Supplementary material to

Two-dimensional structure from random multi-particle X-ray scattering images using cross-correlations

B. Pedrini, A. Menzel, M. Guizar-Sicairos, V. A. Guzenko, S. Gorelick, C. David, B. D. Patterson and R. Abela

Paul Scherrer Institute, 5232 Villigen PSI, Switzerland
Supplementary Figure S1. Diagonalization of 2-point CC matrices. The results are shown for \( n = 6 \) (upper plot) and \( n = 2 \) (lower plot). The plots display the components of the eigenvector \( s_{n,a} \) corresponding to the dominant eigenvalue of the matrix \( C_{n,ab} \), as a function of \( q_a \). The data are rescaled with \( s_0(q_a) \). The following four \( q_a \)-sets were considered: 0.03 – 0.12 nm\(^{-1}\), \( N_{q_a} = 61 \) equidistant points (solid lines), 0.06 – 0.12 nm\(^{-1}\), \( N_{q_a} = 31 \) equidistant points (squares), 0.03 – 0.09 nm\(^{-1}\), \( N_{q_a} = 31 \) equidistant points (diamonds), and 0.03 – 0.12 nm\(^{-1}\), \( N_{q_a} = 31 \) equidistant points (dots). The insets show the full set of eigenvalue ratios \( \lambda_{n,i}/\lambda_{n,1} \), ordered from largest to smallest, as a function of \( 1 - (i-1)/N_{q_a} \), such that the largest eigenvalue appears at unity.
Supplementary Figure S2. Non-valid single-particle diffraction intensity components. (a) Non-valid $s_n$ as a function of $q$, calculated for the $q_a$-range $0.03 - 0.24$ nm$^{-1}$. The data are shown for $n = 2$ (violet) and $n = 4, 8, 10, 14, 16, 20$ (red). The blue dashed line corresponds to $n = 12$ and is shown for comparison. (b)-(e) Complex plane representations of the ratios $R_{nm}(q_a, q_b, q_c)$ (see article) arising from the 3-point CC identity, for fixed $q_a = 0.066$ nm$^{-1}$ and $q_b = 0.115$ nm$^{-1}$, and for a large number of $q_c$. The centroid points in orange are the median of all points, and the orange circles enclose the areas containing the closest 80% of all points. In (b), $n$ and $m$ are multiples of 6. In (c)-(e), at least one of the three integers $n$ and $m$ is not a multiple of six, which corresponds to a non-valid component.
 Supplementary Figure S3. Non-identical particles. (a) Single particle diffraction intensity determined following the CC-based protocol, using as input 1000 simulated scattering images, each from 10 identical, randomly oriented crosses. (b) Reconstructed 2D shape using the data in (a) as input for the phasing algorithm, and imposing the compact support region delimited by the dashed white square. The shape of the cross is given by the ratio \( b : h = 3 : 5 \) between of the arm width \( b \) and the cross half-height \( h \). (c) Same as (a), but with simulations performed by using crosses whose dimensions \( b \) and \( h \) were allowed to vary independently, both following a Gaussian distribution with 5\% FWHM. (d) Same as (b), but using the data in (c) as input for the phasing algorithm.
Supplementary Figure S4. Noise spikes in the 2-point CC. $C^{(2)}(\psi)$ computed from 10 (upper panel) and 1000 (lower panel) different simulated Poisson noise realizations of $S(\phi) = s_0$, for $s_0 = 0.01$ (green), $s_0 = 0.1$ (blue), $s_0 = 1$ (cyan), $s_0 = 10$ (magenta), and $s_0 = 100$ (red). The discretization of the angular variables was performed with $N_\phi = 64$. 
Supplementary Figure S5. Beam profile. (a),(b) The blue lines show a Bragg peak intensity $I(\xi)$ measured by scanning transversally the periodic reference structure in the horizontal direction (a) and in the vertical direction (b), as a function of the transverse coordinate $\xi$. The red graphs are the predictions $I_{\Lambda,\zeta_0}(\xi)$ for a structure of width $\zeta_0=100 \, \mu m$, and assuming the transverse beam profile $\Lambda$ to be Lorentzian with the different indicated FWHMs. (c) SEM image of a corner of the $100 \times 100 \, \mu m$, periodic scattering structure used to assess the beam profiles.
Supplementary Figure S6. Determination of the phases of the coefficients $s_n$. (a) 2D shape of the object used for the simulation. (b) Simulated single-particle diffraction pattern of the object in (a), calculated assuming the same setup as for the experiments. (c) Example of a simulated diffraction pattern of $N = 5$ identical, randomly oriented objects. (d),(e) Amplitudes $|s_n|$ and start phases $\varphi_n^{(0)}$ as a function of $q$ for $n=2$, $n=12$ and $n=22$, calculated exploiting the 2-point CCs from 10000 diffraction images with $N = 5$. In (d), the orange curves correspond to the amplitudes extracted from the single-particle diffraction image. (f) Correction phases $\chi_n$ calculated following the procedure described in the Supplementary Methods. $\chi_2 = 0$ was kept fixed. The error bars account for the small variability of the computed solutions. (g) Corrected phases $\varphi_n$ as a function of $q$ for the same three $n$-values as in (d). The orange curves are for the phases extracted from the single-particle diffraction image (b). (h) Correction phases $\chi_n$ calculated when the full CC procedure was applied to one single-particle diffraction image. (i) Correction phases $\chi_n$ calculated by application of the full CC protocol to the experimental data from the 350 nm sized nanostructures. $\chi_6$ was kept fixed.
Supplementary Note 1

Proof of the cross-correlation identity in 2D, and effect of particle alignment

Assuming the absence of inter-particle interference, the diffraction image of \( N \) particles in 2D with orientations \((\alpha_1, \ldots, \alpha_N)\) is the sum of the images from each particle,

\[
I(q, \phi) = \sum_n e^{-in\phi} \cdot \left( \sum_{i=1}^N e^{-in\alpha_i} \right) \cdot I_n(q),
\]

and is independent of the particles positions. \( I_n(q) \) are the \( \phi \)-Fourier components of the intensity \( I(q, \phi) \) in the images.

An equivalent definition of the 2-point CC to that given in the article is

\[
c^{(2)}(q_1, q_2) = \langle I_n(q_1) I_n(q_2)^* \rangle \quad (n \neq 0).
\]

\( \langle \cdots \rangle \) denotes an average over the images, and is equivalent to the average over the multi-particle orientations. Inserting (S1) into (S2) gives

\[
c^{(2)}(q_1, q_2) = \sum_n s_n(q_1)s_n(q_2)^* \cdot \sum_{i,j=1}^N \langle e^{-in(\alpha_i - \alpha_j)} \rangle = \sum_n s_n(q_1)s_n(q_2)^* \cdot \left( N + \sum_{i,j=1; i \neq j}^N \langle e^{-in(\alpha_i - \alpha_j)} \rangle \right).
\]

For randomly oriented particles, the term \( \varepsilon_n \) vanishes if averaged over the particles’ orientations (or, equivalently, in the limit of an infinite number of images). This is not true in the case of particles with preferred orientation. The term \( \varepsilon_n \) accounts for the amount of partial alignment, and is in general \( n \)-dependent. In the extreme limit of full particle alignment, \( \varepsilon = N(N - 1) \), and the term in square brackets on the right hand side of (S3) becomes \( N^2 \) and is again \( n \)-independent.

Supplementary Note 2

Effect of inter-particle interference on the cross-correlations

Inter-particle interference causes a double spike in the diagonal 2-point CC, which is proportional to the square of the number of particles in a coherence area. Here we prove this for the case of isotropic scatterers, where the single-particle diffraction intensity is given by \( S(q, \phi) = s_0(q) = a_0(q)a_0^*(q) \). In absence of interference, the higher-order correlations vanish. Taking interference into account, \( N \) particles located at positions \((\vec{x}_1, \ldots, \vec{x}_N)\) will produce an
image

\[ I(q, \phi) = a_0(q) a_0^*(q) \left[ N + \sum_{i \neq j} e^{-i q r_{ij} \cos(\alpha_{ij})} \right] = \]
\[ = s_0(q) \left[ N + \sum_{i \neq j} \sum_n c^n e^{-i n \alpha_{ij}} J_n(q r_{ij}) \right]. \quad (S4) \]

Here, \( \vec{x}_{ij} = r_{ij}(\cos(\alpha_{ij}); \sin(\alpha_{ij})) \) is the 2D distance vector between particles \( i \) and \( j \), and \( J_n(x) \) is the \( n \)-th Bessel function. Inserting (S4) into the 2-point CC definition (S2) gives

\[ c_n^{(2)}(q_1, q_2) = s_0(q_1) s_0(q_2) \sum_{i \neq j, k \neq l} \langle e^{-i n \alpha_{ij}} J_n(q_1 r_{ij}) e^{+i n \alpha_{kl}} J_{-n}(q_2 r_{kl}) \rangle \quad . \quad (S5) \]

The image average \( \langle \cdots \rangle \) is equivalent to the average over the particles positions. Nonvanishing contributions appear only if the four indices are pairwise the same. For \( i = k \) and \( j = l \), one obtains

\[ \sum_{i \neq j} \langle J_n(q_1 r_{ij}) J_n(q_2 r_{ij})^* \rangle = N(N - 1) \int_0^\infty dr \, r \, J_n(q_1 r) J_{-n}(q_2 r) = \]
\[ = N(N - 1) \frac{1}{q_1} \delta(q_1 - q_2) \quad , \quad (S6) \]

while for \( i = l \) and \( j = k (\alpha_{ij} = \alpha_{kl} + \pi) \), one obtains

\[ \sum_{i \neq j} \langle J_n(q_1 r_{ij}) J_n(q_2 r_{ij})^* \rangle = N(N - 1) \int_0^\infty dr \, r \, J_n(q_1 r) J_{-n}(q_2 r) \cdot e^{-i \pi} = \]
\[ = N(N - 1) \frac{1}{q_1} \delta(q_1 - q_2) \cdot (-1)^n \quad . \quad (S7) \]

Summing (S6) and (S7) gives the correction term to the 2-point CC identity,

\[ c_{n,\text{interference}}^{(2)}(q_1, q_2) = N(N - 1) \cdot s_0(q_1) s_0(q_2) \cdot \frac{1}{q_1} \delta(q_1 - q_2) (1 + (-1)^n) \quad , \quad (S8) \]

and therefore

\[ C^{(2,\text{interference}}(q_1, q_2, \psi) = N(N - 1) \cdot s_0(q_1) s_0(q_2) \cdot \frac{1}{q_1} \delta(q_1 - q_2) [2\pi \delta(\psi) + 2\pi \delta(\psi - \pi)] \quad , \quad (S9) \]

which for \( q_1 = q_2 \) results in two spikes of equal height at \( \psi = 0 \) and \( \psi = \pi \).

For a general single-particle scattering amplitude \( A(q, \phi) = \sum_{n \in \mathbb{Z}} a_n e^{-i n \phi} \), the single-particle diffraction intensity is \( S(q, \phi) = A(q, \phi) A(q, \phi)^* \), with \( \phi \)-Fourier coefficients given by

\[ s_n(q) = \sum_{m \in \mathbb{Z}} a_{m+n}(q) a_m^*(q). \]

The calculation of the 2-point CC in this general case can be performed along the same lines as above.
Supplementary Note 3

Effect of shot noise on the cross-correlations

Shot and background noise cause a spike in the diagonal 2-point CC,
\[
C^{(2), \text{noise}}(q_1, q_2, \psi) = \nu(q_1) \cdot \delta(q_1 - q_2)/q_1 \cdot \delta(\psi),
\]
where \(\nu(q_1)\) is a measure of the noise level at a fixed \(q\). This is reflected in a noise correction term to the Fourier coefficients
\[
c_n^{(2), \text{noise}}(q_1, q_2) = \nu(q_1) \cdot \delta(q_1 - q_2)/q_1 \quad (n \neq 0).
\]

The generation of the spike is illustrated using the following simple model: Let \(\phi_k\) be the discrete angular variable on a scattering ring \((\phi_k = k \cdot 2\pi/N_\phi)\), and let \(S(\phi_k) = s_0\) be an “isotropic single-particle diffraction intensity”. The intensities of different pixels \(\phi_k\) in different “single-particle images”, \(I(\phi_k) = \text{Poisson}[s_0]\), follow a Poisson distribution, with \(s_0\) the mean value in each pixel. Examples of 2-point CC \(C^{(2)}(\psi_k)\) resulting from different numbers of such “images” are shown in Supplementary Figure S4 and clearly exhibit the spike at \(\psi_k = 0\).

Supplementary Note 4

Accounting for the inhomogeneous illumination function

Let \(S(q, \phi) = \sum_n s_n(q) e^{-in\phi}\) be the scattering intensity of a single particle when it is illuminated by a beam of uniform intensity \(I_0\), and let the real beam intensity \(I\) in the 2D sample plane be modulated by a function \(\Omega\), i.e., \(I(x, y) = I_0 \cdot \Omega(x, y)\). The 1-point CC identity is
\[
C^{(1)}(q) = \langle \sum_i \Omega(x_i, y_i) \rangle \cdot s_0(q) =: N,
\]
where the index \(i\) labels the particles producing a scattering image, \((x_i, y_i)\) are the 2D coordinates of the \(i\)-th particle, and \(\langle \cdots \rangle\) denotes an average over the images. The “effective” number of particles \(N\) in the beam is hereby fixed to be
\[
N = \nu \int dxdy \Omega(x_i, y_i) =: \nu \int dxdy \Omega(x_i, y_i) =: N_k^{(2)}(q),
\]
where \(\nu\) is the 2D surface particle density. Similarly, the 2-point CC identity reads
\[
c_n^{(2)}(q_1, q_2) = \langle \sum_i \Omega(x_i, y_i)^2 \rangle \cdot s_n(q_1) s_n^*(q_2) =: N_k^{(2)}(q_1, q_2),
\]
and a similar discussion applies to the 3-point CC identity. The beam shape dependent constants \(k^{(2)}\) and \(k^{(3)}\) are given by
\[
k^{(j)} = \frac{\langle \sum_i \Omega(x_i, y_i)^j \rangle}{\langle \sum_i \Omega(x_i, y_i) \rangle} = \int dxdy \Omega(x_i, y_i)^j/\int dxdy \Omega(x_i, y_i),
\]
and depend only on the beam shape but not on the beam width (i.e., they are invariant under dilatations).

For a double Lorentzian beam shape

$$\Omega(x, y) = \frac{\delta h \cdot \delta v}{(x^2 + \delta h^2)(y^2 + \delta v^2)}, \quad (S16)$$

the effective number of particles $N$ is related to the 2D surface particle density by

$$\frac{N}{\nu \cdot \delta h \cdot \delta v} = \left( \frac{\pi}{2} \right)^2 \quad . \quad (S17)$$

The values $\kappa(2) = 0.26$ and $\kappa(3) = 0.14$, which are independent of $\delta h$ and $\delta v$, were determined from a numerical integration.
Supplementary Methods

Diagonalization procedure to solve the 2-point CC equations and validation of the single-particle diffraction Fourier coefficients

We use here the experimental data measured with the 350 nm particles, with a surface density of 10 particles per 100 µm². Supplementary Figure S1 shows the results of diagonalizations of the Hermitian matrix $C_{n,ab}^{(2)}(q_a, q_b)$ defined by 2-point CC coefficients for a discrete set of $q_a$. The $a$-th component $s_{n,a}$ of the eigenvector corresponding to the largest eigenvalue $\lambda_{n,1}$ is shown as a function of $q_a$, for four different $q_a$-sets in the range $0.03 - 0.12$ nm$^{-1}$. The inset displays the relative magnitude of the computed eigenvalues. The upper plot is for $n = 6$, which exemplifies the result obtained for all $n$ which are multiples of 6. That the four graphs match perfectly and that there is always one dominant eigenvalue indicate consistency of the experimental data. This is clearly not the case for $n = 2$, as demonstrated by the data shown in the lower plot. The same occurs for all coefficient orders which are not multiples of 6.

Supplementary Figure S2(a) shows the amplitudes of single-particle scattering Fourier coefficients calculated from $s_n(q_a) = s_{n,a}/\sqrt{N}$, where $s_{n,a}$ is the eigenvector corresponding to the largest eigenvalue of the $C_{n,ab}$, calculated for different sets of $q_a$-values in the range $0.03 - 0.24$ nm$^{-1}$. The values for all the coefficients which are not multiples of 6 are smaller than those which are multiples of 6 ($s_{12}$ is shown as reference), with the exception of $n = 2$. Panels (b)-(e) in Supplementary Figure S2 display the ratios $R_{nm}(q_a, q_b, q_c)$ (see article text) as points in the complex plane. In (b), $n$ and $m$ are multiples of 6, and the fact that most of the points cluster at one real positive value is taken as a validation of the calculated $s_n(q_a)$. The plots (c)-(e) are for ratios for which $n$, $m$ or $n + m$ is $\pm 2$, and present precisely the opposite features, which supports the exclusion of $s_n(q)$ to the single-particle diffraction intensity.

Experimental assessment of the X-ray beam profile at the sample

A 100×100 µm periodic array of gold nanostructures, with hexagonal 500 nm unit cell (Supplementary Figure S5(c)), was scanned in the $x$- and $y$-directions, and the intensity $I(\xi)$ of one of the low-order Bragg peaks was determined as a function of the horizontal or vertical displacement $\xi$. $I(\xi)$ is expected to be given by

$$I_{\Lambda,\zeta_0}(\xi) = \int_{-\zeta_0/2}^{\zeta_0/2} d\zeta \, \Lambda(\zeta - \xi), \quad (S18)$$

where $\Lambda(\xi)$ is the beam profile and $\zeta_0$ is the width of the scattering structure. Supplementary Figure 5(a,b) shows comparisons of $I(\xi)$ with $I_{\Lambda,\zeta_0}(\xi)$, for $\zeta_0 = 100$ µm, and Lorentzian beam profiles $\Lambda(\xi) = \delta/(\xi^2 + \delta^2)$, for various $\delta$. The FWHM in the horizontal and vertical directions are estimated to be $2\delta \cdot h = 14 \pm 2$ µm and $2\delta \cdot v = 6 \pm 2$ µm, respectively.
Determination of the $s_n$ phases using the 3-point CC equations

The eigenvector components $s_{n,a}$ are defined up to a phasor that depends only on $n$:

$$s_{n,a} = |s_{n,a}| \cdot e^{-i(\varphi_{n,a}^{(0)} + \chi_n)} , \quad (S19)$$

$\chi_n$ is a phase shift that transforms a set of "starting" phases $\varphi_{n,a}^{(0)}$ into the correct phases $\varphi_{n,a}$:

$$\varphi_{n,a} = \varphi_{n,a}^{(0)} + \chi_n . \quad (S20)$$

In practice, the corrections $\chi_n$ can be determined from the zero phase requirement on the ratio $R_{nm}(q_a, q_b, q_c)$ (see article text), which links different values of $n$:

$$\Phi_{nm,abc} = (\varphi_{n,a}^{(0)} + \chi_n) + (\varphi_{m,b}^{(0)} + \chi_m) - (\varphi_{(n+m),c}^{(0)} + \chi_{(n+m)}) \mod 2\pi . \quad (S21)$$

$\Phi_{nm,abc}$ is the phase of $c_{nm}^{(3)}(q_a, q_b, q_c)$. Solutions of equations (S21), or approximate solutions in case of imperfect experimental data, are easily determined, because the unknowns $\chi_n$ appear to first order, a direct consequence of the generalized definition of the 3-point CCs.

We concretely illustrate the procedure applied to the case of the 2D structure shown in Supplementary Figure S6(a), which is an example of a general shape, devoid of mirror symmetry. The input dataset consisted of 10000 simulated diffraction images (see an example in Supplementary Figure S6(c)), each from $N = 5$ randomly oriented particles. The coefficients $s_n$ were calculated to order $n = 32$, and up to the overall correction phases $\chi_n$, exploiting the 2-point CCs. Examples of amplitudes and phases of such $s_n$ are displayed in Supplementary Figures S6(d) and S6(e), respectively. Next, $\chi_2 = 0$ was set to fix the orientation of the single-particle diffraction pattern to be reconstructed, leaving the unknowns $\chi_4, \chi_6, \ldots, \chi_{30}, \chi_{32}$, whose number is $N_\chi = 15$. Three "high-signal" $q$ values, close to local maxima of $s_n(q)/s_0(q)$, were then selected for each $n$ (see magenta arrows in Supplementary Figure S6(d)). The combinations of the form (S21) for all considered $n, m, q_a, q_b$ and $q_c$ take the form

$$a_{r,4} \chi_4 + \ldots + a_{r,32} \chi_{32} = \Phi_r - (a_{r,4} \varphi_{4,r}^{(0)} + \ldots + a_{32,r} \varphi_{32,r}^{(0)}) - a_{r,2} \varphi_{2,r}^{(0)} \mod 2\pi . \quad (S22)$$

with suitable coefficients $a_{r,n}$ and $\varphi_{n,r}^{(0)}$ that depend on the index $r$ of the combination. For convenience, (S22) can be written in matrix notation using $r$ as the row index:

$$A \cdot \tilde{\chi} = \tilde{B} \mod 2\pi , \quad (S23)$$

The following procedure was then repeated a number $N_{\text{iter}}$ of times. First, $N_\chi$ row indices in (S23) are randomly selected; if the reduced matrix $\tilde{A}$ is invertible, i.e.

$$\text{rank}(\tilde{A}) = N_\chi , \quad (S24)$$

then a possible solution is calculated:

$$\tilde{\chi}_s = \tilde{A}^{-1} \cdot \tilde{B} \mod 2\pi , \quad (S25)$$
where \( \tilde{B} \) is the reduced vector of \( \tilde{B} \). The solution is accepted if all rows of (S23) are fulfilled within a small tolerance. The average of the accepted solutions from the different iterations yields the final phase correction \( \tilde{\chi} \):

\[
e^{-i\tilde{\chi}} = \langle e^{-i\tilde{\chi}_s} \rangle_{s=1..N_{\text{iter}}},
\]

which is shown in Supplementary Figure S6(f) for the for the considered dataset based on the h-shaped object. Supplementary Figure 6(g) displays examples of the corrected phases \( \varphi_n \). The latter match extremely well the values derived directly from the single-particle diffraction image, proving that the presented protocol is in general applicable. The spread in the correction phases of Supplementary Figure 6(f), more marked at large \( n \), is interpreted as due to the fact that full convergence of the 3-point CC is not achieved with the number of images used. This conclusion can be drawn from comparison with Supplementary Figure S6(h), corresponding to the output of the full CC procedure applied to one diffraction image from one h-shaped structure. There, the fact that for a single particle one image is sufficient for accessing all \( n \)-point CCs in 2D is reflected in the minimal spread of the correction phases. For sake of completeness, Supplementary Figure S6(j) is devoted to the experimental case of the 350 nm nanostructures discussed in the article. In this case, \( \chi_6 \) was fixed at the beginning, and the unknown correction phases to be computed were \( \chi_{12}, \chi_{18}, \chi_{24} \) and \( \chi_{30} \).