Solution of the phase problem for coherent scattering from a disordered system of identical particles

2013 New J. Phys. 15 013059

(http://iopscience.iop.org/1367-2630/15/1/013059)

New Journal of Physics
The open access journal for physics

Solution of the phase problem for coherent scattering from a disordered system of identical particles

R P Kurta[1], R Dronyak[1,4], M Altarelli[2], E Weckert[1] and I A Vartanyants[1,3,5]

[1] Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, D-22607 Hamburg, Germany
[2] European X-ray Free-Electron Laser Facility, Notkestraße 85, D-22607 Hamburg, Germany
[3] National Research Nuclear University, 'MEPhI', 115409 Moscow, Russia
E-mail: ivan.vartaniants@desy.de

New Journal of Physics **15** (2013) 013059 (14pp)
Received 12 October 2012
Published 28 January 2013
Online at http://www.njp.org/
doi:10.1088/1367-2630/15/1/013059

Abstract. While the implementation of single-particle coherent diffraction imaging for non-crystalline particles is complicated by current limitations on photon flux, hit rate and sample delivery, the concept of many-particle coherent diffraction imaging offers an alternative way of overcoming these difficulties. In this paper, we present a direct, non-iterative approach for the recovery of the diffraction pattern corresponding to a single particle using coherent x-ray data collected from a two-dimensional disordered system of identical particles; this approach does not require *a priori* information about the particles and can be applied to the general case of particles without symmetry. The reconstructed single-particle diffraction pattern can be directly used in common iterative phase retrieval algorithms to recover the structure of the particle.

[4] Deceased.
[5] Author to whom any correspondence should be addressed.

Contents

1. Introduction 2
2. Basic equations 3
3. Recovery of the projected electron density of a single particle 6
4. Results and discussion 10
5. Conclusions 12
Acknowledgments 12
Appendix 13
References 14

1. Introduction

It was recently realized [1–3] that analysis of diffraction patterns from disordered systems based on intensity cross-correlation functions (CCFs) can provide information on the local symmetry of these systems. A variety of disordered systems, for example colloids and molecules in solution, and liquids or atomic clusters in the gas phase, can be studied by this approach. It becomes especially attractive with the availability of x-ray free-electron lasers (FELs) [4–6]. Essentially, scattering from an ensemble of identical particles can provide, in principle, the same structural information as single-particle coherent imaging experiments [7] that are at present limited in achievable resolution [8]. Clearly, a large number of particles in the native environment can scatter up to a higher resolution, at the same photon fluences, comparing with the experiments on single particles injected into the FEL beam. However, it is still a challenge to recover the structure of individual particles composing a system using the CCF formalism. In the pioneering work of Kam [9, 10] it was proposed to determine the structure of a single particle using scattered intensity from many identical particles in solution. However, this approach, based on spherical harmonics expansion of the scattered amplitudes, was not fully explored until now. Recently, it was revised theoretically [11–13] and experimentally [14]. The possibility to recover the structure of individual particles was demonstrated in systems in two dimensions (2D) [11, 12, 14] and in three dimensions (3D) [13] using additional *a priori* knowledge of the symmetry of the particles. Unfortunately, these approaches, based on optimization routines [11, 13] and iterative techniques [12, 14], do not guarantee the uniqueness of the recovered structure unless strong constraints are applied. This all leads to a high demand for finding direct, non-iterative approaches for recovering the structure of an individual particle in the system.

In this paper, we further develop Kam's ideas and propose an approach enabling direct reconstruction of single-particle intensity distributions using algebraic formalism of measured two- and three-point CCFs without additional constraints. Once the single-particle intensity is obtained, conventional phase retrieval algorithms [15, 16] can recover the projected electron density of a single particle. Our approach is developed for 2D systems of particles, which is of particular interest for studies of membrane proteins. It can be also used to study 3D systems provided that particles can be aligned with respect to the incoming x-ray beam direction.

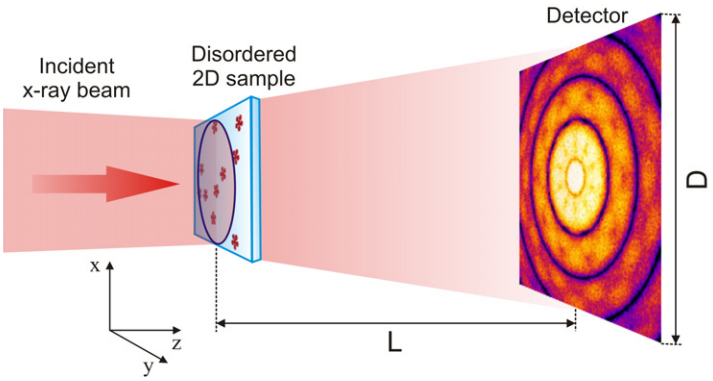

Figure 1. Geometry of the diffraction experiment. The incident x-ray beam coherently illuminates a 2D disordered sample and produces a diffraction pattern on a detector. The direction of the incident beam is defined along the z-axis of the coordinate system.

2. Basic equations

We consider a scattering experiment in transmission geometry (see figure 1), where the direction of the incoming x-ray beam is perpendicular to the 2D sample plane, and simulate a set of diffraction patterns. Our sample consists of an arbitrary small number N of identical, spatially disordered particles (see figure 2(a)). Also, we assume a uniform distribution of orientations of particles in the system. First, for simplicity, we consider a model where the total scattered intensity $I(\boldsymbol{q})$ is represented as an *incoherent* sum of intensities $I_{\psi_i}(\boldsymbol{q})$ corresponding to individual particles in the system

$$I(\boldsymbol{q}) = \sum_{i=1}^{N} I_{\psi_i}(\boldsymbol{q}), \tag{1}$$

where $\boldsymbol{q}$ is the momentum transfer vector and ψ_i is the orientation of the ith particle. This model is a good approximation for dilute systems when the mean distance between particles is much larger than their size [2, 3]. In this approximation we neglect the inter-particle correlations due to coherent interference of scattered amplitudes from individual particles. Later, in simulations, we generalize our approach to the case of *coherent* scattering from a system of particles in the presence of Poisson noise and demonstrate the applicability of our approach.

In the frame of kinematical scattering, the intensity $I_{\psi_0}(\boldsymbol{q})$ scattered from a single particle in some reference orientation ψ_0 is related to the electron density of the particle $\rho_{\psi_0}(\boldsymbol{r})$ by the following relation:

$$I_{\psi_0}(\boldsymbol{q}) = \left| \int \rho_{\psi_0}(\boldsymbol{r}) \exp(\mathrm{i}\boldsymbol{q}\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \right|^2. \tag{2}$$

Once the single-particle intensity $I_{\psi_0}(\boldsymbol{q})$ is determined, conventional phase retrieval algorithms [15, 16] can recover the projected electron density $\rho_{\psi_0}(\boldsymbol{r})$ of a single particle. Our goal is to determine the scattering pattern of a single particle $I_{\psi_0}(\boldsymbol{q})$ using a large number of diffraction patterns $I(\boldsymbol{q})$ corresponding to different realizations of the system.

The intensity $I_{\psi_0}(\boldsymbol{q})$ scattered from a single particle can also be described in a polar coordinate system $\boldsymbol{q} = (q, \varphi)$ as a function of the angular position $(0 < \varphi \leqslant 2\pi)$ on the

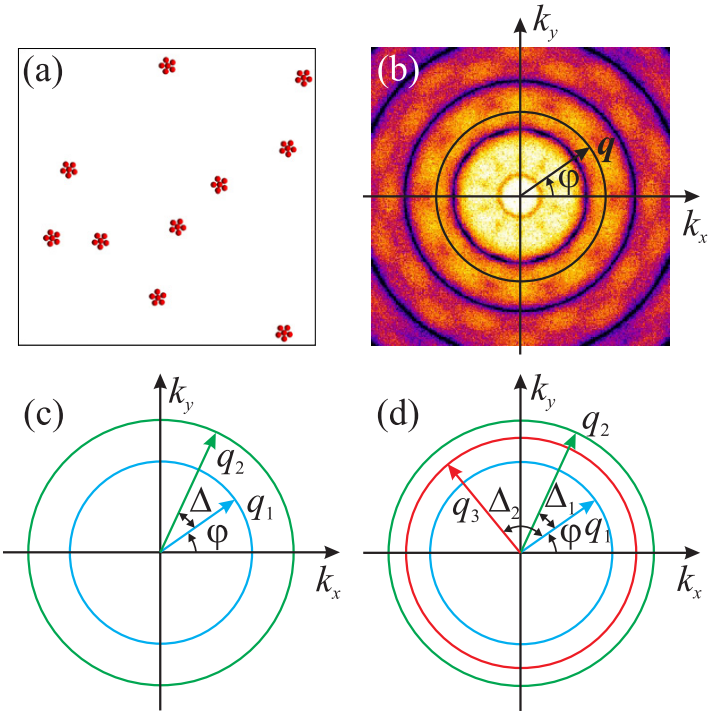

Figure 2. (a) A 2D disordered system composed of $N = 10$ clusters in random positions and orientations. (b) Scattered intensity $I(\mathbf{q})$ in a polar coordinate system $\mathbf{q} = (q, \phi)$ as a function of the angular position ϕ on the resolution ring q. (c) and (d) Definition of the momentum transfer vectors in the derivation of the two-point $C(q_1, q_2, \Delta)$ (c) and the three-point CCFs $C(q_1, q_2, q_3, \Delta_1, \Delta_2)$ (d).

resolution ring q (see figure 2(b)). For each resolution ring q, one can represent the intensity function $I_{\psi_0}(q, \varphi)$ as a Fourier series expansion

$$I_{\psi_0}(q, \varphi) = \sum_{n=-\infty}^{\infty} I^n_{q,\psi_0} \exp(in\varphi), \tag{3}$$

where I^n_{q,ψ_0} are the Fourier components of $I_{\psi_0}(q, \varphi)$.

For a 2D system of identical particles, the intensity $I_{\psi_0}(q, \varphi)$ scattered from a particle in the reference orientation $\psi_0 = 0$ is related to the intensity $I_{\psi_i}(q, \varphi)$ scattered from a particle in an arbitrary orientation ψ_i as $I_{\psi_i}(q, \varphi) = I_{\psi_0}(q, \varphi - \psi_i)$. Applying the shift theorem for the Fourier transforms [17], we obtain for the corresponding Fourier components of the intensities, $I^n_{q,\psi_i} = I^n_{q,\psi_0} \exp(-in\psi_i)$. Using these relations, we can write for the Fourier components I^n_q of the intensity $I(q, \varphi)$ scattered from N particles

$$I^n_q = I^n_{q,\psi_0} \sum_{i=1}^{N} \exp(-in\psi_i) = I^n_{q,\psi_0} \mathbf{A}_n, \tag{4}$$

where $\mathbf{A}_n = \sum_{i=1}^{N} \exp(-in\psi_i)$ is a random phasor sum [18]. According to (3) the intensity scattered from a single particle can be uniquely determined by the set of complex coefficients $\{I^n_{q,\psi_0}\} = \{|I^n_{q,\psi_0}|, \phi^n_{q,\psi_0} = \arg(I^n_{q,\psi_0})\}$. Here we propose a direct approach for the determination

of these Fourier components $\{I^n_{q,\psi_0}\}$ applying two- and three-point CCFs to the measured intensities $I(q,\varphi)$ scattered from N particles.

We start with the two-point CCF defined at two resolution rings q_1 and q_2 [2, 9, 11]

$$C(q_1, q_2, \Delta) = \left\langle \widetilde{I}(q_1,\varphi)\widetilde{I}(q_2,\varphi+\Delta)\right\rangle_\varphi , \tag{5}$$

where $0 \leqslant \Delta \leqslant 2\pi$ is the angular coordinate (see figure 2(c)), $\widetilde{I}(q,\varphi) = I(q,\varphi) - \langle I(q,\varphi)\rangle_\varphi$ is the intensity fluctuation function and $\langle \ldots \rangle_\varphi$ denotes the average over the angle φ. It can be directly shown [2] that the Fourier components $C^n_{q_1,q_2}$ of the CCF $C(q_1, q_2, \Delta)$ for $n \neq 0$ are defined by the Fourier components I^n_q of the intensities $I(q,\varphi)$,[6]

$$C^n_{q_1,q_2} = I^{n*}_{q_1} \cdot I^n_{q_2} \tag{6}$$

and for $n = 0$ Fourier components $C^n_{q_1,q_2} = 0$ according to the definition of the intensity fluctuation function $\widetilde{I}(q,\varphi)$. Using (4) in (6) and introducing statistical averaging $\langle \ldots \rangle_M$ over a large number M of diffraction patterns, one can obtain

$$\left\langle C^n_{q_1,q_2}\right\rangle_M = I^{n*}_{q_1,\psi_0} I^n_{q_2,\psi_0} \cdot \left\langle |\mathbf{A}_n|^2\right\rangle_M = I^{n*}_{q_1,\psi_0} I^n_{q_2,\psi_0} \cdot N. \tag{7}$$

Here, we used the fact that for a uniform distribution of orientations of N particles, $\langle |\mathbf{A}_n|^2\rangle_M$ asymptotically converges to N for a sufficiently large number M of diffraction patterns[7].

Equation (7) can be used to determine both, the amplitudes $|I^n_{q,\psi_0}|$ and phases ϕ^n_{q,ψ_0} (for $n > 0$) of the Fourier components I^n_{q,ψ_0} associated with a single particle. For example, to determine the amplitudes equation (7) can be applied successively to three different resolution rings q_1, q_2 and q_3 connecting each time a pair of q-values. Direct evaluation gives for the amplitudes of the Fourier components of a single particle on the ring q_1

$$|I^n_{q_1,\psi_0}| = \mathcal{I}^n_{q_1,\psi_0}/\sqrt{N}, \tag{8}$$

where $\mathcal{I}^n_{q_1,\psi_0} = \sqrt{|\langle C^n_{q_1,q_2}\rangle_M| \cdot |\langle C^n_{q_3,q_1}\rangle_M| / |\langle C^n_{q_3,q_2}\rangle_M|}$ is an experimentally determined quantity. Applying (8) to different resolution rings q and orders n, all required amplitudes $|I^n_{q,\psi_0}|$ can be determined[8]. Equation (8) should be used with care to avoid possible instabilities due to division by zero. For this purpose one should exclude from consideration the cases when $|\langle C^n_{q_3,q_2}\rangle_M|$ is close to zero. Since the Fourier components of the intensities obey the symmetry condition $I^{n*}_{q,\psi_0} = I^{-n}_{q,\psi_0}$, it is sufficient to determine I^n_{q,ψ_0} for $n \geqslant 0$. The zeroth-order Fourier component by its definition [2, 3] is a real-valued quantity, $I^0_q = \langle I(q,\varphi)\rangle_\varphi$, and can be determined from the experiment as well. Using (4) one can readily find that $I^0_{q,\psi_0} = I^0_q/N$.

Equation (7) also determines the phase difference between two Fourier components $I^n_{q_1,\psi_0}$ and $I^n_{q_2,\psi_0}$ of the same order n, defined at two different resolution rings q_1 and q_2,

$$\arg\left[\left\langle C^n_{q_1,q_2}\right\rangle_M\right] = \phi^n_{q_2,\psi_0} - \phi^n_{q_1,\psi_0}. \tag{9}$$

Note that one can freely assign an arbitrary phase to one of the Fourier components $I^n_{q_j,\psi_0}$, which corresponds to an arbitrary initial angular orientation of a particle. Assuming, for example,

[6] We note that the Fourier components of the intensity fluctuation function $\widetilde{I}^n_q$ can be expressed through the Fourier components of intensity I^n_q using the relation $\widetilde{I}^n_q = I^n_q - I^0_q \cdot \delta_{n,0}$, where $\delta_{n,0}$ is the Kronecker symbol.

[7] Note that in this paper, contrary to [2, 3], we use non-normalized CCFs, which, in particular, results in different asymptotic values of $\langle |\mathbf{A}_n|^2\rangle_M$.

[8] If the number of particles in the system is not known *a priori*, then N in (8) has to be considered as a scaling factor that is obtained on the final stage of reconstruction of a single-particle intensity (see section 3).

$\phi^n_{q_1,\psi_0} = 0$ and using (9) one can directly determine the phases of the Fourier components with the same n-value on all other resolution rings $q_2 \neq q_1$ [12].

To completely solve the phase problem, it is required to obtain additional phase relations between Fourier components with different n values on different resolution rings q. These relations can be determined using a three-point CCF introduced by Kam [9, 10]. An important aspect of our approach is to use the three-point CCF defined on *three* different resolution rings (see figure 2(d)),

$$C(q_1, q_2, q_3, \Delta_1, \Delta_2) = \langle \widetilde{I}(q_1, \varphi)\widetilde{I}(q_2, \varphi + \Delta_1)\widetilde{I}(q_3, \varphi + \Delta_2)\rangle_\varphi, \tag{10}$$

contrary to [11, 12] where the three-point CCF was defined on *two* resolution rings. Similar to $C^n_{q_1,q_2}$ it can be shown (see the appendix) that for the Fourier components of this CCF, the following relation is valid,

$$C^{n_1,n_2}_{q_1,q_2,q_3} = I^{(n_1+n_2)*}_{q_1} I^{n_1}_{q_2} I^{n_2}_{q_3} \tag{11}$$

for $n_1 \neq 0, n_2 \neq 0, n_1 \neq -n_2$. Using (4) in (11) and performing statistical averaging, we obtain for the averaged Fourier components of the three-point CCF

$$\left\langle C^{n_1,n_2}_{q_1,q_2,q_3}\right\rangle_M = I^{(n_1+n_2)*}_{q_1,\psi_0} I^{n_1}_{q_2,\psi_0} I^{n_2}_{q_3,\psi_0} \cdot \langle \mathbf{A}_{n_1,n_2}\rangle_M, \tag{12}$$

where $\mathbf{A}_{n_1,n_2} = \sum_{i,j,k=1}^{N} \exp\{\mathrm{i}[(n_1+n_2)\psi_i - n_1\psi_j - n_2\psi_k]\}$. Our analysis shows (see the appendix) that the statistical average $\langle \mathbf{A}_{n_1,n_2}\rangle_M$ converges to N for a sufficiently large number M of diffraction patterns, i.e. $\arg[\langle \mathbf{A}_{n_1,n_2}\rangle_M] = 0$, and we obtain from (12) the following phase relation:

$$\arg\left[\left\langle C^{n_1,n_2}_{q_1,q_2,q_3}\right\rangle_M\right] = \phi^{n_1}_{q_2,\psi_0} + \phi^{n_2}_{q_3,\psi_0} - \phi^{(n_1+n_2)}_{q_1,\psi_0}. \tag{13}$$

Equation (13) determines the phase shift between three Fourier components $I^{(n_1+n_2)}_{q_1,\psi_0}$, $I^{n_1}_{q_2,\psi_0}$ and $I^{n_2}_{q_3,\psi_0}$ of different order n defined on three resolution rings. If $n_1 = n_2 = n$, equation (13) reduces to a particular form, giving the phase relation between Fourier components of only two different orders n and $2n$,

$$\arg\left[\left\langle C^{n,n}_{q_1,q_2,q_3}\right\rangle_M\right] = \phi^{n}_{q_2,\psi_0} + \phi^{n}_{q_3,\psi_0} - \phi^{2n}_{q_1,\psi_0}. \tag{14}$$

Equations (8), (9) and (13) constitute the core of our approach and allow us to directly determine the complex Fourier components I^n_{q,ψ_0} using measured x-ray data from a disordered system of N particles. The obtained Fourier components I^n_{q,ψ_0} can be used in (3) to recover the scattered intensity $I_{\psi_0}(q, \varphi)$ corresponding to a single particle.

3. Recovery of the projected electron density of a single particle

We demonstrate our approach for the case of a coherent illumination of a disordered system of particles, in the presence of Poisson noise in the scattered signal. We recover diffraction patterns and projected electron densities for two different particles, a centered pentagonal cluster (figure 3(a)) that has fivefold rotational symmetry and an asymmetric cluster (figure 3(b)). Both clusters have a size of $d = 300\,\mathrm{nm}$ and are composed of polymethylmethacrylate (PMMA) spheres of 50 nm radius.

In both the cases, we consider kinematical coherent scattering of x-rays with wavelength $\lambda = 1\,\text{Å}$ from a system of $N = 10$ clusters in random positions and orientations, distributed within a sample area of $5 \times 5\,\mu\mathrm{m}^2$ (see figure 2(a)). Diffraction patterns are simulated for a

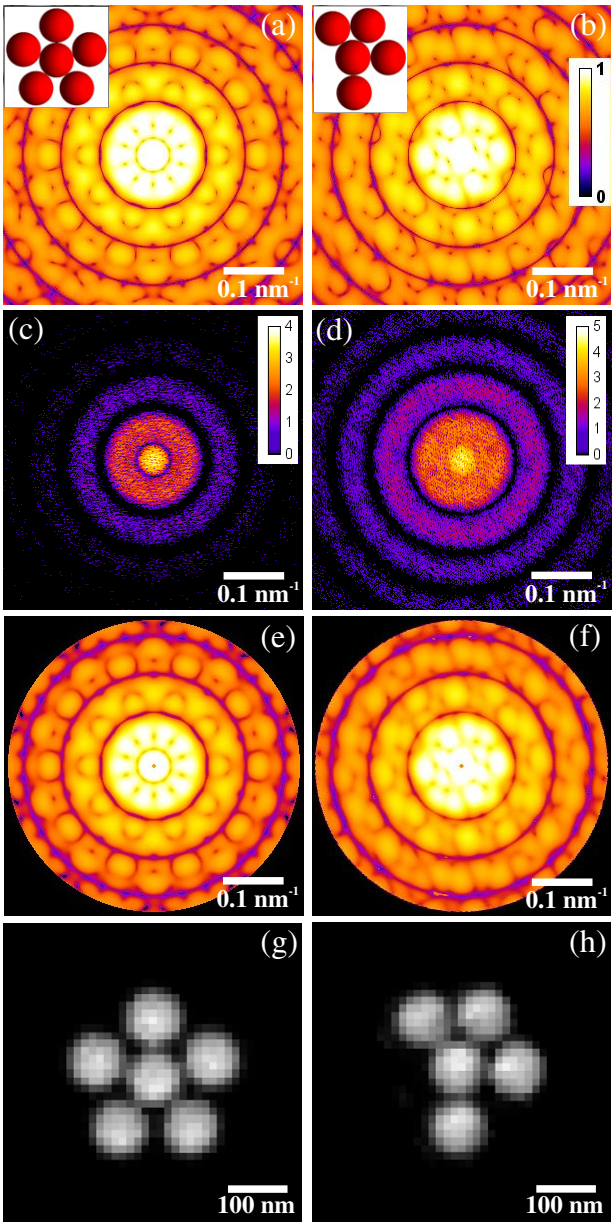

Figure 3. (a), (b) Scattered intensity (logarithmic scale) calculated for a single pentagonal cluster (a) and an asymmetric cluster (b) (clusters are shown in the insets). (c), (d) Coherently scattered intensity from a disordered system consisting of $N = 10$ clusters in random position and orientation. Scattered signal corresponds to the incident fluence of 10^{12} and 10^{13} photons per 25 μm^2 for pentagonal and asymmetric clusters, respectively, and contains Poisson noise. (e), (f) Scattered intensity corresponding to a single pentagonal (e) and asymmetric (f) clusters recovered from $M = 10^5$ diffraction patterns of the form (c) and (d), respectively. (g), (h) Structure of a single cluster reconstructed by an iterative phase retrieval algorithm using the diffraction patterns shown in (e) and (f). The intensity in (a), (b), (e) and (f) is given in arbitrary units (see the scale bar in (b)) and in (c) and (d) in photon counts.

2D detector of size $D = 24\,\text{mm}$ (with pixel size $p = 80\,\mu\text{m}$), positioned in the transmission geometry at $L = 3\,\text{m}$ distance from the sample (see figure 1). This experimental geometry corresponds to scattering to a maximum resolution of $0.25\,\text{nm}^{-1}$. For given experimental conditions the speckle size corresponding to the illuminated area is below the pixel size of the detector. At the same time the speckle size corresponding to the size of a single particle is about 12 pixels, which provides sufficient sampling for phase retrieval.

The coherently scattered intensities simulated for single realizations of the systems[9] are shown in figures 3(c) and (d). We note that in simulations the intensity $I(q, \varphi)$ scattered from a system of N particles was calculated as a coherent sum of the scattered amplitudes $A_{\psi_i}(q, \varphi)$ from each particle, i.e. $I(q, \varphi) = |\sum_{i=1}^{N} A_{\psi_i}(q, \varphi)|^2$. The incident fluence was considered to be 10^{12} and 10^{13} photons per $25\,\mu\text{m}^2$ for pentagonal and asymmetric clusters, respectively. Poisson noise was included in both cases. Since the asymmetric cluster has one PMMA sphere less than the pentagonal cluster, higher photon fluence was considered in the latter case to have a sufficient number of scattered photons at high resolution. The Fourier components of two-point and three-point CCFs (equations (7) and (12)) were averaged over $M = 10^5$ diffraction patterns[10].

Once all the amplitudes $\left|I^n_{q,\psi_0}\right|$ and phases ϕ^n_{q,ψ_0} of the Fourier components I^n_{q,ψ_0} are determined, the diffraction pattern corresponding to a single particle can be recovered by performing the Fourier transform (3). Using experimentally accessible quantities $\left\langle\langle I(q, \varphi)\rangle_\varphi\right\rangle_M = N I^0_{q,\psi_0}$ and $\mathcal{I}^n_{q,\psi_0} = \sqrt{N}\left|I^n_{q,\psi_0}\right|$, we can rewrite (3) as

$$I_{\psi_0}(q, \varphi) = \frac{\left\langle\langle I(q, \varphi)\rangle_\varphi\right\rangle_M}{N} + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{\mathcal{I}^n_{q,\psi_0}}{\sqrt{N}} \exp(\mathrm{i}\phi^n_{q,\psi_0}) \exp(\mathrm{i}n\varphi). \tag{15}$$

Note that even in the case of coherent scattering the intensity $\left\langle\langle I(q, \varphi)\rangle_\varphi\right\rangle_M$ averaged over a large number of diffraction patterns asymptotically converges to $N I^0_{q,\psi_0}$, which is the same value as for the case of incoherent scattering.

Here we present a detailed description of how the amplitudes and phases of the Fourier components I^n_{q,ψ_0} associated with a single particle, using x-ray data from a 2D system of pentagonal clusters (figure 3(a)) described above. Applying expression (8) to different q we find all required amplitudes $\sqrt{N}|I^n_{q,\psi_0}|$ (for $n > 0$) scaled by the factor $\sqrt{N}$. The zeroth-order Fourier component is determined as $I^0_{q,\psi_0} = \left\langle\langle I(q, \varphi)\rangle_\varphi\right\rangle_M / N$. In figure 4 the values $\sqrt{N}|I^n_{q,\psi_0}|$ derived from (8) normalized by $\langle\langle I(q, \varphi)\rangle_\varphi\rangle_M$ are shown for $n \leqslant 40$ at three different resolution rings $q_1 = 0.21\,\text{nm}^{-1}$, $q_2 = 0.23\,\text{nm}^{-1}$ and $q_3 = 0.25\,\text{nm}^{-1}$. Due to fivefold symmetry of the pentagonal cluster, in the q-range accessible in our model experiment we need to determine the phases of the Fourier components I^n_{q,ψ_0} only with $n = 10l$, $1 \leqslant l \leqslant 4$, where l is an integer. Below, we present the algorithm for the phase determination using equations (9), (13) and (14).

1. First, we assign an arbitrary phase to one of the Fourier components I^n_{q,ψ_0}, which corresponds to an arbitrary initial angular orientation of a particle. It is convenient to assign

[9] All simulations of diffraction patterns were performed using the computer code MOLTRANS.

[10] According to our simulations of dilute systems, the inter-particle contribution in the two-point and three-point CCFs defined on different resolution rings q_1, q_2 and q_3 (see (5) and (10)) even in the case of coherent illumination of a disordered system is negligible. Contrary to that, as was shown in [2, 3], this is not the case for two-point CCFs defined on the same resolution ring $q_1 = q_2$, when inter-particle contributions and noise can give a substantial contribution. Therefore, in this work we avoid the last case.

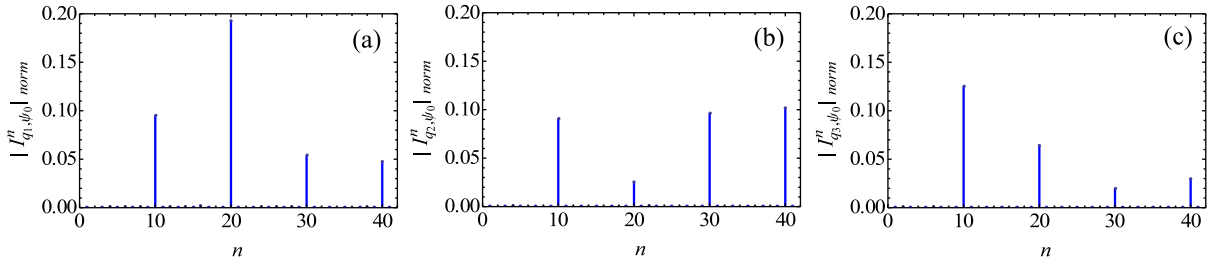

Figure 4. Normalized amplitudes $|I^n_{q,\psi_0}|_{\rm norm} = |I^n_{q,\psi_0}|/(|I^0_{q,\psi_0}|\sqrt{N})$ determined using (8) at three different resolution rings: (a) $q_1 = 0.21\,{\rm nm}^{-1}$, (b) $q_2 = 0.23\,{\rm nm}^{-1}$ and (c) $q_3 = 0.25\,{\rm nm}^{-1}$.

this phase to the Fourier component with the smallest available n value. In our case, we assign a zero phase $\phi^{10}_{q_1,\psi_0} = 0$ to the Fourier component $I^{10}_{q_1,\psi_0}$.

2. The phases of the Fourier components with $n = 10$ on other resolution rings q_2 ($j = 1, 2$) were obtained by using (9), i.e. $\phi^{10}_{q_2,\psi_0} = \phi^{10}_{q_1,\psi_0} + \arg[\langle C^{10}_{q_1,q_2}\rangle_M] = \arg[\langle C^{10}_{q_1,q_2}\rangle_M]$.
3. The phase of the Fourier component $I^{20}_{q_1,\psi_0}$ on the resolution ring q_1 was determined by using (14) with $n = 10$, $\phi^{20}_{q_1,\psi_0} = \phi^{10}_{q_2,\psi_0} + \phi^{10}_{q_3,\psi_0} - \arg[\langle C^{10,10}_{q_1,q_2,q_3}\rangle_M]$.
4. The phases of the Fourier components with $n = 20$ on other resolution rings q_2 ($j = 1, 2$) were obtained from (9) $\phi^{20}_{q_2,\psi_0} = \phi^{20}_{q_1,\psi_0} + \arg[\langle C^{20}_{q_1,q_2}\rangle_M]$.
5. The phase of the Fourier component $I^{30}_{q_1,\psi_0}$ on the resolution ring q_1 was determined by using (13) with $n_1 = 10$ and $n_2 = 20$, $\phi^{30}_{q_1,\psi_0} = \phi^{10}_{q_2,\psi_0} + \phi^{20}_{q_3,\psi_0} - \arg[\langle C^{10,20}_{q_1,q_2,q_3}\rangle_M]$.

The process was continued in a similar way until all phases at each q-value were determined. The same procedure has been applied for the case of a 2D system of asymmetric clusters (figure 3(b)). Due to asymmetric structure of particles, in the same q-range we need to determine the phases of a significantly larger set of the Fourier components with $n = 2 \cdot k$, $1 \leqslant k \leqslant 24$, where k is an integer. Assigning a zero phase $\phi^2_{q_1,\psi_0} = 0$ to the Fourier component with $n = 2$, we successively determine the phases of the Fourier components of the higher orders up to $n = 48$. We note that the data redundancy intrinsic to $\langle C^n_{q_1,q_2}\rangle_M$ and $\langle C^{n_1,n_2}_{q_1,q_2,q_3}\rangle_M$ offers a lot of flexibility in determining possible ways of solving the phases of the Fourier components I^n_{q,ψ_0}. We used different choices for the sets of the Fourier components of two- and three-point CCFs in our phasing algorithm and obtained, in fact, the same reconstructed diffraction intensities of a single cluster. However, we note that one should avoid using in the phase determination algorithm those subsets of the Fourier components (both two- and three-point CCFs) which are defined on the resolution rings where the average intensity $\left\langle\langle I(q,\varphi)\rangle_\varphi\right\rangle_M$ has vanishing values. The phases of such Fourier components are poorly defined and may cause significant errors in the reconstruction.

If the number of particles N in the system is not known (see equation (15)), it can be determined using the positivity constraint $I_{\psi_0}(q,\varphi) \geqslant 0$ for the recovered intensity. One can slightly relax this constraint and allow a small fraction of pixels with negative values in order to get an image with a higher contrast, where the contrast can be defined as $\sum_{q,\varphi} |I_{\psi_0}(q,\varphi) - \langle I_{\psi_0}(q,\varphi)\rangle_\varphi|/\langle I_{\psi_0}(q,\varphi)\rangle_\varphi$. These negative values can be substituted by zeros before using this diffraction pattern in the iterative phase retrieval algorithms for the reconstruction of the

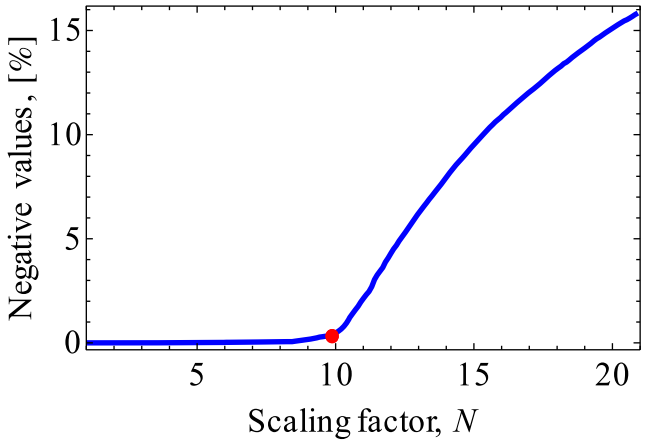

Figure 5. Dependence of the relative number of pixels with negative values (normalized to the total number of pixels on the detector) in the recovered diffraction pattern $I_{\psi_0}(q, \varphi)$ as a function of the scaling factor N. The red point indicates the value of $N = 9.8$ used in the final reconstruction of intensity $I_{\psi_0}(q, \varphi)$.

particle structure. This procedure can be justified by inaccuracies which arise due to statistical estimate of the CCF and also different sources of error (e.g. interpolation) in calculations. For example, for the case of pentagonal clusters in our simulations we applied the value of $N = 9.8$, which corresponds to 0.33% of negative pixels on the detector (see figure 5). This value is very close to the theoretically predicted $N = 10$ for a uniform distribution of orientations of particles.

4. Results and discussion

The diffraction patterns corresponding to single particles recovered by our approach are presented in figures 3(e) and (f). As one can see from these figures, the recovered single-particle diffraction patterns reproduce the diffraction patterns of individual clusters shown in figures 3(a) and (b) very well. The structure of the clusters reconstructed from these recovered diffraction patterns was obtained by a standard phase retrieval approach and is presented in figures 3(g) and (h). In our reconstructions we used an alternative sequence of hybrid input–output and error-reduction algorithms [15]. The comparison of the single-cluster structures obtained by our approach (figures 3(g) and (h)) with the initial model shown in the insets of figures 3(a) and (b) confirms the correctness of our reconstruction. Both clusters were reconstructed with a resolution up to 25 nm. These results clearly demonstrate the ability of our approach to recover the single-particle structure from noisy data obtained in coherent x-ray scattering experiments.

The ability of the presented approach to recover the diffraction pattern of a single particle relies on the accuracy of the two- and three-point CCFs determination. The statistical properties of the CCFs and their convergence to the average values strongly depend on the number of particles N in the system, their density and the distribution of their orientations [2, 3]. In particular, the value of the scaling factor N in the expression of the Fourier components of

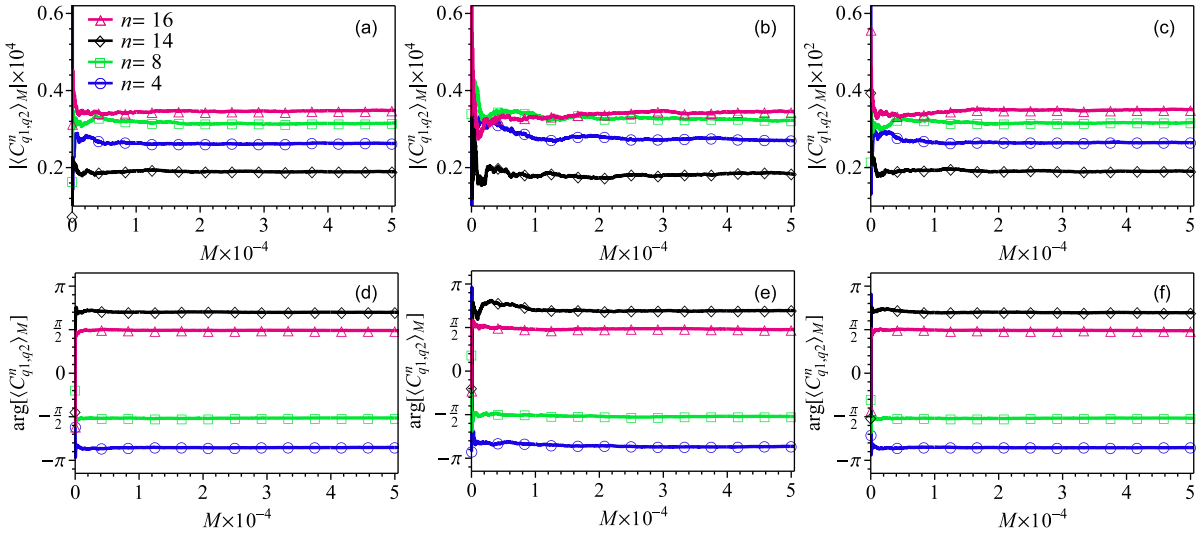

Figure 6. Statistical convergence of the Fourier components $\langle C^n_{q_1,q_2}\rangle_M$ at $q_1 = 0.25\,\mathrm{nm}^{-1}$, $q_2 = 0.09\,\mathrm{nm}^{-1}$, $n = 4, 8, 14$ and 16 calculated for the case of asymmetric clusters as a function of the number of diffraction patterns M used in the averaging. The amplitudes $|\langle C^n_{q_1,q_2}\rangle_M|$ and phases $\arg[\langle C^n_{q_1,q_2}\rangle_M]$ were calculated for the incident photon fluence 10^{12} photons per $25\,\mu\mathrm{m}^2$: (a), (d) without noise, (b), (e) with Poisson noise and (c), (f) for the photon fluence 10^{13} photons per $25\,\mu\mathrm{m}^2$ with Poisson noise.

the CCF is defined by the statistical distribution of particles in the system. We performed a few simulations with varying numbers N of particles in the system: (a) for Gaussian distribution of the number of particles with an average value $\langle N\rangle = 20$ and standard deviation $\sigma = 2$; and (b) for Poisson distribution[11] of the number of particles with an average value $\langle N\rangle = 30$. In both cases, the CCFs statistically converge to the same average values as for the systems with a fixed number of particles, $N = 20$ and 30 (for Gaussian and Poisson distributions, correspondingly). The only difference we observed is a slightly slower convergence of CCFs in the case of statistical distribution of particles compared to the case with a fixed number N. This means that in the case of systems with varying numbers of particles one needs to measure a larger number of diffraction patterns.

Particle density also affects the statistical behavior of the CCFs. In the limiting case of a very dilute system, the CCFs calculated for a single diffraction pattern are defined by independent structural contributions of N individual particles. In this case statistical averaging over diffraction patterns acts, in fact, on the fluctuating terms $|\mathbf{A}_n|^2$ in (7) and $\mathbf{A}_{n_1,n_2}$ in (12) and convergence of these terms to their statistical estimates determines the number M of diffraction patterns required for averaging [3]. In the case of a dense system, the CCFs contain a significant inter-particle contribution in the same range of q-values, where the structural contribution of individual particles is observed [2, 3]. This additional contribution slows down the convergence of CCFs (especially of the three-point CCF) and the number of measured

[11] Do not mix up this Poisson distribution of the number of particles with Poisson distribution used for noise simulation in the scattered intensities.

diffraction patterns should be increased. In the presence of Poisson noise in the scattered signal, one needs to accumulate even more diffraction patterns. Additional effects of resolution, solution scattering, etc on the measured CCFs were considered in detail in [19]. In the disordered systems considered in this work, the particle density is characterized by $\langle R\rangle/d \approx 5.0$, where $\langle R\rangle$ is the average inter-particle distance. In these conditions it was sufficient to use 10^5 diffraction patterns with Poisson noise to achieve convergence of the two- and three-point CCFs and perform a successful reconstruction of the particle structure.

In practical applications, the convergence of the CCFs can be directly controlled as a function of the number of the diffraction patterns M considered in the averaging [3]. The results presented in figure 6 demonstrate the evolution of the Fourier components $\left\langle C^n_{q_1,q_2}\right\rangle_M$ at $q_1 = 0.25\,\mathrm{nm}^{-1}$ and $q_2 = 0.09\,\mathrm{nm}^{-1}$ as a function of M, calculated for the case of coherent scattering from $N = 10$ asymmetric clusters. The amplitudes and phases of $\left\langle C^n_{q_1,q_2}\right\rangle_M$ for $n =$ 4, 8, 14 and 16 were calculated for the incident photon fluence 10^{12} photons per 25 μm^2 without noise (figures 6(a) and (d)) and with Poisson noise (figures 6(b) and (e)), and for the photon fluence 10^{13} photons per 25 μm^2 with Poisson noise (figures 6(c) and (f)). As one can see, after averaging correlation functions over a sufficiently large number M of diffraction patterns, the amplitudes and phases of $\langle C^n_{q_1,q_2}\rangle_M$ asymptotically converge to their statistical estimates. Comparing figures 6(a), (b) and 6(d), (e), one can see that for the same photon fluence the CCFs converge more slowly in the case of the presence of Poisson noise, and one needs to consider more patterns on the average. This effect is even more prominent for the three-point CCFs. That explains our approach to perform the averaging of the CCFs over $M = 10^5$ diffraction patterns.

5. Conclusions

In conclusion, we present here a direct, non-iterative approach for the recovery of the scattering pattern corresponding to a single particle (molecule, cluster, etc) using scattering data from a disordered system of these particles including noise. Our simulations demonstrate the successful application of this approach to 2D systems composed of particles with and without rotational symmetry. It can be of particular interest for studies of membrane proteins which naturally form 2D systems. Our approach can be also applied to 3D systems of particles to recover their projected electron density if a specific alignment of the particles along the direction of the incoming x-ray beam can be achieved. We have shown that our approach is robust to noise and intensity fluctuations that arise due to coherent interference of waves scattered from different particles. We foresee that this method will find wide application in studies of disordered systems at newly emerging FELs.

Acknowledgments

We acknowledge H Franz for a careful reading of the manuscript. Part of this work was supported by the BMBF proposal number 05K10CHG ‘Coherent Diffraction Imaging and Scattering of Ultrashort Coherent Pulses with Matter’ in the framework of the German–Russian collaboration ‘Development and Use of Accelerator-Based Photon Sources’ and the Virtual Institute VH-VI-403 of the Helmholtz Association.

Appendix

The Fourier series expansion of the CCF (10) can be written as

$$C(q_1, q_2, q_3, \Delta_1, \Delta_2) = \sum_{\substack{n_1=-\infty \\ n_1\neq 0}}^{\infty} \sum_{\substack{n_2=-\infty \\ n_2\neq 0}}^{\infty} C_{q_1,q_2,q_3}^{n_1,n_2} \mathrm{e}^{\mathrm{i}n_1\Delta_1} \mathrm{e}^{\mathrm{i}n_2\Delta_2}, \quad n_1 \neq -n_2, \tag{A.1}$$

$$\begin{aligned} C_{q_1,q_2,q_3}^{n_1,n_2} &= \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi}\int_0^{2\pi} C(q_1, q_2, q_3, \Delta_1, \Delta_2)\mathrm{e}^{-\mathrm{i}n_1\Delta_1}\mathrm{e}^{-\mathrm{i}n_2\Delta_2}\,\mathrm{d}\Delta_1\,\mathrm{d}\Delta_2 \\ &= I_{q_1}^{(n_1+n_2)*} I_{q_2}^{n_1} I_{q_3}^{n_2}, \quad n_1 \neq 0, n_2 \neq 0, n_1 \neq -n_2, \end{aligned} \tag{A.2}$$

where $C_{q_1,q_2,q_3}^{n_1,n_2}$ are the Fourier components of the CCF $C(q_1, q_2, q_3, \Delta_1, \Delta_2)$. In general, (A.2) determines the relation between three different Fourier components of intensity I_q^n of the order n_1, n_2 and n_1+n_2, defined on three resolution rings, q_1, q_2 and q_3.

Considering incoherent scattering from N particles, we can rewrite (A.2) in terms of the Fourier components of intensity I_{q,ψ_0}^n associated with a single particle

$$\left\langle C_{q_1,q_2,q_3}^{n_1,n_2} \right\rangle_M = I_{q_1,\psi_0}^{(n_1+n_2)*} I_{q_2,\psi_0}^{n_1} I_{q_3,\psi_0}^{n_2} \cdot \langle \mathbf{A}_{n_1,n_2} \rangle_M, \tag{A.3}$$

where $\langle \ldots \rangle_M$ denotes the statistical averaging and we introduce a new random phasor sum

$$\mathbf{A}_{n_1,n_2} = \sum_{i,j,k=1}^{N} \exp\{\mathrm{i}[(n_1+n_2)\psi_i - n_1\psi_j - n_2\psi_k]\}. \tag{A.4}$$

To understand the statistical properties of $\left\langle \mathbf{A}_{n_1,n_2} \right\rangle_M$ we split it into three contributions

$$\langle \mathbf{A}_{n_1,n_2} \rangle_M = \langle \mathbf{B}_{n_1,n_2} \rangle_M + \langle \mathbf{C}_{n_1,n_2} \rangle_M + \langle \mathbf{D}_{n_1,n_2} \rangle_M, \tag{A.5}$$

where $\langle \mathbf{B}_{n_1,n_2} \rangle_M$, $\langle \mathbf{C}_{n_1,n_2} \rangle_M$ and $\langle \mathbf{D}_{n_1,n_2} \rangle_M$ are defined for different combinations of the subscripts i, j and k:

$$\langle \mathbf{B}_{n_1,n_2} \rangle_M = \left\langle \sum_{i=j=k} \exp\{\mathrm{i}[(n_1+n_2)\psi_i - n_1\psi_j - n_2\psi_k]\} \right\rangle_M = N, \tag{A.6}$$

$$\begin{aligned} \left\langle \mathbf{C}_{n_1,n_2} \right\rangle_M &= \left\langle \sum_{i=j,j\neq k} \cdots + \sum_{i=k,j\neq k} \cdots + \sum_{i\neq j,j=k} \cdots \right\rangle_M \\ &= 2\left\langle \sum_{i>j} \cos[n_1(\psi_i - \psi_j)] \right\rangle_M + 2\left\langle \sum_{i>j} \cos[n_2(\psi_i - \psi_j)] \right\rangle_M \\ &\quad + 2\left\langle \sum_{i>j} \cos[(n_1+n_2)(\psi_i - \psi_j)] \right\rangle_M, \end{aligned} \tag{A.7}$$

$$\left\langle \mathbf{D}_{n_1,n_2} \right\rangle_M = \left\langle \sum_{i\neq j\neq k} \exp\{\mathrm{i}[(n_1+n_2)\psi_i - n_1\psi_j - n_2\psi_k]\} \right\rangle_M. \tag{A.8}$$

Now, we determine the statistical estimates of the terms $\langle \mathbf{C}_{n_1,n_2} \rangle_M$ and $\langle \mathbf{D}_{n_1,n_2} \rangle_M$. We consider a uniform distribution of orientations of particles ψ on the interval $(-\pi, \pi)$, where two different orientations ψ_i and ψ_j are independent and equally probable. We also consider that the sum (or difference) of two angles ψ, as well as a product $n\psi$ (where n is an arbitrary number), is also distributed on the interval $(-\pi, \pi)$. In other words, we deal with a wrapped angular distribution [20]. In this case, we can use the following two arguments. Firstly, the difference $\bar{\psi} = \psi_i - \psi_j$ of two uniformly distributed independent variables ψ_i and ψ_j also has a uniform distribution. In this case the probability density function (PDF) of $\bar{\psi}$ is a convolution of the corresponding PDFs of ψ_i and ψ_j [21], which is just a constant after wrapping to the interval $(-\pi, \pi)$. Secondly, using the rules of probability theory for transformation of random variables [21, 22], it can be shown that the product $n\psi$ is also a uniformly distributed variable. Using a combination of these two arguments in (A.7) and (A.8), we find that $\langle \mathbf{C}_{n_1,n_2} \rangle_M = 0$ and $\langle \mathbf{D}_{n_1,n_2} \rangle_M = 0$. Therefore, the statistical estimate in (A.5) becomes a real number, $\langle \mathbf{A}_{n_1,n_2} \rangle_M = N$. Using this result in (A.3) we obtain (13).

References

[1] Wochner P, Gutt C, Autenrieth T, Demmer T, Bugaev V, Diaz-Ortiz A, Duri A, Zontone F, Grübel G and Dosch H 2009 *Proc. Natl Acad. Sci. USA* **106** 11511–14

[2] Altarelli M, Kurta R P and Vartanyants I A 2010 *Phys. Rev.* B **82** 104207

[3] Kurta R P, Altarelli M, Weckert E and Vartanyants I A 2012 *Phys. Rev.* B **85** 184204

[4] Emma P *et al* 2010 *Nature Photon.* **4** 641–7

[5] Ishikawa T *et al* 2012 *Nature Photon.* **6** 540–4

[6] Altarelli M *et al* 2007 *The European X-Ray Free-Electron Laser Technical Design Report* http://xfel.desy.de/tdr/tdr

[7] Gaffney K J and Chapman H N 2007 *Science* **316** 1444–8

[8] Seibert M M *et al* 2011 *Nature* **470** 78–82

[9] Kam Z 1977 *Macromolecules* **10** 927–34

[10] Kam Z 1980 *J. Theor. Biol.* **82** 15–39

[11] Saldin D K *et al* 2010 *New J. Phys.* **12** 035014

[12] Saldin D K, Poon H C, Shneerson V L, Howells M, Chapman H N, Kirian R A, Schmidt K E and Spence J C H 2010 *Phys. Rev.* B **81** 174105

[13] Saldin D K, Poon H C, Schwander P, Uddin M and Schmidt M 2011 *Opt. Express* **19** 17318–35

[14] Saldin D K, Poon H C, Bogan M J, Marchesini S, Shapiro D A, Kirian R A, Weierstall U and Spence J C H 2011 *Phys. Rev. Lett.* **106** 115501

[15] Fienup J R 1982 *Appl. Opt.* **21** 2758–69

[16] Elser V 2003 *J. Opt. Soc. Am.* **20** 40–55

[17] Oppenheim A V, Sharfer R W and Buck J R 1999 *Discrete-Time Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall)

[18] Goodman J W 2007 *Speckle Phenomena in Optics: Theory and Applications* (Greenwood Village, CO: Roberts and Company)

[19] Kirian R A, Schmidt K E, Wang X, Doak R B and Spence J C H 2011 *Phys. Rev.* E **84** 011921

[20] Mardia K V and Jupp P E 2000 *Directional Statistics (Wiley Series in Probability and Statistics)* (Chichester: Wiley)

[21] Goodman J W 2000 *Statistical Optics* (New York: Wiley)

[22] Rohatgi V K, Saleh A K and MD E 2000 *An Introduction to Probability and Statistics* 2nd edn (*Wiley Series in Probability and Statistics*) (Hoboken, NJ: Wiley)