Grant Nos. MCB-1158138 and STC-1231306), and the UWM High Performance Computing Center (HPC) for the use of the Avi cluster of networked computers. We are grateful to Dr. R. Kirian for useful discussions.

- ¹K. Moffat, "Time-resolved macromolecular crystallography," *Annu. Rev. Biophys. Biophys. Chem.* **18**, 309 (1989).
- ²M. Schmidt, "Structure based enzyme kinetics by time-resolved x-ray crystallography," *Ultrashort Laser Pulses in Medicine and Biology*, edited by W. Zinth, M. Braun, and P. Gilch (Springer, Berlin, New York, Germany, 2008).
- ³W. A. Barletta, J. Bisognano, J. N. Corlett, P. Emma, Z. Huang, K. J. Kim, R. Lindberg, J. B. Murphy, G. R. Neil, D. C. Nguyen, C. Pellegrini, R. A. Rimmer, F. Sannibale, G. Stupakov, R. P. Walker, and A. A. Zholents, "Free electron lasers: Present status and future challenges," *Nucl. Instrum. Methods Phys. Res., Sect. A* **618**, 69 (2010).
- ⁴P. Emma *et al.*, "First lasing and operation of an angstrom-wavelength free-electron laser," *Nat. Photonics* **4**, 641–647 (2010).
- ⁵K. Tanaka and T. Shintake, "SCSS x-fel conceptual design report," RIKEN Harima Institute, Hyogo, Japan, 2005.
- ⁶J. C. H. Spence, U. Weierstall, and H. N. Chapman, "X-ray lasers for structural and dynamic biology," *Rep. Prog. Phys.* **75**, 102601 (2012).
- ⁷R. Neutze and K. Moffat, "Time-resolved structural studies at synchrotrons and x-ray free electron lasers: Opportunities and challenges," *Curr. Opin. Struct. Biol.* **22**, 651–659 (2012).
- ⁸H. N. Chapman *et al.*, "Femtosecond x-ray protein nano crystallography," *Nature* **470**, 73–81 (2011).
- ⁹J. Tenboer *et al.*, "Time-resolved serial crystallography captures big-resolution intermediates of photoactive yellow protein," *Science* **346**, 1242–1246 (2014).
- ¹⁰A. McPherson, *Crystallization of Biological Macromolecules* (Cold Spring Harbor Laboratory Press, NY, 1999).
- ¹¹M. Schmidt, V. Srajer, R. Hening, H. Ihee, N. Purwar, J. Tenboer, and S. Tripathi, "Protein energy landscapes determined by five-dimensional crystallography," *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **69**, 2534 (2013).
- ¹²O. F. Lange, "Determination of solution structures of proteins up to 40 kda using cs-rosetta with sparse nmr data from deuterated samples," *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10873–10878 (2012).
- ¹³C. M. T. Spahn and P. A. Penczek, "Exploring conformational modes of macromolecular assemblies by multi particle cryo-em," *Curr. Opin. Struct. Biol.* **19**, 623–631 (2009).
- ¹⁴R. M. Glaser and R. J. Hall, "Reaching the information limit in cryo-em of biological macromolecule: Experimental aspects," *Biophys. J.* **100**, 2331–2337 (2011).
- ¹⁵Z. Kam, "Determination of macromolecular structure in solution by spatial correlation of scattering fluctuations," *Macromolecules* **10**, 927–934 (1977).
- ¹⁶D. Svergun and H. B. Stuhmann, "New developments in direct shape determination from small-angle scattering 1. Theory and model calculations," *Acta Crystallogr., Sect. A: Found. Crystallogr.* **47**, 736–744 (1991).
- ¹⁷D. K. Saldin, V. L. Shneerson, R. Fung, and A. Ourmazd, "Structure of isolated biomolecules obtained from ultrashort x-ray pulses: Exploiting the symmetry of random orientations," *J. Phys.: Condens. Matter* **21**, 134014 (2009).
- ¹⁸H. C. Poon, P. Schwander, M. Uddin, and D. K. Saldin, "Fiber diffraction without fibers," *Phys. Rev. Lett.* **110**, 265505 (2013).
- ¹⁹H. Liu, B. K. Poon, A. J. E. M. Janssen, and P. H. Zwart, "Computation of fluctuation scattering profiles via three-dimensional zernike polynomials," *Acta Crystallogr., Sect. A: Found. Crystallogr.* **68**, 561–567 (2012).
- ²⁰H. Liu, B. K. Poon, D. K. Saldin, J. C. H. Spence, and P. H. Zwart, "Three-dimensional single-particle imaging using angular correlations from x-ray laser data," *Acta Crystallogr., Sect. A: Found. Crystallogr.* **69**, 365–373 (2013).
- ²¹K. Pande, P. Schwander, M. Schmidt, and D. K. Saldin, "Deducing fast electron density changes in randomly oriented uncrystallized biomolecules in a pump-probe experiment," *Philos. Trans. R. Soc., B* **369**, 20130332 (2014).
- ²²S. Krinsky and R. Gluckstern, "Analysis of statistical correlations and intensity spiking in the self-amplified spontaneous-emission free-electron laser," *Phys. Rev. Spec. Top.—Accel. Beams* **6**, 050701 (2003).
- ²³R. A. Kirian, K. E. Schmidt, X. Wang, R. B. Doak, and J. C. H. Spence, "Signal, noise, and resolution in correlated fluctuations from snapshot small angle x-ray scattering," *Phys. Rev. E* **84**, 011921 (2011).
- ²⁴L. Lovisolo and E. A. B. da Silva, "Uniform distribution of points on a hyper-sphere with applications to vector bit-plane encoding," *IEE Proc. Vis. Image Signal Process.* **148**, 187–193 (2001).
- ²⁵P. C. Hansen, "The truncated svd as a method for regularization," *BIT Numer. Math.* **27**, 534–553 (1987).
- ²⁶K. Pearson, "Notes on regression and inheritance in the case of two parents," *Proc. R. Soc. London* **58**, 240–242 (1895).
- ²⁷G. Geloni, E. Saldin, L. Samoylova, E. Schneidmiller, H. Sinn, Th. Tschentscher, and M. Yurkov, "Coherent properties of the European XFEL," *New J. Phys.* **12**, 035021 (2010).
- ²⁸M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Oxford, 1980).
- ²⁹M. M. Siebert *et al.*, "Single mimivirus particles intercepted and imaged with an x-ray laser," *Nature* **470**, 78–82 (2011).
- ³⁰P. H. Zwart, private communication (2015).
- ³¹R. Fung, V. Shneerson, D. K. Saldin, and A. Ourmazd, "Structure from fleeting illumination of faint spinning objects in flight," *Nat. Phys.* **5**, 64–67 (2009).
- ³²D. Armlund *et al.*, "Visualizing a protein quake with time-resolved x-ray scattering at a free electron laser," *Nat. Methods* **11**, 923 (2014).