Computation of Fluctuation Scattering profiles via 3D Zernike Polynomials

Haiguang Liu,[a,c] Billy K. Poon,[a] Augustus J.E.M. Janssen[b] and Peter H. Zwart [a*]

[a]Physical Biosciences Division, Lawrence Berkeley National Laboratories, One Cyclotron Road, Berkeley CA 94720 USA, [b]Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, 5600MB Eindhoven, The Netherlands, and [c]Current address: Physics Department, Arizona State University, Tempe, AZ 85287 USA. E-mail: PHZwart@lbl.gov

Abstract

Ultrashort X-ray pulses from Free Electron Laser (FEL) X-ray sources make it feasible to conduct small and wide angle scattering experiments on biomolecular samples in solution at sub-picosecond time scales. During these so-called fluctuation scattering experiments, the absence of rotational averaging, typically induced by Brownian motion in classic solution scattering experiments, increases the information content of the data. In order to perform shape reconstruction or structure refinement from such data, it is essential to compute the theoretical profiles from 3D models. Based on the 3D Zernike polynomial expansion models, we derived a fast method to compute

the theoretical fluctuation scattering profiles. The theoretical profiles have been validated against simulated results obtained from 300,000 scattering patterns for several representative biomolecular species.

1. Introduction

The overwhelming majority of known protein structures at the atomic scale have been determined by X-ray crystallography. The availability of these high-resolution structural models is the foundation of a deeper understanding of fundamental processes in biology (Orengo *et al.*, 1999), and the development of new therapeutic drugs (Schneider & Fechner, 2005) and novel classes of nano-materials (Barth *et al.*, 2005) . Due to difficulties in obtaining crystals of large macromolecular complexes, like membrane proteins or large molecular machines, X-ray crystallography is rarely the technique of choice for deriving structure information of such biomolecules or complexes (Miao *et al.*, 2008). Current techniques used to study large macromolecular complexes, like electron microscopy (Medalia *et al.*, 2002; Frank, 2002), derive structural information of complexes in non-native environments and do not easily allow for the investigation of time-dependent, large-scale structural changes. Techniques like small- and wide-angle X-ray scattering (SAXS/WAXS) are suitable and allow for time-resolved studies, but have the drawback that the data have relatively low information content (Volkov & Svergun, 2003). A possible route for increasing the information content in solution scattering experiments while monitoring the large-scale structural changes of macromolecular complexes in aqeuous environments is fluctuation X-ray scattering (fXS), a method proposed by Kam in the late 1970's (Kam, 1977; Kam *et al.*, 1981). A fluctuation X-ray scattering experiment is performed by collecting scattering patterns of a dilute sample of scatterers at exposure times shorter than the time required for particles to reorient themselves via rotational diffusion. These

experiments are ideally performed on free electron lasers such as the Linac Coherent Light Source (LCLS; Emma *et al.* (2010)), the European X-ray Free Electron Laser (XFEL; Vartanyants *et al.* (2007)), or the (to be constructed) Next Generation Light Source (NGLS; Schoenlein (2011)). Under aforementioned conditions, the measured scattering patterns will no longer be angularly isotropic, but will contain speckles. It can be shown that a large number of these scattering patterns can be used to estimate the average angular autocorrelation of the scattering pattern of a single particle (Kam *et al.*, 1981; Saldin *et al.*, 2010; Saldin *et al.*, 2011).

Earlier work by Saldin *et al.* (2009) has shown that the angular autocorrelations can be used in the determination of low resolution structural envelopes in a manner similar to which SAXS data are used to determine structural envelopes (Saldin *et al.*, 2009). By parametrizing a macromolecular envelope with spherical harmonics, a straightforward procedure is obtained that allows for the optimization of real space expansion coefficients given observed angular correlations. Although the spherical harmonics expansion used to parameterize the macromolecular structure has been widely used, its main drawback is that it cannot properly describe shapes containing cavities or pores. A more appropriate polynomial expansion, capable of describing complex shapes not limited to star-shaped objects, can be obtained by using 3D Zernike Polynomials (Novotni & Klein, 2003; Mak *et al.*, 2008; Liu *et al.*, 2012*b*). 3D Zernike polynomials are extensions of the well-known 2D Zernike polynomials frequently used in describing optics (Wang & Silva, 1980) and have similar properties. A basic introduction to 3D Zernike polynomials can be found in Canterakis (1999).

3D Zernike polynomials are useful tools in solution scattering because with a relatively low expansion order (20, for instance), one is able to faithfully reproduce a macromolecular shape and its associated scattering function. This fact has lead to novel ways to compute SAXS profiles (Liu *et al.*, 2012*b*) and accelerated shape recov-

ery from SAXS data (unpublished). In this article, we report an economic route of computing the 2-point angular autocorrelations, $C_2(q, \Delta\phi)$, from the Zernike expansion coefficients. This work is a first step towards *ab initio* structure solution from fluctuation scattering data at low to medium resolution ranges. In a forthcoming paper, details on how to utilize this method in 3D model reconstruction will be reported.

2. Methods

The average angular autocorrelation can be extracted from N experimental patterns, acquired in femtosecond X-ray scattering experiments, via

$$C_2(q, \Delta\phi) = \frac{1}{N} \sum_i \sum_\phi I_i(q, \phi) I_i(q, \phi + \Delta\phi) \tag{1}$$

where $I_i(q, \phi)$ is the scattering intensity of pattern i at a pixel position corresponding to reciprocal space point (q, ϕ). Kam has shown that, given a sufficient number of scattering patterns, this autocorrelation will converge to a fixed value (Kam *et al.*, 1981). The expected value of $C_2(q, \Delta\phi)$ can be computed from the associated molecular structure as outlined below.

2.1. 3D Zernike polynomial models

The 3D Zernike model is a compact description of 3D models using convenient orthogonal polynomials. After scaling down to fit in a unit sphere, the electron density of a 3D model can be represented by

$$\rho(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{l=0}^{n} \sum_{m=-l}^{l} c_{nlm} R_{nl}(r) Y_{lm}(\omega_r) \tag{2}$$

in which $\mathbf{r}$ is a 3D vector (r, ω_r). The original 3D model can be obtained by scaling up the unit sphere by r_{max}, the radius of a sphere circumscribing the model. $R_{nl}(r)$ is the 3D Zernike Polynomial radial function, and the orders l and n have the same parity, i.e., $(n - l)$ is an even number.

The expansion coefficients, c_{nlm}, can be computed via the following procedure: the 3D model is firstly scaled down to fit into a unit sphere by dividing the r_{max} to obtain the scaled density distribution function, $\rho(\mathbf{r})$; and then the c_{nlm} coefficients can be computed utilizing the orthogonality,

$$c_{nlm} \quad = \quad \frac{3}{4\pi} \int_{|\mathbf{r}|<1} \rho(\mathbf{r})R_{nl}(r)Y^*_{lm}(\mathbf{r})d\mathbf{r} \tag{3}$$

The normalization again r_{max} is necessary because the Zernike polynomials are orthogonal only in the unit sphere. Novotni & Klein (2003) have derived a more efficient way of computing the Zernike moments using the geometry moments, instead of the direct integration.

It is worthwhile to point out that the computation of the Zernike Moments, c_{nlm}, are the time-limiting step. The details of the implementation can be found in the original papers (Novotni & Klein, 2003; Mak *et al.*, 2008). In short, the PDB models are mapped to 3D grids to generate a voxelized object, and from which the c_{nlm} are subsequently calculated. The grid spacing is set to 0.7 Å, so the number of non-zero voxel is proportional to the number of atoms. For very large molecular systems, the grid size can be increased to speed up the computation. This makes the Zernike approximation advantageous over other methods, especially when high resolution data is not required or available.

Following Liu *et al.* (2012b) and Appendix A, the Fourier transform of this model is equal to

$$A(\mathbf{q}) \quad = \quad 4\pi \sum_{n}^{\infty} \sum_{l}^{n} \sum_{m=-l}^{+l} i^l(-1)^{(n-l)/2}c_{nlm}Y^*_{lm}(\omega_q)b_n(qr_{max}) \tag{4}$$

with

$$b_n(qr_{max}) = \frac{j_n(qr_{max}) + j_{n+2}(qr_{max})}{2n+3} \tag{5}$$

We now aim to expand $A(\mathbf{q})$ in shells of fixed q, and only vary the angles.

$$A_q(\omega_q) = 4\pi \sum_n^\infty \sum_l^n \sum_{m=-l}^{+l} i^l(-1)^{(n-l)/2}b_n(qr_{max})c_{nlm}Y_{lm}^*(\omega_q)$$

$$= 4\pi \sum_n^\infty \sum_l^n \sum_{m=-l}^{+l} w_{nl}(qr_{max})c_{nlm}Y_{lm}^*(\omega_q) \qquad (6)$$

$$w_{nl}(qr_{max}) = i^l(-1)^{(n-l)/2}b_n(qr_{max}) \qquad (7)$$

If we now regroup constants and set the maximum expansion order to n_{max}, we get

$$A_q(\omega_q) = 4\pi \sum_l^{n_{max}} \sum_{m=-l}^{+l} Y_{lm}^*(\omega_q) \sum_n^{n_{max}} w_{nl}(qr_{max})c_{nlm}$$

$$(8)$$

or

$$A_q(\omega_q) = 4\pi \sum_l^{n_{max}} \sum_{m=-l}^{+l} a_{lm}(qr_{max})Y_{lm}^*(\omega_q)$$

$$a_{lm}(qr_{max}) = \sum_n^{n_{max}} w_{nl}(qr_{max})c_{nlm} \qquad (9)$$

The outlined route to compute coefficients a_{lm} using available Zernike moments can be seen to depend linearly on expansion order n_{max}. This is in contrast to the traditional method (Kam, 1977) when a_{lm} is computed directly from the atomic coordinates (Liu *et al.*, 2012*b*, expression 6) and depends linearly on the number of atoms N_{atoms}:

$$a_{lm}(q) = \sum_{j=1}^{N_{atoms}} f_j(q)j_l(qr_j)Y_{lm}^*(\omega_j) \qquad (10)$$

2.2. Intensities

The relation between the Zernike coefficients to the complex expansion coefficients for the structure factor at any fixed resolution, q, can be utilized to get spherical harmonics based expansion coefficients for the intensity at this fixed resolution.

$$I_q(\omega_q) = A_q^*(\omega_q)A_q(\omega_q)$$

$$= 16\pi^2 \sum_l^{n_{max}} \sum_{m=-l}^{+l} \sum_{l'}^{n_{max}} \sum_{m'=-l'}^{+l'} a_{l'm'}^*(qr_{max})a_{lm}(qr_{max})Y_{lm}^*(\omega_q)Y_{l'm'}(\omega_q)$$

(11)

Although the above expression is not overtly complicated, we need to re-expand it in a spherical harmonics series:

$$I_q(\omega_q) = 16\pi^2 \sum_l^{n_{max}} \sum_{m=-l}^{+l} I_{lm}(q)Y_{lm}(\omega_q)$$

(12)

The expansion coefficients, $I_{lm}(q)$, can be expressed as a function of $a_{lm}(qr_{max})$ via a Gaunt series (Gaunt, 1928), using:

$$\int Y_{lm}Y_{l'm'}Y_{l''m''}d\Omega = G_{ll'l''}^{mm'm''}$$

(13)

with

$$G_{ll'l''}^{mm'm''} = (-1)^m \sqrt{\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}} \times \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix} \times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}$$

(14)

One thus obtains

$$I_{lm}(q) = \sum_{l'} \sum_{l''} \sum_{m'} \sum_{m''} a_{l'm'}(qr_{max})a_{l''m''}(qr_{max})G_{ll'l''}^{mm'm''}$$

(15)

where $\begin{pmatrix} l & l' & l'' \\ m & m' & m'' \end{pmatrix}$ and $\begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}$ are Wigner 3j-symbols.

The latter expansion facilitates easy computation of the autocorrelation (Saldin *et al.*, 2009) as shown in the the next section.

2.3. The expansion of autocorrelations

According to Saldin *et al.* (2009) and following their notation, the correlation function, C_2, at a given q value can be expressed as a weighted sum of Legendre polyno-

mials:

$$C_{2,q}(\Delta\phi) \quad = \quad \sum_l F_l(\Delta\phi)B_l(q) \tag{16}$$

with

$$F_l(\Delta\phi) \quad = \quad \frac{1}{4\pi}P_l\left[\cos^2\theta(q) + \sin^2\theta(q)\cos(\Delta\phi)\right]$$
$$B_l(q) \quad = \quad \sum_m |I_{lm}(q)|^2 \tag{17}$$

Where $P_l(\cdot)$ denotes a Legendre polynomial and

$$\theta(q) = \pi/2 - \sin^{-1}(q/2\kappa) \tag{18}$$

κ is equal to the wavenumber $2\pi/\lambda$ with λ the wavelength of the incident radiation. The expansion coefficients $B_l(q)$, can thus be computed directly from the Zernike moments by considering relations (9), (15) and (17).

3. Results

In this section, the theoretical fXS profiles calculated using the 3D Zernike polynomial method is compared to the results calculated using the spherical harmonics approach, as well as the fXS profiles extracted from simulated scattering patterns of single molecules. After the validation of the method, we will show the relation between the resolution and the Zernike expansion order. The computing complexity and speed will also be discussed.

3.1. Validation

In order to verify the above expressions, scattering patterns of proteins in random orientations were simulated up to q values of 0.5 $\mathring{A}^{-1}$ as described in Appendix B. Besides model data generated via expressions (9), (15) and (17), $B_l(q)$ were computed via a spherical harmonics expansion of the intensities (Saldin *et al.*, 2009). Using three

different example proteins, the fXS profiles obtained from three distinct approaches are compared in Figure 1. The $B_l(q)$ curves obtained with the Zernike expansion are in excellent agreement with the curves calculated using the spherical harmonics approach, revealing the theoretical fXS profiles under ideal conditions (noise-free with infinitely small pixel size). The agreement between the Zernike-based method and the data obtained via numerical simulation is satisfactory at low resolution but quickly gets worse as the resolution increases. The discrepancy between the theoretical fXS profiles and the data extracted from the simulated diffraction patterns might be due to the finite resolution of the detectors (see Discussion and Conclusions).

3.2. Expansion order

The effect of the 3D Zernike expansion order on the calculated data is shown in Figure 2. The truncation effects depicted in Figure 2 are comparable to those seen when computing SAXS profiles as described by Liu *et al.* (2012*b*). As described in the previous study, the resolution at which truncation ripples appear is largely a function of the expansion order, n_{max}, and size, R_{max}, of the particle due to the scale-free nature of the Zernike polynomials (Novotni & Klein, 2003; Mak *et al.*, 2008; Liu *et al.*, 2012*b*).

3.3. Complexity and timing

As discussed elsewhere (Liu *et al.*, 2012*b*), the complexity of the Zernike method for computing SAXS data scales linearly with the number of atoms (N) and is separately dependent of the number of data points computed (M); therefore, the computing complexity is $O(M + N)$. As mentioned, the time-limiting step is the computation of the Zernike moments, and the dependence on M is due to the precomputation of spherical bessel functions. The resulted computing complexity can be approximated by

$O(N)$, because the computation of bessel functions only compromises a small overhead on the total runtime. On the other hand, the spherical harmonics based method, depends linearly on the number of atoms and the number of data points computed, resulting a complexity of $O(NM)$. For computing fluctuation scattering data, both the spherical harmonics based method and the Zernike based method show linear dependence on the number of data points and number of atoms, Figure 3a. The reason for the linear dependence when computing $B_l(q)$ curves using the Zernike method can be seen from equation (9) in which a q-dependent weight is computed. As can be seen from equation (9), the 3D Zernike moments c_{nlm} need only be computed once for a given model. In contrast, the q-dependent expansion coefficients, $a_{lm}(q)$, as obtained from a spherical harmonics expansion (see equation 10) depends on the number of atoms. This reduction in numerical complexity provides a significant time advantage in the calculation of $B_l(q)$ coefficients via the Zernike method. These effects are illustrated in Figure 3b.

4. Discussions and Conclusions

The results from Figure 1 indicate that the proposed route for computing model fXS data is effective as the curves obtained via simulation, spherical harmonics and the Zernike based method are all in agreement. The main benefit of the proposed method is the gain in computational efficiency, a prerequisite for real-time structure refinement or iterative real-reciprocal space *ab-initio* structure determination methods.

Although a notable difference between simulated and calculated data is found, especially at higher q-values, the discrepancy can be attributed to smearing of data due to finite pixel size as well as sampling errors and other experimental phenomena.

The timing results indicate that the Zernike based method is more efficient than the spherical harmonics based method because a time consuming step that depends on

the number of atoms (caluclating 3D Zernike moments) is performed only once. For the spherical harmonics based method, a loop over all atoms has to be performed for each data point, making this method more computationally costly. This is especially critical when applying a local perturbation approach for structural reconstruction or refinement, where only part of the system will be changed at a time. In this case, since the Zernike moments are a summation of contributions from all parts, only the perturbed part of system needs to be updated to obtain the fXS profiles for the newly generated model. Although this principle applies to the spherical harmonics based method as well (see Eq 10), the memory requirements and computational efficiency for the Zernike method are more favourable.

Furthermore, the proposed Zernike method can, in principle, be improved by removing the q-dependence. This can be accomplished by obtaining a series expansion form of $I_{lm}(q)$ that depends only on the 3D Zernike moments and r_{max}, likely along the lines of work outlined by Pavelcik *et al.* (2002). For the SAXS intensity, which is closely related to $B_0(q)$, this expression is quite straightforward (Liu *et al.*, 2012*b*), indicating that for higher orders of l, similar simple expressions can be found. This is beyond the scope of this work.

To conclude, an efficient method based on 3D Zernike polynomials for fluctuation X-ray scattering profile calculations is presented. The performance and accuracy is validated against the data obtained from the spherical harmonics expansion method and simulation results. This method provides an opportunity in *ab initio* model reconstruction based on the fXS profiles, which are embedded in the fluctuation X-ray scattering experimental data.

Appendix A
Fourier transform of 3D Zernike Radial functions

Expression 17 in Liu *et al.* (2012*b*) provides details of the Fourier transform of a Zernike polynomial. A formal proof of the expression (40) as stated by Mathar (2008) could not be found and is clarified below. It is required to show that

$$I_{nl}(q) = \int_0^1 j_l(qr)R_{nl}(r)r^2 dr = (-1)^{\frac{n-l}{2}}\sqrt{2n+3}\frac{j_{n+1}(q)}{q}$$

(19)

A concise argument for the validity of (19) follows. We have from Abramowitz & Stegun (1972), 10.1.1 and 9.1.10

$$j_l(z) = \Gamma(\frac{3}{2})(\frac{1}{2}z)^l \sum_{k=0}^{\infty} \frac{(-\frac{1}{4}z^2)^k}{k!\Gamma(k+l+\frac{3}{2})}$$

(20)

using this in (19) yields

$$\begin{aligned} I_{nl}(q) &= \Gamma(\frac{3}{2})(\frac{1}{2}q)^l \sum_{k=0}^{\infty} \frac{(-\frac{1}{4}z^2)^k}{k!\Gamma(k+l+\frac{3}{2})} \\ &\times \int_0^1 r^{l+2k+2} R_{nl}(r) dr \end{aligned}$$

(21)

To evaluate the integrals on the right-hand side of (21), we use the definition of $R_{nl}(r)$, see Eq. (43) in Mathar (2009),

$$R_{nl}(r) = \sqrt{2n+3}\ r^l P_p^{(0,l+\frac{1}{2})}(2r^2-1), p = \frac{1}{2}(n-l)$$

(22)

where $P_j^{(\alpha,\beta)}(x)$ is the general Jacobi polynomial. With the substitution $x = 2r^2 - 1$, we then obtain

$$\begin{aligned} \int_0^1 r^{l+2k+2} R_{nl}(r) dr &= \sqrt{2n+3}\ 2^{-l-k-\frac{5}{2}} \\ &\times \int_{-1}^1 (1+x)^{k+l+\frac{1}{2}} P_p^{(0,l+\frac{1}{2})}(x) dx \end{aligned}$$

(23)

Next, using Rodrigues' formula in Abramowitz & Stegun (1972) (Eq. 21.11.1), so that

$$(1+x)^{l+\frac{1}{2}} P_p^{(0,l+\frac{1}{2})}(x) = \frac{(-1)^p}{2^p p!}(\frac{d}{dx})^p[(1-x)^p(1+x)^{p+l+\frac{1}{2}}]$$

(24)

Upon p partial integrations in (23), one obtains

$$
\begin{aligned}
\int_0^1 r^{l+2k+2} R_{nl}(r)dr &= \sqrt{2n+3}\, 2^{-p-l-k-\frac{5}{2}} \frac{k!}{p!(k-p)!} \\
&\times \int_{-1}^1 (1-x)^p(1+x)^{k+l+\frac{1}{2}}dx
\end{aligned}
\tag{25}
$$

when $k \geq p$, and (25) vanishes when $0 \leq k \leq p$. The remaining integral can be expressed in terms of the Beta integral as in Abramowitz & Stegun (1972) (Eq. 6.2). Using the substitution $t = \frac{1}{2}(1+x)$, the integral becomes

$$
\int_0^1 r^{l+2k+2} R_{nl}(r)dr = \frac{1}{2}\sqrt{2n+3}\frac{k!\Gamma(k+l+\frac{3}{2})}{(k-p)!\Gamma(p+k+l+\frac{5}{2})}
\tag{26}
$$

for $k \geq p$, while (26) vanishes for $0 \leq k \leq p$. Using this in (21) gives

$$
I_{nl}(q) = \frac{1}{2}\Gamma(\frac{3}{2})(\frac{1}{2}q)^l \sum_{k=p}^{\infty} \frac{(-\frac{1}{4}q^2)^k}{(k-p)!\Gamma(p+k+l+\frac{5}{2})}
\tag{27}
$$

Replacing k by $(j+p), j = 0, 1, ...,$ one obtains

$$
\begin{aligned}
I_{nl}(q) &= (-1)^p \frac{1}{2}\Gamma(\frac{3}{2})(\frac{1}{2}q)^{l+2p} \\
&\times \sqrt{2n+3}\sum_{j=0}^{\infty}\frac{(-\frac{1}{4}q^2)^j}{j!\Gamma(j+l+2p+\frac{5}{2})}
\end{aligned}
\tag{28}
$$

Then (19) follows from (20) and the fact that $n = l + 2p$.

Appendix B
Simulation of Scattering Data

Scattering data was simulated for molecules in vacuum, without any sources of noise. Using atomistic models from the Protein Data Bank (PDB; http://www.pdb.org), a single molecule was randomly rotated and placed in a 1 μm^3 box. A full scattering image, using physically relevant detector parameters that captured q values up to 0.5 $\mathring{A}^{-1}$, was then calculated via direct summation,

$$
A(\mathbf{q}) = \sum_i^{n_{atoms}} f_i(\mathbf{q})\exp(i\mathbf{q}\cdot\mathbf{r_i})
$$

$$I(\mathbf{q}) \;=\; |A(\mathbf{q})|^2 \tag{29}$$

where f_i and $\mathbf{r_i}$ are the scattering form factor and coordinate of atom i, respectively. Pixels in the image were then binned for specific q and ϕ values, and the autocorrelations calculated with (1). Simulations were repeated for 300,000 images to ensure convergence of the autocorrelations.

After the autocorrelations were calculated from the simulated images, a series of Legendre polynomials was used to fit the autocorrelations according to (16). The coefficients were calculated with the general expression for expanding any function with a Fourier-Legendre series,

$$
\begin{aligned}
B_l(q) &= \frac{2l+1}{2} \int_{-1}^{1} P_l(x) f(x) dx \\
x &= \cos^2 \theta(q) + \sin^2 \theta(q) \cos(\Delta\phi)
\end{aligned}
\tag{30}
$$

where l is the order of the Legendre polynomial, $f(x)$ is the autocorrelation function, $C_2(q, \Delta\phi)$, and x is used to map $\Delta\phi$ onto [-1,1] for the Legendre polynomial, P_l,. Since the integral in (30) was evaluated numerically, the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm was used to minimize the least squares difference between the simulated autocorrelation and its series expansion form.

All simulations and calculations were implemented in Python and C/C++ using the Small Angle Scattering Toolbox (Liu *et al.*, 2012*a*) as part of the Computational Crystallography Toolbox (http://cctbx.sourceforge.net) (Grosse-Kunstleve *et al.*, 2002). Source code is available upon request and from *http://sastbx.als.lbl.gov*.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. New York: Dover Publications.

Barth, J. V., Costantini, G. & Kern, K. (2005). *Nature*, **437**(7059), 671–679.

Canterakis, N. (1999). In *Scandinavian Conference on Image Analysis*.

Emma, P., Akre, R., Arthur, J., Bionta, R., Bostedt, C., Bozek, J., Brachmann, A., Bucksbaum, P., Coffee, R., Decker, F. J., Ding, Y., Dowell, D., Edstrom, S., Fisher, A., Frisch, J., Gilevich, S., Hastings, J., Hays, G., Hering, P., Huang, Z., Iverson, R., Loos, H., Messerschmidt, M., Miahnahri, A., Moeller, S., Nuhn, H. D., Pile, G., Ratner, D., Rzepiela, J., Schultz, D., Smith, T., Stefan, P., Tompkins, H., Turner, J., Welch, J., White, W., Wu, J., Yocky, G. & Galayda, J. (2010). *Nat Photon*, **4**(9), 641–647.

Frank, J. (2002). *Annual Review of Biophysics and Biomolecular Structure*, **31**, 303–319.

Gaunt, J. (1928). *Proceedings of the Royal Society of London Series A-Containing Papers of a Mathematical and Physical Character*, **122**(790), 513–532.

Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W. & Adams, P. D. (2002). *Journal of Applied Crystallography*, **35**(1), 126–136.

Kam, Z. (1977). *Macromolecules*, **10**(5), 927–934.

Kam, Z., Koch, M. & Bordas, J. (1981). *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **78**(6), 3559–3562.

Liu, H., Hexemer, A. & Zwart, P. H. (2012*a*). *Journal of Applied Crystallography*, p. in press.

Liu, H., Morris, R. J., Hexemer, A., Grandison, S. & Zwart, P. H. (2012*b*). *Acta Crystallographica Section A*, **68**(2).

Mak, L., Grandison, S. & Morris, R. J. (2008). *J Mol Graph Model*, **26**(7), 1035–45.

Mathar, R. J. (2008). *Baltic Astronomy*, **17**(3-4), 383–398.

Mathar, R. J. (2009). *Serbian Astronomical Journal*, **179**, 107–120.

Medalia, O., Weber, I., Frangakis, A., Nicastro, D., Gerisch, G. & Baumeister, W. (2002). *Science*, **298**(5596), 1209–1213.

Miao, J., Ishikawa, T., Shen, Q. & Earnest, T. (2008). *Annual Review of Physical Chemistry*, **59**, 387–410.

Novotni, M. & Klein, R. (2003). In *Proceedings of the eighth ACM symposium on Solid modeling and applications*, SM '03, pp. 216–225. New York, NY, USA: ACM.

Orengo, C. A., Todd, A. E. & Thornton, J. M. (1999). *Current Opinion in Structural Biology*, **9**(3), 374 – 382.

Pavelcik, F., Zelinka, J. & Otwinowski, Z. (2002). *Acta Crystallographica Section D*, **58**(2), 275–283.
URL: *http://dx.doi.org/10.1107/S0907444901019710*

Saldin, D. K., Poon, H. C., Bogan, M. J., Marchesini, S., Shapiro, D. A., Kirian, R. A., Weierstall, U. & Spence, J. C. H. (2011). *Phys Rev Lett*, **106**(11), 115501.

Saldin, D. K., Shneerson, V. L., Fung, R. & Ourmazd, A. (2009). *J Phys Condens Matter*, **21**(13), 134014.

Saldin, D. K., Shneerson, V. L., Howells, M. R., Marchesini, S., Chapman, H. N., Bogan, M., Shapiro, D., Kirian, R. A., Weierstall, U., Schmidt, K. E. & Spence, J. C. H. (2010). *New Journal of Physics*, **12**, 035014.

Schneider, G. & Fechner, U. (2005). *Nat Rev Drug Discov*, **4**(8), 649–63.

Schoenlein, R. (2011). In *Frontiers in Optics*, p. FMC2. Optical Society of America.

Vartanyants, I. A., Robinson, I. K., McNulty, I., David, C., Wochner, P. & Tschentscher, T. (2007). *J Synchrotron Radiat*, **14**(Pt 6), 453–70.

Volkov, V. & Svergun, D. (2003). *Journal of Applied Crystallography*, **36**, 860–864.

Wang, J. Y. & Silva, D. E. (1980). *Appl Opt*, **19**(9), 1510–8.

Fig. 1. **Comparisons of fluctuation scattering profiles.** The fXS profiles are calculated using three methods: Zernike-based method (solid curve), spherical harmonics approach (open square), and extracted from simulations (cross). For clarity, only L=0 and L=8 curves are shown as representatives here. Curves depicted are for models (A) 1EE2, (B) 2LAO and (C) 2E2G. The theoretical curves from Zernike-based method agree with the spherical harmonics based method. In spite of discrepancies between the simulations results and the theoretical curves, the main features are preserved. The expansion order n_{max} was set to 40.

Fig. 2. **Model resolution dependence on expansion order.** Accurate modeling of higher resolution fXS profile requires more polynomials, i.e., a higher Zernike expansion order. The donut shaped protein (PDB ID: 2E2G) is used as an example to demonstrate the relationships between real space resolution, reciprocal space scattering vector, and the maximum expansion order. A higher expansion order needs to be used when truncation errors are observed.

Fig. 3. **Computing complexity and speed.** The computing time comparison for the spherical harmonic and Zernike based approaches. Both methods depict a linear computational complexity with respect to the number of q-values in the dataset (left) as well as the number of atoms in the model (right). The speed increase for the Zernike based method (circles) is approximately a factor 70 as compared to the spherical harmonics based method (squares).

Synopsis

An algorithm for fast fluctuation scattering parameter calculation via 3D zernike polynomial expansion method.