

## Theoretical estimation for correlations of diffraction patterns from objects differently oriented in space

B. Ziaja<sup>a,\*</sup>, A.V. Martin<sup>a</sup>, F. Wang<sup>a</sup>, H.N. Chapman<sup>a,b</sup>, E. Weckert<sup>c</sup>

<sup>a</sup> Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, Notkestr. 85, D-22607 Hamburg, Germany

<sup>b</sup> Department of Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany

<sup>c</sup> Hamburger Synchrotronstrahlungslabor, Deutsches Elektronen-Synchrotron, Notkestr. 85, D-22607 Hamburg, Germany

### ARTICLE INFO

Available online 5 January 2011

#### Keywords:

Free-electron laser

Coherent diffraction imaging

Classification

### ABSTRACT

Coherent diffraction imaging of single biomolecules is expected to open unique opportunities for studies of non-crystalline samples. There are, however, still many technical and physical issues that need to be resolved in a more quantitative manner, especially if one aims for structural information at high resolution. Signal recorded from an object after a single shot is low. As primarily proposed in Spence and Doak (2004) and Huldt et al. (2003) [1,2], averaging over the diffraction patterns from many different shots is necessary, in order to achieve a signal-to-noise ratio sufficient for image reconstruction. The images of the randomly oriented molecules have to be sorted out in order to identify those corresponding to the similar spatial orientations of the objects. This procedure is called the classification of diffraction images. Here we approach the classification in the framework of pattern-to-pattern correlations, and analyse theoretically the correlations between diffraction images of differently oriented objects.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

X-ray free-electron lasers (FELs) are expected to open new horizons for structural studies of biological systems, especially for studies of non-crystalline samples, such as viruses or living cells. In general, radiation damage limits the accuracy of the structure determination of biological particles in standard diffraction experiments. However, computer simulations of damage formation have strongly suggested [3–7] that radiation tolerance may be extended to very high doses with ultrafast exposures, as will be possible with the presently operating and developing X-ray FELs, such as LCLS, SCSS, and the European XFEL. This is due to their photon pulses being of a shorter duration than the time atoms require to move a distance comparable to the required resolution. This improved radiation tolerance indicates the possibility of recording images of single biological particles at high resolution without the need to amplify scattered radiation through Bragg reflections. This application of FELs could have a tremendous impact on structural studies at both the molecular and cellular level, with profound implications for biology and medicine. Recent experiments performed at FLASH [8,9] have demonstrated the proof of this imaging principle.

There are, however, still many technical and physical issues that need to be resolved in a more quantitative manner, especially if one aims for structural information at high resolution. Signal recorded from an object after a single shot is low. As primarily proposed in Refs. [1,2], averaging over the diffraction patterns from many different shots is necessary, in order to achieve a signal-to-noise ratio sufficient for image reconstruction. The alternative way is to diffract light from many identical objects during one shot. In both cases it is crucial that the imaged objects have the same spatial orientation towards the laser beam.

If the molecules cannot be aligned, the images of randomly oriented molecules have to be sorted out in order to identify those corresponding to the similar spatial orientations of the objects. This procedure is called the classification of diffraction images. Various techniques of classification have been proposed so far [2,10–12]. Here we will restrict to the framework of simple pattern-to-pattern correlations. Correlations between diffraction images give an estimate of an ‘orientational distance’ between differently oriented objects. In what follows we address the theoretical question about the origin of the correlations between different patterns. In particular, we predict the Gaussian profile of the average correlation function which is a consequence of underlying Wilson statistics [13], and show that this profile is maintained in simulation of diffraction images obtained from a realistic biological sample (a small virus). Moreover, we demonstrate that the width of the Gaussian profile that corresponds to the angular (orientational) resolution is determined only by the

\* Corresponding author at: Institute of Nuclear Physics, Radzikowskiego 152, 31-342 Kraków, Poland.

E-mail address: [ziaja@mail.desy.de](mailto:ziaja@mail.desy.de) (B. Ziaja).

virus structure. It is unaffected by a random water layer surrounding the virus or a Poissonian noise. This indicates that the classification of diffraction images within the pattern-to-pattern scheme is not significantly perturbed by these effects, and therefore still possible.

## 2. Analysis of correlations between diffraction patterns

We consider angular correlations of two diffraction patterns from objects in different spatial orientations. Our aim is to show that a simple model based on random atom positions is sufficient to account for essential features of these angular correlations.

The intensity diffracted from an object can be calculated with the following formula:

$$I(\mathbf{q}) = r_e^2 \Phi_{in} \Omega_{pix} \left| \sum_{i=1}^N f_i(q) e^{i\mathbf{q} \cdot \mathbf{r}_i} \right|^2, \quad (1)$$

where  $\mathbf{q}$  is the wave-transfer vector;  $f_i$  are the atomic form factors for each element species;  $r_e$  is the classical radius of the electron;  $\Phi_{in}$  is the photon fluence, and  $\Omega_{pix}$  is the solid angle for one pixel.

In order to quantify the possible loss of orientational and structural information, the  $q$ -dependence of the correlation of 2D diffraction patterns for different relative angular object orientations,  $\alpha$ , is calculated by analogy to a similar analysis in Ref. [12], as

$$G(\alpha, \mathbf{q}) \equiv \frac{\tilde{I}(\alpha, \mathbf{q}) \cdot \tilde{I}(0, \mathbf{q})}{\sqrt{\langle \tilde{I}(\alpha, \mathbf{q})^2 \rangle} \cdot \sqrt{\langle \tilde{I}(0, \mathbf{q})^2 \rangle}}, \quad (2)$$

with  $\tilde{I}(\alpha, \mathbf{q}) = I(\alpha, \mathbf{q}) - \langle I(\alpha, \mathbf{q}) \rangle$ . The average is defined as  $\langle I(\alpha, \mathbf{q}) \rangle = (2\pi)^{-1} \int_0^{2\pi} d\phi_q I(\alpha, \mathbf{q})$ . The angle  $\alpha$  denotes the rotation angle around an arbitrary axis defined by a unit vector,  $\mathbf{n}$ . The angle,  $\phi_q$  is the azimuthal angle of vector  $\mathbf{q}$  in the  $\mathbf{q}$  plane. Fig. 1 shows the geometry. Let us assume that the object consists of  $N$  atoms, located at positions  $\mathbf{r}_i$ , where  $i = 1, \dots, N$ . For simplicity we consider a monoatomic object, where scattering factors:  $f_i(\mathbf{q}) \equiv f(q)$  for  $i = 1, \dots, N$ . The intensity of diffracted signal then reads:

$$I(\mathbf{q}) = |f(q)|^2 r_e^2 \Phi_{in} \Omega_{pix} \left| \sum_{i=1}^N e^{i\mathbf{q} \cdot \mathbf{r}_i} \right|^2, \quad (3)$$

and can be further rewritten:

$$I(\mathbf{q}) = |f(q)|^2 r_e^2 \Phi_{in} \Omega_{pix} \sum_{i,j=1}^N e^{i\mathbf{q} \cdot \mathbf{r}_{ij}}, \quad (4)$$

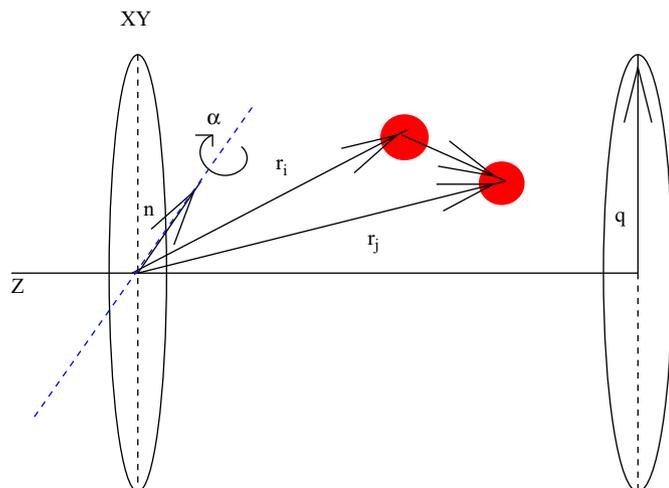


Fig. 1. Geometry of diffraction imaging. Beam axis corresponds to z-axis. Wave-transfer vector  $\mathbf{q}$  lies in the XY-plane (small angle approximation).

where  $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ , and the scalar product of  $\mathbf{q}$  and  $\mathbf{r}_{ij}$ :  $\mathbf{q} \cdot \mathbf{r}_{ij} = q \cdot r_{ij} \sin\theta_{ij} \cos(\phi_q - \phi_{ij})$ , is evaluated, using the spherical coordinates of the vector  $\mathbf{r}_{ij}$ .

Below we estimate the average correlation between the diffraction pattern of an object and the diffraction pattern of an object rotated around the beam axis (z-axis) by an angle  $\alpha$ . A similar derivation can be performed for the rotation by an angle around an arbitrary axis with more complicated formula resulting (not shown here). The average correlation,  $\langle G(\alpha, \mathbf{q}) \rangle$ , can be approximated by the ratio of averages:

$$\langle G(\alpha, \mathbf{q}) \rangle = \frac{\langle \tilde{I}(\alpha, \mathbf{q}) \cdot \tilde{I}(0, \mathbf{q}) \rangle}{\sqrt{\langle \tilde{I}(\alpha, \mathbf{q})^2 \rangle} \cdot \sqrt{\langle \tilde{I}(0, \mathbf{q})^2 \rangle}}. \quad (5)$$

This is further evaluated with the help of relations:

$$\langle I(0, \mathbf{q}) \rangle = \langle I(\alpha, \mathbf{q}) \rangle = |f(q)|^2 r_e^2 \Phi_{in} \Omega_{pix} \sum_{i,j=1}^N J_0(q r_{ij} \sin\theta_{ij}), \quad (6)$$

and

$$\langle I(0, \mathbf{q}) I(\alpha, \mathbf{q}) \rangle = (|f(q)|^2 r_e^2 \Phi_{in} \Omega_{pix})^2 \times \sum_{i,j,t,s=1}^N J_0 \left( q \sqrt{r_{ij}^2 \sin^2\theta_{ij} + r_{ts}^2 \sin^2\theta_{ts} + 2r_{ij}r_{ts} \sin\theta_{ij} \sin\theta_{ts} \cos(\phi_{ij} - \phi_{ts} - \alpha)} \right). \quad (7)$$

Below we evaluate numerically the average correlation, Eq. (5), for the object consisting of 100 atoms randomly located (i) on a sphere (hollow sphere) of a fixed radius and (ii) within a ball (solid sphere) of a fixed radius. The results are shown in Fig. 2. As the average distance between these randomly distributed 100 atoms is large, reliable fits can be obtained only for small values of  $q$  ( $q < 0.5$ ).

Fig. 2 shows that the Gaussian-like dependence of the average correlation function can be parametrized as

$$\langle G(\alpha, q) \rangle = a(q) e^{-\alpha^2 / 2b(q)^2} + c(q), \quad (8)$$

with  $q$ -dependent fitting parameters  $a$ ,  $b$  and  $c$ . This exponential behaviour can be understood within the framework of our simple theoretical model, and is directly related to the underlying Wilson statistics [13]. If the number of atoms within an object is large, their positions can be treated as 'random', and the central limit theorem can be applied to a global distribution of random variables depending on the object positions [13].

Assuming the large number of atoms, and their random locations within the object, one can introduce a continuous probability

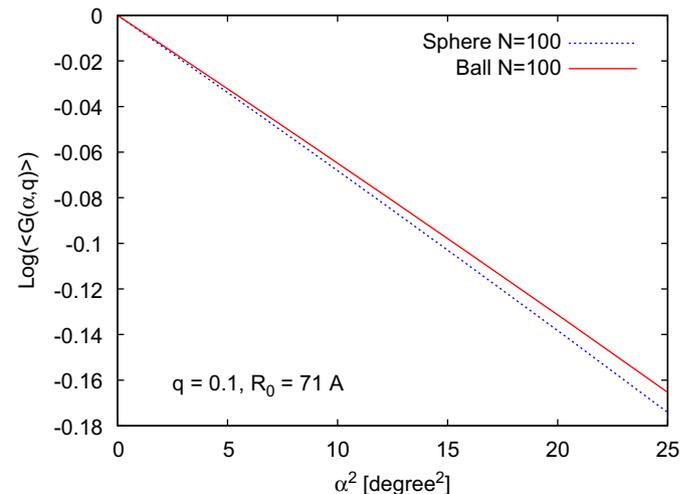
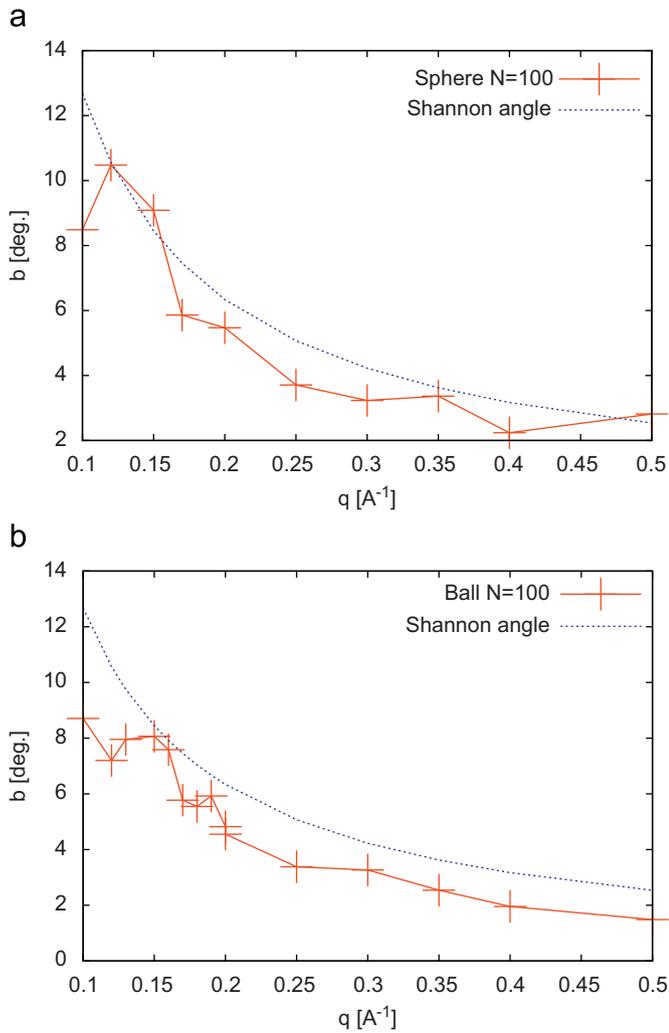


Fig. 2. Gaussian-like correlations between different diffraction patterns for object consisting of  $N=100$  atoms.



**Fig. 3.** The  $q$ -dependence of the fitting parameter  $b(q)$  for object consisting of  $N=100$  atoms. The fit is approaching the  $(4R_0q/2\pi)^{-1}$  (half of the Shannon angle) at the radius,  $R_0 \sim 71$  Å.

distribution for describing the atomic positions, and estimate the average value of pattern-to-pattern correlation to follow Eq. (8) (for details see Appendix). In this limit the estimated parameter  $b(q)$  then corresponds accurately to the half of the Shannon angle,  $\alpha_S$ ,  $\alpha_S = (2R_0q/2\pi)^{-1}$ , where  $R_0$  is the molecule radius. The Shannon angle denotes the  $q$ -dependent angular span of a Shannon speckle on the detector. The minimum rotation distance required to distinguish two speckles corresponds to half of the Shannon angle. The Shannon angle is therefore the natural unit of angular resolution (similarly to the definition of resolution in optical microscopy).

Fig. 3 shows the plots of the  $q$ -dependence of the fitting parameter  $b(q)$  for 100 atom-object. In this case statistical fluctuations are large, and the value of  $b(q)$  only approaches half of the Shannon angle.

In the next section we simulate diffraction images from a biological sample in order to show that also the diffraction images obtained with a realistic large molecule are correlated, following Eq. (8). The obtained angular resolution is determined with a very good accuracy by half of the Shannon angle.

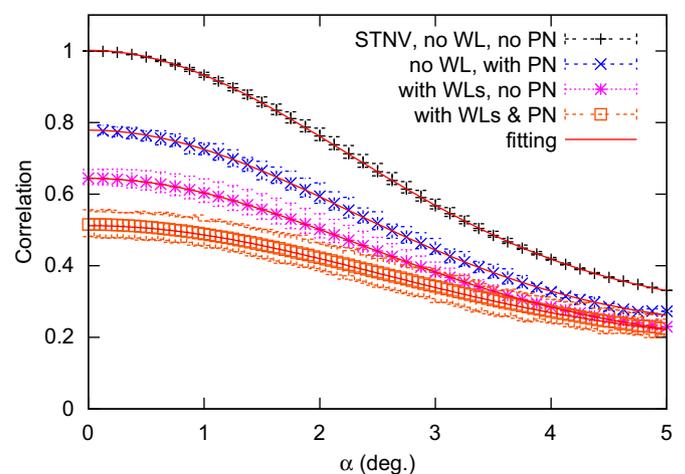
### 3. Correlations between diffraction images of a rotated virus

So far we have tried to understand the origin of correlations, considering the simplest case of a monoatomic sample. For this

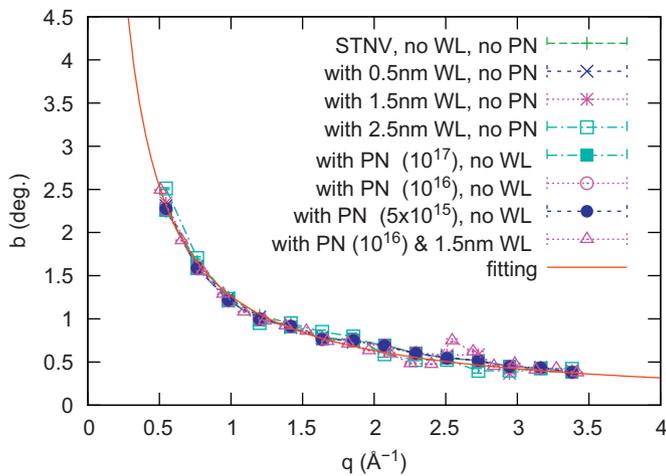
specific case we used Eqs. (6) and (7). For more complicated systems direct simulations of diffraction patterns are necessary. Below we study the correlations between diffraction images of a realistic biosample, as considered in [14].

The test object for the simulations was the Satellite Tobacco Necrosis virus (STNV), whose capsid structure has been solved by X-ray crystallography [15] (Protein Data Bank ID: 2BUC): object size  $\sim 17$  nm,  $\sim 0.18$  M atoms, icosahedral symmetry. Figs. 4 and 5 show that for a pure virus the average correlation between different patterns is Gaussian, and it follows Eq. (8). The obtained Gaussian width, i.e. the angular resolution, is determined with a very good accuracy by the half of the Shannon angle.

In a realistic case the information contained in the diffraction images is additionally perturbed by the Poissonian noise. Also, as discussed in [14], the bioparticles injected into an FEL beam by spraying techniques are covered by an evaporating water layer [16]. A thick layer of water around the imaged object is considered to be a method of slowing down the radiation damage, i.e. slowing the movement of ions due to repulsive Coulomb forces within the irradiated sample [17]. In order to test the possible loss of structural information due to the presence of a water layer (WL) surrounding the imaged object, this virus was covered with water layers of different thickness. The loss of orientational and structural information due to the presence of a water layer and its varying molecular structure was quantified theoretically, using molecular dynamics (MD) and coherent diffraction imaging simulations. The effect of Poissonian noise (PN) for scattered photons was also taken into account. In this scheme we kept the positions of the equilibrated virus atoms fixed, and only allow the surrounding water molecules to move during the simulations, as we considered the effects of “random” water layers on diffraction images. The related effects of slightly different protein conformations within the virus shell as well as effects due to the radiation damage processes were not considered in this work. We expect that they would further lower the correlations between the patterns. Finally, in our diffraction simulations, we simulated only the empty capsid without filling it with RNA genome. The RNA structure inside a virus capsid appears to be disordered as determined from single particle EM analyses. Similar to these



**Fig. 4.** Examples of rotation-angle dependent correlation curves at the fixed range of  $q = 2\pi \times [0.069, 0.104]$  Å $^{-1}$ : (1) correlation between noise-free patterns of pure STNV at different orientations; (2) STNV particle with water layers of 1.5 nm average thickness each from a different MD simulation; correlation between noise-free patterns for different orientations; (3) correlation as in (1), but patterns include Poisson noise at level of photon flux  $10^{15}$  photons/ $\mu\text{m}^2$  per pulse; (4) WL + PN contributions are included. Error bars show estimated variance obtained for different absolute orientations and due to Poisson noise. The fitting function presented in Eq. (8) fits well the averaged correlation curves.



**Fig. 5.** The  $q$ -dependence of the fitting parameter  $b(q)$  for pure virus structure, and with WL and with PN included. The fit is  $b(q) = (4R_0q/2\pi)^{-1}$  (half of the Shannon angle) at the average virus radius,  $R_0 \sim 71 \text{ \AA}$ .

methods, it is expected that its contribution will average out for the large number of exposures considered here.

In Fig. 4, the diffraction pattern correlations between the STNV particles with/without WLs or PN are shown as function of relative rotation angle at a fixed small  $q$  range. The radiation wavelength was  $1.5 \text{ \AA}$ . In a purely mathematical treatment of the calculated diffraction patterns, for an ideal virus capsid, the correlation is trivially one for the same orientation, and it decreases with the increasing angular difference in orientation (Fig. 4). In a more realistic model, after introducing Poisson noise and water layers of different thickness, the correlation of the diffraction images is significantly reduced. This increases the difficulty of identifying the orientation of the particles.

However, the averaged correlation curves can be parameterized by a Gaussian fit, Eq. (8), with  $q$ -dependent fitting parameters  $a$ ,  $b$  and  $c$ . We fitted the three parameters, using the correlation curves obtained from our simulations with fitting errors less than 2%. The parameters  $a(q)$  and  $c(q)$  that determine the magnitude of the correlation, strongly depend on the imaged structure (not shown). They show the highest values for pure virus structure, and decrease significantly as soon as WL or PN is included. The rotation-independent background  $c(q)$  always decreases strongly for high values of  $q$ . When “random” water layers are included, the parameters  $a(q)$  and  $c(q)$  are significantly suppressed at the range of the liquid water peak. In the water layer case with average  $2.5 \text{ nm}$  thickness, the parameter  $a(q)$  is decreasing to less than  $0.1$ , which is comparable to the correlation variance. After including WL+PN, the fitted  $a(q)$  and  $c(q)$  are smaller than in the cases with a WL or with PN only, implying a much reduced correlation.

Fig. 5 shows the parameter  $b(q)$  for different cases. As it can be seen, this parameter is determined only by the size of virus. The value of  $b(q)$  corresponds to half of the Shannon angle  $\alpha_s$ ,  $\alpha_s = (2R_0q/2\pi)^{-1}$ , where  $R_0$  is the average virus radius. The randomness introduced by WL or PN does not affect the angular resolution. This implies that the limiting angular resolution given by  $b(q)$  is determined only by the virus structure.

In summary, we studied pattern-to-pattern correlations between diffraction images obtained from objects differently oriented in space. Assuming large number of atoms in an object and their random positions, we predicted the Gaussian profile of the average correlation function, and showed that this profile is maintained in simulations of diffraction images obtained from a realistic biological sample (a small virus). Further, we demonstrated that the obtained

angular (orientational) resolution is determined only by the virus structure. It is unaffected by a random water layer surrounding the virus or a Poissonian noise. This indicates that the classification of diffraction images within the pattern-to-pattern scheme is not perturbed much by these effects, and therefore still possible.

## Acknowledgments

The authors thank Veit Elser and Abbas Ourmazd for illuminating comments on the classification schemes. B. Z. thanks R. Santra for discussions.

## Appendix

Assuming a large number of atoms, and their random locations within the object, one can introduce a continuous probability distribution for describing the atomic positions, and estimate the average value of pattern-to-pattern correlation to follow Eq. (8). In this limit the estimated parameter  $b(q)$  then corresponds accurately to the half of the Shannon angle,  $\alpha_s$ ,  $\alpha_s = (2R_0q/2\pi)^{-1}$ , where  $R_0$  is the molecule radius. In order to prove this, we recall again Eq. (7):

$$\langle I(0, \mathbf{q})I(\alpha, \mathbf{q}) \rangle = [f(q)]^2 r_e^2 \phi_{in} \Omega_{pix}^2 \times \sum_{i,j,s,t=1}^N J_0 \left( q \sqrt{r_{ij}^2 \sin^2 \theta_{ij} + r_{st}^2 \sin^2 \theta_{st} + 2r_{ij}r_{st} \sin \theta_{ij} \sin \theta_{st} \cos(\phi_{ij} - \phi_{st} - \alpha)} \right).$$

In what follows we restrict only to the dominant  $\alpha$  dependent terms in Eq. (7). We have proved (not shown) that the dominant  $\alpha$  dependent contribution to correlation function comes from the terms, where  $i=t$ ,  $j=s$ . If we neglect all other terms, Eq. (7) then becomes:

$$\langle I(0, \mathbf{q})I(\alpha, \mathbf{q}) \rangle \approx [f(q)]^2 r_e^2 \phi_{in} \Omega_{pix}^2 \sum_{(ij)} J_0 \left( 2qr_{ij} \sin \theta_{ij} \sin \frac{\alpha}{2} \right). \quad (9)$$

We assume that the atoms are randomly located within a ball so a pair of atoms gives a random value of  $\theta_{ij}$  between  $0$  and  $\pi$ , while it takes a random valued  $r_{ij}$  between  $0$  and  $2R_0$ . Every  $r_{ij}$  value is not equally likely, but follows a probability distribution  $P(r_{ij})$ . We can replace the sum over atom-pairs with an integral over  $r \equiv r_{ij}$  and  $\theta \equiv \theta_{ij}$ :

$$\langle I(0, \mathbf{q})I(\alpha, \mathbf{q}) \rangle \approx [f(q)]^2 r_e^2 \phi_{in} \Omega_{pix}^2 \int_0^\pi \int_0^{2R_0} J_0 \left( 2qr \sin \theta \sin \frac{\alpha}{2} \right) \times P(r) r^2 \sin \theta dr d\theta. \quad (10)$$

The probability distribution has the radial dependence of the autocorrelation function of a ball and is given by:

$$P(r) \sim \frac{\pi r^3}{24} + \frac{4\pi R_0^3}{3} - \pi r R_0^2. \quad (11)$$

If we make the small angle approximation, the expression (10) simplifies to

$$\langle I(0, \mathbf{q})I(\alpha, \mathbf{q}) \rangle \approx [f(q)]^2 r_e^2 \phi_{in} \Omega_{pix}^2 \int_0^\pi \int_0^{2R_0} J_0(qr \sin \theta \alpha) P(r) r^2 \sin \theta dr d\theta. \quad (12)$$

It can be shown by comparing terms in the Taylor series expansion to 4th order that,

$$\int_0^\pi \int_0^{2R_0} J_0(qr \sin \theta \alpha) P(r) r^2 \sin \theta dr d\theta \approx e^{-q^2 R_0^2 \alpha^2 / 5}. \quad (13)$$

Using the parameterisation from Eq. (8), we predict:

$$b(q) = \frac{\sqrt{2.5}}{qR_0} \approx \frac{1.58}{qR_0}. \quad (14)$$

Comparing this to half of Shannon angle:

$$b(q) = \frac{2\pi}{4qR_0} \approx \frac{1.57}{qR_0}, \quad (15)$$

we find that these two values are in a very good agreement.

## References

- [1] J.C.H. Spence, R.B. Doak, Phys. Rev. Lett. 92 (2004) 198102.
- [2] G. Huld, et al., J. Struct. Biol. 144 (2003) 219.
- [3] R. Neutze, et al., Nature 406 (2000) 752.
- [4] Z. Jurek, et al., Eur. Phys. J. D 29 (2004) 217.
- [5] Z. Jurek, et al., Europhys. Lett. 65 (2004) 491.
- [6] S.P. Hau-Riege, R.A. London, A. Szoke, Phys. Rev. E 69 (2004) 051906.
- [7] S.P. Hau-Riege, et al., Phys. Rev. E 71 (2005) 061919.
- [8] H. Chapman, et al., Nat. Phys. 2 (2006) 839.
- [9] M.J. Bogan, et al., Nanoletters 8 (2008) 310.
- [10] N.-T. Loh, V. Elser, Phys. Rev. E 80 (2009) 026705.
- [11] R. Fung, et al., Nat. Phys. 5 (2009) 64.
- [12] G. Bortel, et al., J. Struct. Biol. 166 (2009) 226.
- [13] U. Shmueli, Theories and Techniques of Crystal Structure Determination, Oxford University Press, 2007, p. 164.
- [14] F. Wang, et al., submitted to Phys. Rev. E.
- [15] T.A. Jones, L. Liljas, J. Mol. Biol. 177 (1984) 735.
- [16] A. Patriksson, et al., Biochemistry 46 (2007) 933.
- [17] S.P. Hau-Riege, et al., Phys. Rev. Lett. 98 (2007) 198302.