BNL-114233-2017-JA

Robust X-ray Angular Correlations for the Study of Meso-structures

K. G. Yager

Submitted to the Journal of Applied Crystallography

March 2017

Center for Functional Nanomaterials

Brookhaven National Laboratory

**U.S. Department of Energy
USDOE Office of Science (SC),
Basic Energy Sciences (SC-22)**

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No. DE- SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Robust X-ray Angular Correlations for the Study of Meso-structures

JULIEN R. LHERMITTE,^a CHENG TIAN,^a AARON STEIN,^a ATIKUR RAHMAN,^a
YUGANG ZHANG,^b LUTZ WIEGART,^b ANDREI FLUERASU,^b OLEG GANG^{a,c,d} AND
KEVIN G. YAGER^{a*}

^a*Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, New York 11973 US*, ^b*National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York 11973 US*, ^c*Department of Chemical Engineering, Columbia University, New York, NY, 10027, USA* and ^d*Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, 10027 US*.

E-mail: kyager@bnl.gov

Abstract

As self-assembling nanomaterials become more sophisticated, it is becoming increasingly important to measure structural order of finite-size assemblies of nano-objects. These mesoscale clusters represent an acute challenge to conventional structural probes, owing to the range of implicated size scales (10 nm to several microns), the weak scattering signal, and the dynamic nature of meso-clusters in native solution environments. The high x-ray flux and coherence of modern synchrotrons presents an opportunity to extract structural information from these challenging systems, but conventional ensemble x-ray scattering averages out crucial information about local particle configurations. Conversely, a single meso-cluster scatters too weakly to recover the full

diffraction pattern. Using x-ray angular cross correlation analysis (XCCA), it is possible to combine multiple noisy measurements to obtain robust structural information. We explore the key theoretical limits and experimental challenges that constrain the application of these methods to probing structural order in real nanomaterials. We present a metric to quantify the signal-to-noise of angular correlations, and use it to identify several experimental artifacts that arise. In particular, we find that background scattering, data masking, and inter-cluster interference profoundly affect the quality of correlation analyses. We demonstrate a robust workflow for mitigating these effects, and extracting reliable angular correlations from realistic experimental data.

1. Introduction

Assembly of nanoparticles into nanoscale constructs is an emerging strategy for designed nanomaterials. Enormous progress has been made in the assembly of nanoparticles into extended, three-dimensional superlattices, exploiting nanoparticle ligands (Ye *et al.*, 2015; Murray *et al.*, 2000; Redl *et al.*, 2003) or programmable DNA linkers (Mirkin *et al.*, 1996; Park *et al.*, 2008; Nykypanchuk *et al.*, 2008; Zhang *et al.*, 2013). Recently, these methods have advanced towards the creation of discrete, finite-sized assemblies of particles—that is, meso-clusters containing a well-defined number of nanoparticles in a precise spatial arrangement (Tian *et al.*, 2015; Tan *et al.*, 2014; Zhang *et al.*, 2014; Sharma *et al.*, 2009; Lo *et al.*, 2010). Continued progress in materials design, however, is limited by the challenge of measuring structural order in these ultra-small meso-clusters, especially in their native environment (e.g. water solution). Probing *in-situ* is especially important in the context of dynamic and reconfigurable nanomaterials, which exploit the soft linkers to program material response (Zhang *et al.*, 2015; Pal *et al.*, 2015; Maye *et al.*, 2010)). While scanning or transmission electron microscopy (SEM and TEM) have sufficient resolution, their application to

liquid environments is non-trivial. Scanning probe microscopy can be performed in liquid environments, but can only probe relatively static, surface-attached objects. X-ray scattering methods have unique advantages with respect to rapidly probing nanomaterials in liquid environments, having sufficient flux and penetration to accommodate macroscopic sample cells.

Small-angle X-ray Scattering (SAXS) has been demonstrated to be a powerful technique for resolving structural order in extended nanoscale superlattices (Förster *et al.*, 2005; Yager *et al.*, 2014; Senesi & Lee, 2015). However, the powder-averaging inherent to ensemble SAXS obscures some aspects of order. For instance, local particle packing motifs can only be determined indirectly—by assigning a super-crystalline unit cell and fitting the data to ascribe particle positions. For finite-size meso-clusters, such approaches generally fail, since the isotropically-averaged scattering intensity (the form factor of the meso-cluster) lacks sufficient definition to unambiguously assign local particle geometries. Conversely, coherent diffraction imaging can reconstruct the exact meso-cluster structure (Sun *et al.*, 2012; Putkunz *et al.*, 2011; Vartanyants & Robinson, 2001; Takahashi *et al.*, 2013; Williams *et al.*, 2006) but requires a static sample removed from its native environment. An alternative approach is to avoid ensemble averaging during a SAXS measurement, by probing a very small number of scattering entities within the beam cross-section. In such experiments far from the ensemble limit, the signal-to-noise ratio (S/N) is generally exceedingly low, making direct analysis of scattering patterns problematic. The high peak flux of X-ray Free Electron Lasers (XFELs) has been used for “diffract and destroy” experiments, where scattering snapshots can in principle be assembled to reconstruct the 3D reciprocal-space. This measurement mode may not be suitable for all in-situ studies; moreover, even with high flux some systems may present highly noisy data-frames. However, correlation analyses can be used to extract statistical information from each noisy

scattering frame; the average of a particular correlation signal across a great number of frames (even with a dynamic sample) becomes a robust reporter of structural order. In particular, angular correlations along a scattering ring encode information about the local symmetry and particle packing in meso-clusters. This X-ray Cross Correlation Analysis (XCCA) has gained considerable attention recently, owing to its promise with respect to probing local particle configurations in both ordered and disordered nanomaterials (Donatelli *et al.*, 2015; Mancini *et al.*, 2016; Zaluzhnyy *et al.*, 2015).

Angular correlations were introduced for electron microscopy (Kam, 1977*a*), where calculation of two- and three-point angular correlation functions enabled reconstruction of the 3D structure of randomly oriented samples. The idea was adapted to x-ray measurements, first through simulations (Saldin *et al.*, 2010*a*; Elser, 2011; Saldin *et al.*, 2011*b*; Saldin *et al.*, 2010*b*; Saldin *et al.*, 2010*c*) and subsequently experimentally for nano rods (Saldin *et al.*, 2011*a*), dumbbells (Chen *et al.*, 2012; Starodub *et al.*, 2012) and triangles (Pedrini *et al.*, 2013). Considerable theoretical development has gone into better understanding how the structure of samples relates to their angular correlation functions (Kurta *et al.*, 2012; Altarelli *et al.*, 2010). There has been corresponding progress in reconstructing real space images of samples from the measured angular correlations (Donatelli *et al.*, 2015; Fienup, 1982). However, reconstruction is possible only in the limit of high-fluence measurements; that is, for strongly-scattering, static samples measured using a long x-ray exposure.

Despite considerable progress in the theoretical development of angular correlation analysis, these methods have seen little uptake in x-ray scattering experiments. Here, we explore the key theoretical limits and experimental challenges that constrain the application of XCCA methods to probing real nanomaterials. Using numerical simulations and synchrotron experiments, we identify several specific artifacts that are likely to arise when applying correlation analyses, and present methods—applicable during

data collection and during post-analysis—for mitigating these effects. We identify the regimes where the technique can be expected to work, and present the experimental and analysis workflows necessary in order to obtain reliable data about local order in meso-clusters.

In order to highlight certain key artifacts to the analysis, results from real and simulated experiments will be presented. The simulated experiments will contain the expected artifactual noise expected in experimental data. Finally, the goodness of correlation will be then compared to a known model. The model should not be confused with the simulated results, which are a measurements of the model with noise introduced. To distinguish them, the former will be referred to as model and the latter as synthetic data.

2. Methods

2.1. Workflow

The general workflow of the experiment is presented in figure 1. A series of far field xray diffraction patterns are measured on a 2-dimensional area detector (top left). These series are averaged together to obtain an average image (top right). The series of images and the average image are re-partitioned into a polar grid (second row from top). The polar grid of the average image is subtracted from the polar grid of the series (third row from top). The angular correlation is computed on this polar grid for each polar grid in the series (bottom left). These are averaged to obtain the final correlation functions (bottom right). In order to score the quality of the analysis, the analyzed data is compared to a known model. An example of a comparison of a cross section of real data (magenta) and a known model (blue) is presented in the bottom right panel. In this case, the model was taken by transforming measured real space images into the expected diffraction patterns. When a more accurate score is needed,

a more precise model will be used. Throughout the process, there is missing or invalid data. These are masked and thus ignored throughout the process. This pipeline was written in the Python language (Oliphant, 2007; Lhermitte, 2017).

This paper focuses on various components of this pipeline. The following subsections will detail the experiment and analysis. The next section will outline in detail the importance of various steps in the pipeline, through either measured or simulated data.

Fig. 1. Workflow for computing angular correlation functions. First, detector images are masked (m) and re-meshed into a $I(q, \phi)$ map using a selected partitioning scheme (p). Partitioning is finer than the example (blue grid) shown in the figure. Concurrently, the masked detector images are averaged (a) to obtain an estimate of the ensemble scattering. This average is subtracted (s) from each image, which removes the background and highlights the correlation signal. Each image is then correlated (c) along the angular direction (ϕ), resulting in a $C(\Delta\phi)$ 2D correlation function. Finally, these correlation functions are averaged (a) to obtain a robust correlation function. A cross-section is plotted (magenta) with a comparison of the correlation computed by running the same process on Fourier transformed SEM images of the same sample (blue). The SEM image can be found in figure 2c. The mask (white regions) is also processed in parallel through each step, and used to properly account for the number of pixels averaged per $(q, \Delta\phi)$ pair (not shown).

2.2. Synchrotron measurements

Synchrotron experiments were performed at the Coherent Hard X-ray (CHX, 11-ID) undulator beamline at the National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory (Fluerasu *et al.*, 2011). X-ray energy was set to 8.9 keV (1.4 Å wavelength) using a Si111 double-crystal monochromator, yielding an energy resolution of $\frac{\Delta E}{E} \approx 10^{-4}$. A set of 1D Be compound refractive lenses and a Si kinoform lens were used to focus the beam to a spot size of approximately $10 \mu\text{m} \times 12 \mu\text{m}$ on the sample in the horizontal and vertical directions (Chubar *et al.*, 2013). Samples were measured in transmission, with the small-angle scattering (TSAXS) collected on a solid-state pixel-array area detector (Dectris Eiger 4MP), featuring a pixel size of $75 \mu\text{m}$. The sample-detector distance was set to 4.87 m.

A reference sample of disordered porous glass (ColarPor[®], Schott) was measured to establish the coherence of the setup; a 10% contrast was measured, suggesting that the coherence volume of the sample is at least 1/10th of the diffraction volume (Sutton, 2008). At small angle, the longitudinal coherence length is set by the path-length difference $s < \frac{\lambda^2}{\Delta\lambda} \frac{2L^2}{x^2} \approx 3 \text{ mm}$ where L is the sample detector distance and x is the maximum transverse length on the detector ($2,000 \text{ pixels} \times 75 \mu\text{m}$) (Als-Nielsen & McMorrow, 2011). Thus, in the longitudinal direction, the entire sample is firmly within the coherence length, while in the transverse direction we estimate a coherence length of at least $5 \mu\text{m} \times 5 \mu\text{m}$. All samples investigated herein are $< 1 \mu\text{m}$ in spatial extent; thus the entire sample is within a single coherence volume. This was confirmed by measuring a sample $1.3 \mu\text{m}$ in extent and confirming the appearance of the expected interference fringes. Coherence across the full sample size is necessary in order to obtain structural information at these larger meso-scales. Coherence at scales larger than the meso-clusters being investigated is not necessary to probe their internal structure; in fact—as will be described in detail below—coherence at these larger

scales is essentially detrimental as the inter-cluster interference introduces confounding scattering features.

2.3. Sample preparation

A variety of samples were fabricated using electron-beam lithography (JEOL JBX6300-FS), in order to systematically test the performance of XCCA methods. Gold features were patterned onto silicon wafers (150 μm thick). The patterning was performed by first applying a ≈ 100 nm layer of photoresist (ZEP520A), and exposure with an ebeam of accelerating voltage 100 kV at 2 nA beam current, with dose of $250 - 750 \mu\text{C}/\text{cm}^2$. The exposed resist was then developed in amyl acetate at room temperature for 90 sec and rinsed in isopropyl alcohol for 45 sec. Pattern transfer to gold was accomplished by deposition via electron beam evaporation and lift-off in *n*-methyl pyrrolidone at $\approx 80^\circ\text{C}$ for 2–4 hours. Two patterns were used: square posts (of side length ≈ 150 nm) and hexagonally-packed dots (diameter 60 nm and nearest neighbor spacing 120 nm), shown in part figure **2b** and **2c**, respectively.

Fig. 2. The samples used in the analysis pipeline: **a**: a random arrangement of oriented square posts, **b**: a sparse arrangement of hexagonal clusters, **c**: a hexagonal arrangement of gold dots with no long range order, **d**: a simulation of randomly arranged trimers. The scattering of the **a-c** are measured via synchrotron experiments whereas the scattering of **d** is computed numerically. Note that the scale bar in **d** is not relevant for simulations, since **d** is not a physical sample.

An additional set of samples was prepared using self-assembly, in order to create a material where the beam volume is filled with a distribution of randomly-oriented grains (2D powder) that nevertheless exhibit well-defined local packing motifs. Hexagonally-packed nanoparticle arrays were fabricated using a block-copolymer patterning method similar to that previously reported (Rahman *et al.*, 2015). Briefly, a cylinder-forming polystyrene-*block*-poly(methyl methacrylate) (PS-*b*-PMMA) diblock-copolymer was annealed on a ‘neutral’ (with respect to block wetting behavior) substrate. Under the selected conditions, the block-copolymer forms an array of vertically-oriented nanoscale cylinders (cylinder long-axis perpendicular to substrate plane), arranged locally into a hexagonal lattice. The PMMA cylinder cores were converted into alumina replicas using selective infiltration synthesis (Peng *et al.*, 2010), and used as a mask for etching into the underlying silicon substrate. The Si posts were decorated with Au caps using glancing-angle physical vapor deposition. This protocol gives rise

to Si posts capped with Au nanoparticles, with an average diameter of 20 ± 3 nm and an average post-post spacing of 40 ± 3 nm. This is seen in figure 2c.

2.4. Simulations

Numerical simulations of coherent x-ray scattering patterns were computed using the Python programming language, and exploiting existing libraries for efficient numerical computation (numpy; (Oliphant, 2007)) and plotting (matplotlib; (Hunter *et al.*, 2007)). Candidate meso-cluster configurations were assessed by projecting the 3D electron density into a 2D grid. The predicted scattering was obtained as the square of the Fourier transform of this projected density. Additional details regarding the simulation methods are presented in the supplemental materials. The clusters used are shown in figure 2d, where the color represents the projected density of the object. The clusters are computed in a 2000×2000 pixel computational grid with a radius of 4 pixels, and nearest neighbor separation 12 pixels.

In one particular case, 1D simulations were used. This is due to the fact there was no loss of generality for this case with the gain of a significant speedup in computation time. For this case, the curve is shown and explained later.

2.5. Correlation function

The internal structure of a single meso-cluster can be obtained by computing the angular correlation function. The benefit of this cross-correlation is that it is independent of meso-cluster orientation, which naturally accounts for the random orientation distribution expected in real experiments of self-assembled meso-clusters. This use of angular correlations was first proposed by (Kam, 1977b), and subsequently extended by many groups, including Vartanyants (Kurta *et al.*, 2012; Altarelli *et al.*, 2010), who performed an in-depth analysis on the Fourier components of angular correlation

functions using simulations and theory. The Fourier components are necessary for the reconstruction of the total diffraction pattern of a single object from the measurement of many objects (Pedrini *et al.*, 2013). However, here we instead explore the limit where such a reconstruction may not be possible. In particular, *in-situ* measurements of weakly-scattering samples will yield a low per-frame photon count. For dynamic samples (e.g. meso-clusters tumbling in solution) longer exposures are not possible as this will simply yield the isotropic SAXS pattern, where the angular correlations are averaged out. Although a single frame may have very poor statistics, by averaging the correlation functions computed across a great many frames, one can obtain a robust estimate of the angular correlations within a particular sample—and thereby a measurement of the structure within a dynamic meso-cluster. Thus, our aim in this work is to study the scaling of the signal-to-noise ratio (S/N) of angular correlation functions.

For some measured intensity $I(q, \phi)$, the angular correlation function is:

$$\mu = \frac{1}{N_m} \sum_{k=1}^{N_m} I_k(q, \phi) M(q, \phi) \quad (1)$$

$$C_{q,q}(\Delta\phi) = \frac{1}{N_m N_{\Delta\phi}} \sum_{j=1}^{N_\phi} \sum_{k=1}^{N_m} (I_k(q, \phi_j) - \mu) M(q, \phi_j) \times \\ (I_k(q, \phi_j + \Delta\phi) - \mu) M(q, \phi_j + \Delta\phi) \quad (2)$$

$$c_{q,q}(\Delta\phi) = \frac{C_{q,q}(\Delta\phi)}{C_{q,q}(\Delta\phi \approx 0)} \quad (3)$$

$$N_{\Delta\phi} = \sum_{j=1}^{N_\phi} M(q, \phi_j) M(q, \phi_j + \Delta\phi), \quad (4)$$

where C and c refer to the unnormalized and normalized versions of the correlation functions, respectively. $\Delta\phi \approx 0$ signifies the first non-zero $\Delta\phi$. Here, $M(q, \phi_j)$ refers to a binary mask, which will be discussed in later sections. For now, it may be assumed to be 1. $N_{\Delta\phi}$ refers to the number of non-masked points summed in ϕ , whereas N_ϕ refers to the number of points in ϕ on the polar grid (the $I(q, \phi)$ map in figure 1). For

the case of a unity binary mask, $N_{\Delta\phi} = N_\phi$. Background subtraction is represented by subtracting the average of the measured intensity μ . For notational simplicity, we will suppress the q dependence of the correlation functions $C_{q,q}$ and $c_{q,q}$, and write C and c , respectively. The choice of $\Delta\phi \approx 0$ being the first non-zero $\Delta\phi$ as opposed to $\Delta\phi = 0$ is chosen to suppress additional variances arising from shot noise statistics (for low count rate measurements). Note that in the case of a perfectly measured correlation curve $c(\Delta\phi)$ its average should be centered around zero, bounded by ± 1 .

2.6. Signal-to-noise metric

In order to quantify the performance of various workflows, we require a metric that assesses the strength of the correlation signal. Here, we assess the quality of a particular correlation function by direct comparison with the ‘ideal’ or ‘model’ correlation function. The model is generated using idealized 2D simulations, and thus is a representation of the correlation function expected for a perfect sample under perfect measurement conditions. For a particular measured correlation function $c_d(\Delta\phi)$, we compute a model $c_m(\Delta\phi)$, and define our correlation signal, correlation noise and signal-to-noise ratio (S/N) as:

$$\text{signal} = \sqrt{\sum_{i=1}^{N_{\Delta\phi}} (c_m(\Delta\phi_i) - \langle c_m \rangle_{\Delta\phi})^2} \quad (5)$$

$$\text{noise} = \sqrt{\sum_{i=1}^{N_{\Delta\phi}} (c_m(\Delta\phi_i) - c_d(\Delta\phi_i))^2} \quad (6)$$

$$S/N = \frac{\text{signal}}{\text{noise}} \quad (7)$$

where i corresponds to the i^{th} correlation in $\Delta\phi$ (for all $N_{\Delta\phi}$ points). The quantity of interest (S/N) is essentially the robustness of the correlation function, across ϕ . Thus, we estimate the strength of the correlation signal from the amount of structure (deviation from the average) in the model, and we estimate the correlation noise from

the mismatch between the experimental and ideal correlation functions. From here on, correlation signal and correlation noise refer to these signal and noise quantities, respectively, whereas signal and noise alone refer to the signal and noise in the traditional sense (measured intensity). The signal-to-noise will always refer to this metric defined in Eq 7. In particular, the correlation noise is defined to be the $\sqrt{\chi^2}$ error of the fit to the model. Intuitively, the more the measured correlation function deviates from the ideal one, the larger the χ^2 error, and the lower the S/N . In some cases, we may be interested in how well the overall form of the correlation function fits the model. For example, a correlation function could consistently well represent the overall six-fold symmetry of a sample, yet always be off by a constant factor. This would bias the correlation noise in our metric, which is defined as the $\sqrt{\chi^2}$ difference between the model and measured correlation function. To compensate for this, we fit the model c_m to the measured correlation functions (from either measurements or synthetic data) with free parameters the amplitude and baseline:

$$c_m(\Delta\phi) = Ac'_m(\Delta\phi) + B, \quad (8)$$

where c_m and c'_m are the fitted and original computed perfect model and where A and B are free parameters. A non-linear least squares optimization scheme is used. This allows for changes in background (bias) and amplitude to be taken into account.

3. Results and Discussion

3.1. Concept

We will discuss various problems/artifacts encountered throughout the analysis pipeline (Fig. 1), each followed by how to resolve them. Real measurements and synthetic data will be used to highlight these artifacts. The results of these real measurements and synthetic data will be scored against model data, using the metric introduced in section 2.6.

3.2. Background Noise

Problem : Noise in correlations

Background scattering is a perennial complication in x-ray scattering experiments, with contributions from detector noise, sample cell windows, air-path scattering, and the sample itself (substrate, water solution, etc.). In correlation analysis, the background represents a noise that pollutes the desired correlation signal. An isotropic background decreases the S/N of the extracted correlation function, whereas a structured background can in fact introduce artifactual correlations. To quantitatively assess the impact of background on the robustness of angular correlation functions, we use a simple 1D model to represent the scattering intensity (along ϕ) for a sample exhibiting 6-fold symmetry (red curve in Figure 3). The orientation of this model is allowed to vary randomly. The noise from the influence of finite photon fluence is simulated by sampling random numbers following a Poisson distribution:

$$P(k; \mu_s, \mu_{bg}) = \mu^k \frac{e^{-\mu}}{k!} = (\mu_s + \mu_{bg})^k \frac{e^{-(\mu_s + \mu_{bg})}}{k!} \quad (9)$$

where $\mu = \mu_s + \mu_{bg}$ is the ideal underlying average intensity versus time, and μ_s and μ_{bg} are the ideal average scattering of the sample and background versus time, respectively. Each of these ideal quantities are assumed averaged over an infinite number of measurements N_m , assuming the sample is static. The sample average scattering versus time μ_s varies with ϕ whereas the background μ_{bg} does not. The ϕ dependence will be suppressed for notational brevity. This results in a noisy curve (black curve in the upper left panel of figure 3), which is tweaked by varying μ_s and μ_{bg} . In order to better understand the correlation noise trends, the scattering intensity is described using two parameters: the overall intensity averaged over time and azimuth, $\langle \mu \rangle_\phi$, and the fraction of sample to total scattering $X = \frac{\langle \mu_s \rangle_\phi}{\langle \mu_s \rangle_\phi + \mu_{bg}}$, where $\langle A(\phi) \rangle_\phi$ refers to an average in angle ϕ $\frac{1}{2\pi} \int_0^{2\pi} A(\phi) d\phi$. 100×100 different combinations of $(\langle \mu \rangle_\phi, X)$

are simulated each for 1000 random instances of the model, and the signal-to-noise computed as outlined in Sec. 2.6 for each curve, and averaged for each $(\langle\mu\rangle_\phi, X)$ pair.

Figure 3 shows the scaling of the signal-to-noise. At low overall count rates $\langle\mu\rangle_\phi$, the S/N is found to fit to the trend $\langle\mu\rangle_\phi X^2$, as can be seen in the left and bottom panels of Fig 3. As expected, the robustness of the correlation functions increases as the total scattering intensity ($\langle\mu\rangle_\phi$) is increased, since Poisson fluctuations are diminished. One can thus experimentally improve the correlation measurement weakly ($S/N \sim \langle\mu\rangle_\phi$) by increasing photon flux or exposure time. In the event that the background dominates the sample ($\mu_{bg} \gg \langle\mu_s\rangle_\phi$, and $X \approx \frac{\langle\mu_s\rangle_\phi}{\mu_{bg}}$), a stronger scaling can be achieved by increasing the sample scattering, or decreasing the background scattering ($S/N \sim X^2$). This scaling highlights the importance of eliminating—as much as possible—all sources of background scattering from an experimental setup. We note that previous estimates of the signal-to-noise of correlation functions used a different metric from the one described here (Kirian *et al.*, 2011); nevertheless, the scaling predictions from previous work are in agreement with the work presented here (further discussion can be found in appendix A).

Fig. 3. Top: One-dimensional simulations of x-ray scattering line cuts (for a 6-fold symmetric sample) are generated (black curve depicts a noisy curve, red depicts a high-signal curve). The associated correlation function is computed (c), and averaged over multiple realizations. The average correlation function (magenta) is compared to the idealized scattering (blue) in order to compute a signal-to-noise (S/N). Bottom: Filled contour plot depicting the scaling of S/N as a function of total scattering intensity ($\langle\mu\rangle_\phi$), and background scattering. X is the ratio of sample scattering to total scattering; i.e. $X \rightarrow \infty$ represents no background scattering (high S/N), while $X \rightarrow 0$ represents background-dominated measurements (low S/N). The labels 1-3 represent the $X, \langle\mu\rangle_\phi$ values for the magenta correlation plots in the top row. The adjacent graphs highlight the precise form of the scaling at low count rates: $S/N \sim \langle\mu\rangle_\phi X^2$.

Another important consideration in real experiments is that the background scattering may not be isotropic. For instance, air scattering can be shadowed by the sample holder, giving rise to a diffuse scattering exhibiting the symmetry of the experimental apparatus. Similarly, parasitic slit scattering streaks can introduce an apparent symmetry (typically 4-fold, for a four-blade slit). More generally, *any* angular structure in the non-sample scattering (air scatter, windows, substrate, etc.) will necessarily give rise to distinct structure in the angular correlation function. Thus, a critical aspect of using XCCA experimentally is to avoid the measurement of unintended correlations arising from the experimental geometry.

We take the analysis from equation 2 and ignore the background subtraction step (set $\mu = 0$) and set the mask to unity. Assuming that the measured scattering intensity is a sum of x-rays from the sample I_s plus some background I_{bg} , the correlation function is (from Eq. 2):

$$C_{q,q}(\Delta\phi)_{\mu=0} = \frac{1}{N_m N_\phi} \sum_{j=1}^{N_\phi} \sum_{k=1}^{N_m} (I_{k,s}(q, \phi_j) + I_{k,bg}(q, \phi_j)) \times (I_{k,s}(q, \phi_j + \Delta\phi) + I_{k,bg}(q, \phi_j + \Delta\phi)) \quad (10)$$

$$= \frac{1}{N_m N_\phi} \sum_{j=1}^{N_\phi} \sum_{k=1}^{N_m} (I_s(\phi) I_s(\phi + \Delta\phi) + \quad (11)$$

$$I_s(\phi) I_{bg}(\phi + \Delta\phi) + I_s(\phi + \Delta\phi) I_{bg}(\phi) + \quad (12)$$

$$I_{bg}(\phi) I_{bg}(\phi + \Delta\phi)) \quad (13)$$

which can be grouped into four terms. The first term is the desired correlation from the sample. The last three contribute to extra covariances that corrupt the structure of the correlation functions, for small number of measurements N_m . However, we can obtain an approximation of the background by computing the ensemble average—that is, by averaging over independent realizations of the experiment, wherein the orientation of the sample (meso-clusters) varies randomly, but the background scattering is essentially constant. In the limit of averaging large numbers of orientations, the ensemble average can be thought of as a summation of the background and the isotropic average of the sample:

$$\langle I \rangle_o = \langle I_s(\phi) + I_{bg}(\phi) \rangle_o \quad (14)$$

$$= \langle \mu_s(\phi) \rangle_\phi + \mu_{bg}(\phi) \quad (15)$$

where the subscript o denotes an average over sample orientations, keeping the background fixed (whereas $\langle \rangle_\phi$ represents an azimuthal average).

Solution : Robust Subtraction

Since the sample is averaged over all possible orientations, it appears as a term constant in ϕ , and does not contribute structure to the correlation functions. Subtracting

this correlation function with this term leads to:

$$C_{q,q}(\Delta\phi)_{\mu=0} = \frac{1}{N_m N_\phi} \sum_{j=1}^{N_\phi} \sum_{k=1}^{N_m} (I_{k,s}(q, \phi_j) + I_{k,bg}(q, \phi_j) - \langle I \rangle_o) \times \\ (I_{k,s}(q, \phi_j + \Delta\phi) + I_{k,bg}(q, \phi_j + \Delta\phi) - \langle I \rangle_o) \quad (16)$$

$$= \frac{1}{N_m N_\phi} \sum_{j=1}^{N_\phi} \sum_{k=1}^{N_m} (\Delta I_{k,s}(q, \phi_j) + \Delta I_{k,bg}(q, \phi_j)) \times \\ (\Delta I_{k,s}(q, \phi_j + \Delta\phi) + \Delta I_{k,bg}(q, \phi_j + \Delta\phi)) \quad (17)$$

$$= \frac{1}{N_m N_\phi} \sum_{j=1}^{N_\phi} \sum_{k=1}^{N_m} \left(\Delta I_{k,s}(\phi) \Delta I_{k,s}(\phi + \Delta\phi) \right. \\ \left. + \Delta I_{k,s}(\phi) \Delta I_{k,bg}(\phi + \Delta\phi) + \Delta I_{k,s}(\phi + \Delta\phi) \Delta I_{k,bg}(\phi) \right. \\ \left. + \Delta I_{k,bg}(\phi) \Delta I_{k,bg}(\phi + \Delta\phi) \right) \quad (18)$$

$$\Delta I_{k,s}(\phi) = I_{k,s}(\phi_j) - \langle \mu_s \rangle_\phi \quad (19)$$

$$\Delta I_{k,bg}(\phi) = I_{k,bg}(\phi_j) - \mu_{bg}(\phi). \quad (20)$$

For a very large number of independent measurements, $\langle I_{k,s}(\phi) \rangle_{N_m} = \langle \mu_s(\phi) \rangle_\phi$ and $\langle I_{k,bg}(\phi) \rangle_{N_m} = \mu_{bg}(\phi)$, where $\langle A_k \rangle_{N_m}$ refers to an average over a very large number of measurements. Assuming that the sample and background are independent, this yields:

$$\langle \Delta I_{k,s}(\phi) \Delta I_{k,bg}(\phi + \Delta\phi) \rangle_{N_m} = \langle \Delta I_{k,s}(\phi) \rangle_{N_m} \langle \Delta I_{l,bg}(\phi + \Delta\phi) \rangle_{N_m} \quad (21)$$

$$= 0 \quad (22)$$

$$\langle \Delta I_{k,bg}(\phi) \Delta I_{k,bg}(\phi + \Delta\phi) \rangle_{N_m} = \langle \Delta I_{k,bg}(\phi) \rangle_{N_m} \langle \Delta I_{l,bg}(\phi + \Delta\phi) \rangle_{N_m} \quad (23)$$

$$= 0. \quad (24)$$

$$(25)$$

Eq. 18 is now independent of the background $I_{bg}(\phi)$.

The artifactual influence of a structured background can be addressed by subtracting the ensemble average from each individual scattering image, before performing a correlation analysis (outlined in Figure 1). This procedure directly subtracts the average measurement background, eliminating this source of artifacts. Subtracting the

ensemble average also subtracts the average (isotropic) scattering from the sample. However, the individual experimental frames will exhibit deviations from this average, which correctly encode the sample's angular correlations (Chen *et al.*, 2012). A correct background subtraction is absolutely essential to the success of x-ray angular correlation analysis.

3.3. Masking Artifacts

Problem : Spurious Correlations

X-ray scattering experiments generically recover imperfect images. Detectors may have untrustworthy ('hot' or 'bad') pixels. Some parts of the image will be blocked by the beamstop and beamstop holder, or shadowed by sample holders or flight-path tubes. Other image regions may be dominated by intense artifacts such as slit scattering. Modern photon-counting detectors typically have gaps between adjacent modules, leading to dark bands in the detector image. Thus, it is inevitable that the scattering image will contain regions that must be ignored—i.e. masked and excluded from the analysis. In this section, we consider the influence of detector image masking on the resulting correlation functions. In particular, we demonstrate that these masked regions will introduce artifacts in the angular correlation functions of, but that appropriate experimental protocol can eliminate these artifacts quite easily.

Masking is taken into account by multiplying the intensities by a binary image $M(q, \phi)$ (Eq 2). Values of 1 result in pixels that are considered in the correlation whereas values of 0 result in pixels that are not. Whenever the mask contains zero values, $N_\phi \neq N_{\Delta\phi}$, and the correlation function is modulated by a new term, which is $\Delta\phi$ dependent. This modulation will contribute to structure measured in the angular correlations. It may be decoupled from the correlation of the sample if multiple orientations are measured. Averaging independent, random measurements of the sample

is then equivalent to an average over orientation, which results in:

$$\langle C(\Delta\phi) \rangle_o = \frac{1}{N_{\Delta\phi} N_m} \sum_{i=1}^{N_\phi} \sum_{k=1}^{N_m} (I_k(\phi_i) - \mu)(I_k(\phi_i + \Delta\phi) - \mu) \times \quad (26)$$

$$M(\phi_i)M(\phi_i + \Delta\phi) \quad (27)$$

$$= \frac{1}{N_{\Delta\phi}} \sum_{i=1}^{N_\phi} M(\phi_i)M(\phi_i + \Delta\phi) \times \quad (28)$$

$$\frac{1}{2\pi} \int_0^{2\pi} (I_k(\phi_i + \phi') - \mu)(I_k(\phi_i + \phi' + \Delta\phi) - \mu) d\phi' \quad (29)$$

$$= \frac{1}{N_{\Delta\phi}} \sum_{i=1}^{N_\phi} M(\phi_i)M(\phi_i + \Delta\phi)C(\Delta\phi) \quad (30)$$

$$= C(\Delta\phi), \quad (31)$$

where an average over orientation is equivalent to an average over large number of independent frames:

$$C(\Delta\phi) = \frac{1}{2\pi} \int_0^{2\pi} (I_k(\phi_i + \phi') - \mu)(I_k(\phi_i + \phi' + \Delta\phi) - \mu) d\phi' \quad (32)$$

$$= \frac{1}{N_m N_\phi} \sum_{i=1}^{N_\phi} \sum_{k=1}^{N_m} (I_k(\phi_i + \phi') - \mu)(I_k(\phi_i + \phi' + \Delta\phi) - \mu). \quad (33)$$

Importantly, by averaging multiple different sample orientations, we recover the desired correlation function for the sample alone. We validated this approach using a simple and controlled experiment. Figure 4 shows three possible workflows for analyzing the scattering from a sample of square-shaped posts (by construction, the posts should give rise to a clean 4-fold angular correlation). The scattering is transformed into a $I(q, \phi)$ map and then correlated to obtain an angular correlation function. For simplicity, only one ring of the angular correlation function is examined. When the correlation function of a single image is computed (Fig. 4, top), it exhibits substantial deviations from the ideal model correlation function (top right, blue line). This occurs because the mask itself introduces covariance in the data, which are translated into structure in the correlation map.

Solution : Independent Sampling

Masking effects can, in principle, be minimized by decoupling the mask and sample orientation information. That is, by collecting multiple scattering frames where the sample reorients randomly, the contributions from the sample and mask (which is necessarily static) can be disentangled. For the example shown in Figure 4, the sample was systematically rotated through 10 different orientations (physically rotated about the beam axis). With multiple frames covering many sample orientations, one can compute (Fig. 4, bottom) the individual correlation functions, and average these functions. This yields a robust correlation function that matches the ideal model (bottom right). The averaging reinforces the sample correlations, while averaging-out the spurious correlations from the mask.

Fig. 4. Candidate workflows for extracting angular correlations from x-ray scattering data. Example data were collected from figure 2a. Three different methods were tried, represented in the three rows a)-c). The columns from left to right involve repartitioning, masking and background subtracting (left column) the images, to correlating (middle column). White regions represent the mask. The rightmost column compares a cross section of one row of the correlation function (magenta) with a known model (blue). The correlations are only plotted from 0 to 180° for ease of viewing. Row a) computes the correlation of just one measurement. Row b) computes the correlation from a series of 12 partitioned images (left inset). Row c) correlates the series of 12 images before averaging them. Of all these techniques, the results from row c) are superior. The dashed curve in the rightmost figure of row c) is the same procedure applied without background subtraction. It has been rescaled and shifted so that its form may be compared to the

This improvement in correlation signal is maximized by maximizing the number of independent measurements (N_m). We note that one could also imagine averaging the N_m independent measurements together prior to averaging. However, the cross terms will introduce extra covariances, corrupting the correlation signal (Fig. 4, middle row). Thus, the data pipeline must be correctly designed to eliminate masking artifacts.

The improvement in the signal-to-noise versus number of measurements may be seen in an experimental measurement (figure 5). The S/N metric for experimental data (square-shaped nano-posts) is computed using different subsets of the data. In this case, the computed correlation functions rapidly converge towards the model correlation function. Experimentally, it is straightforward to avoid masking artifacts by collecting a series of orientation-independent frames for any given sample. For a static sample, the sample can be manually rotated (along the beam axis); for dynamic samples (e.g. meso-clusters in free solution), the inherent motion of the sample will naturally introduce the required orientation distribution.

A special case that must be considered is spans where $N_{\Delta\phi} = 0$; that is, the case where for a particular $\Delta\phi$, there are no data points that can be averaged. These spans of course have undefined correlations. For example, if the mask eliminates more than half the angular space (at a particular q), then one cannot meaningfully compute the 180° correlation. Furthermore, regions where the fraction of available pixels ($\frac{N_{\Delta\phi}}{N_\phi}$) is small will also converge to the true correlation function more slowly than those with ample pixels. It is thus desirable to reduce the number of masked pixels wherever possible.

3.3.1. Average of fluctuating signals In the case of uniform backgrounds, I_{bg} , it may be tempting to ignore the background subtraction step since at first glance the contribution would be expected to be a constant offset. However, there remains a subtle

effect, which may be understood by decomposing the intensity as the sum of a constant part (averaged over azimuth) and a fluctuating part: $I(q, \phi) = \langle I(q, \phi) \rangle_\phi + \delta I_s(q, \phi)$ where $\delta I_s(q, \phi) = I_s(q, \phi) - \langle I_s(q, \phi) \rangle_\phi$ and $\langle I(q, \phi) \rangle_\phi = \langle I_s(q, \phi) \rangle_\phi + I_{bg}$ and I_s and I_{bg} refer to the sample and background intensity, respectively. Ignoring background subtraction ($\mu = 0$) equation 2 then becomes:

$$C(\Delta\phi)_{\mu=0} = \frac{1}{N_{\Delta\phi}N_m} \sum_{i=1}^{N_\phi} \sum_{k=1}^{N_m} (\langle I_k(q, \phi) \rangle_\phi + \delta I_k(q, \phi)) \quad (34)$$

$$\times (\langle I_k(q, \phi) \rangle_\phi + \delta I_k(q, \phi)) M(\phi_i) M(\phi_i + \Delta\phi) \quad (35)$$

$$= \langle I \rangle_\phi^2 + \frac{\langle I \rangle_\phi}{N_{\Delta\phi}N_m} \sum_{i=1}^{N_\phi} \sum_{k=1}^{N_m} \delta I_k(q, \phi) M(\phi_i) M(\phi_i + \Delta\phi) \quad (36)$$

$$+ \frac{\langle I \rangle_\phi}{N_{\Delta\phi}N_m} \sum_{i=1}^{N_\phi} \sum_{k=1}^{N_m} \delta I_k(q, \phi + \Delta\phi) M(\phi_i) M(\phi_i + \Delta\phi) \quad (37)$$

$$+ \frac{1}{N_{\Delta\phi}N_m} \sum_{i=1}^{N_\phi} \sum_{k=1}^{N_m} \delta I_k(q, \phi) \delta I_k(q, \phi + \Delta\phi) M(\phi_i) M(\phi_i + \Delta\phi). \quad (38)$$

Equation 38 contains four terms. The first is an additive constant and the fourth term would yield the expected background subtracted correlation function for large N_m . However, it also contains a second and third term, which are dependent on $\Delta\phi$. In the limit of large numbers of frames, these terms average to zero. However, in the case of small numbers of frames, these terms may contribute to corrupting the data. This difference is seen in our data, as can be seen in the bottom right plot of Figure 4 comparing the analysis without (dotted magenta) and with (solid magenta) background subtraction.

Fig. 5. Scaling of the S/N metric as a function of the number of independent measurements (different sample orientations) averaged together. Data were collected on a collection of randomly-positioned square-shaped nanoposts (example scattering image in upper-left, with SEM data inset). The signal-to-noise (S/N) is computed by comparing to the simulated scattering pattern for an idealized version of the sample (upper-right). As can be seen, the computed correlation functions (bottom row, magenta) rapidly converge towards the ideal correlation function (bottom row, blue). Thus, the S/N of the correlation analysis grows rapidly, saturating to an robust value after only ~ 10 measurements. The corresponding correlation function and model are plotted for some number of orientations (points 1-4), in the bottom row.

3.4. Inter-cluster Interference

Problem : Unwanted Speckle

An important consideration in the measurement of a small number of scattering objects is the inter-object interference contribution. From the point of view of resolving the internal structure of a meso-cluster, the cluster-cluster scattering contribution represents an undesired complication: this interference term gives rise to high-frequency oscillations in the scattering data that encode the relative position and orientation of the various meso-clusters within the measurement volume. The total scattering

intensity for N_c meso-clusters may be written as:

$$I(q, \phi) = \left(\sum_{j=1}^{N_c} F_j(q, \phi) e^{i\vec{q} \cdot \vec{r}_j} \right)^2 \quad (39)$$

$$= \sum_{j=1}^{N_c} |F_j(q, \phi)|^2 + \sum_{\substack{j,k=1 \\ j \neq k}}^{N_c} F_j(q, \phi) F_k^*(q, \phi) e^{i\vec{q} \cdot (\vec{r}_j - \vec{r}_k)} \quad (40)$$

The second term is due to inter-cluster interference and will not lead to a structured signal in the correlation function. This is due to the fact that the positions and orientations of the clusters are random, which means that the inter-cluster interference peaks do not appear with a consistently preserved spacing along ϕ . Although it does not contribute to the cross correlations, it does contribute to the variance (noise) of these correlations, which has the overall effect of reducing the signal-to-noise. This correspondingly increases the number of independent measurements required to suppress this effect and achieve a desired signal-to-noise.

For measurements on a dilute collection of meso-clusters, the average separation distance between clusters tends to be larger than the internal correlation lengths of the sample. These large distances introduce high-frequency oscillations in the detector image (reciprocal-space). This can be seen in Figure 6 (middle row, left), where the broad meso-cluster scattering peaks are modulated by high-frequency oscillations encoding the relative positions and orientations of the adjacent meso-clusters. In the presented example, each meso-cluster is a patch of 23 particles arranged into a hexagonal lattice (which gives rise to the broad structural peaks), with 20 such clusters within the measurement volume. The high-frequency inter-cluster interference gives rise to spurious angular correlations (Fig. 6 bottom left).

Fig. 6. Smoothing x-ray scattering data can improve the signal-to-noise of a correlation analysis. In this example, correlation functions are computed from experimental data for groups of lithographically-patterned meso-clusters. Each meso-cluster is a 5×5 array of Au nanodots arranged in a hexagonal lattice (23 nanoparticles per meso-cluster). In the presented measurement, 20 meso-clusters were within the beam cross-section, arranged randomly (positionally and orientationally) with respect to one another (SEM image in Fig. 2b). An example scattering image is shown in the upper-left, with a zoomed-in view (region denoted by blue box) shown below. The scattering shows a series of broad peaks arising from the hexagonal order within the meso-clusters. These peaks are modulated by high-frequency oscillations that arise due to inter-cluster interference. The computed correlation function (bottom left, magenta curve, computed at the q -value denoted by the blue ring) exhibits artifactual deviations from the idealized correlation function (blue). This results from spurious angular correlations introduced by the inter-cluster interference. These high-frequency artifacts can be suppressed by smoothing the x-ray scattering data, which correspondingly improves the correlation function (bottom center shows smoothing with 3 pixel kernel). Smoothing too aggressively (bottom right shows smoothing with 16 pixel kernel), corrupts the correlation function by smearing the structural peaks. The smoothing kernel size is shown by the insets, and the red circles.

Solution : Smoothing

For these dilute systems, the inter-cluster interference may be reduced by smoothing the image. The smoothing method chosen here is through convolving the pattern with a Gaussian kernel. Smoothing with an appropriate kernel size removes the undesired

high-frequency patterns, without eliminating the lower-frequency patterns that encode the meso-cluster symmetry. This leads directly to an improvement in the correlation function (Fig. 6 bottom center), with the spurious correlations eliminated. Of course, smoothing with a very large kernel will smear even the peaks associated with the sample's structure; in such cases (bottom right), the correlations are averaged out. There is thus an optimal regime of smoothing, where the high-frequency (inter-cluster interference) terms are washed out, improving the visibility of the desired structural peaks, without blurring these peaks themselves. Intuitively, this smoothing size-scale should be somewhat smaller than the width of the structural peaks from the meso-cluster. This is borne out in Figure 6 (top row, right), where an intermediate kernel size (smaller than the structural peaks) maximizes S/N .

In the small-angle limit, smoothing in detector space is equivalent to the loss of transverse coherence (Vartanyants & Robinson, 2003):

$$I_p(\vec{q}) = I_{coh}(\vec{q}) \otimes \mu_{in}(\vec{q})$$

where $\mu_{in}(\vec{q})$ is the Fourier transform of the mutual coherence function which can be modeled by a Gaussian (the Schell Model). The role of partial coherence in angular correlation functions has been investigated by (Latychevskaia *et al.*, 2015). If the mutual coherence function is written as a Gaussian in real space:

$$e^{-\frac{x^2+y^2}{2\sigma^2}}$$

then this is equivalent to a smoothing of the diffraction in reciprocal space but where the kernel in reciprocal space, $\sigma_q = \frac{1}{\sigma}$. For the example shown in Figure 6, the optimal kernel was 3 pixels, which is $2.04 \times 10^{-4} \text{ \AA}^{-1}$, corresponding to a real space length of $3.08 \text{ \mu m}$. This is ≈ 5 times the overall meso-cluster size ($\approx 600 \text{ nm}$). Thus, as a general guideline one can use a smoothing kernel of width $\sigma_q \approx 2\pi/(5r_{eff})$, where r_{eff} is the effective radius of the meso-object one is attempting to probe. The equivalence between

data smoothing and experimental transverse coherence suggests that one could equally avoid spurious inter-cluster correlations by tuning the beamline coherence length. Ideally, the beamline coherence should be set to be slightly larger than the meso-cluster size. This allows the internal structure of the meso-structure to interfere coherently, while eliminating coherent interference between different meso-clusters (i.e. the cluster scattering simply add incoherently). However, it is far easier (and provides greater experimental flexibility) to simply smooth the detector image data.

Smoothing effectively eliminates inter-cluster interference when the meso-clusters are relatively dilute; that is, where the average inter-cluster length scale is large compared to the intra-cluster correlation lengths. A more challenging case is where a substantial fraction of meso-clusters are sufficiently close that the resulting inter-cluster interference terms have frequency components comparable to the desired meso-cluster scattering peaks. In such cases, smoothing is still useful, as it at least eliminates a subset of the spurious correlations. In the next sections, we explore in greater detail the confounding effect of inter-cluster interference as a function of meso-cluster density.

3.5. Meso-cluster Density

Problem : High Meso-cluster Density

The influence of particle density on the signal-to-noise in coherent scattering has been studied extensively with (Kurta *et al.*, 2012; Altarelli *et al.*, 2010) setting up the theoretical footing, and (Latychevskaia *et al.*, 2015) demonstrating the role of partial coherence. Consistent with our results, it was found that inter-object interference corrupts the correlation signal. For a sample of randomly positioned meso-clusters, one should expect a decrease in meso-cluster separation when density is increased, and thus an overall decrease in signal-to-noise functions when increasing the number of meso-clusters within the beam cross-section.

Fig. 7. Signal-to-noise for simulations of N_c meso-clusters within the measurement volume. Each meso-cluster is a simulated trimer as in Fig. 2d. The spatial density of the meso-clusters is given by ρ ; increasing density implicitly decreases the average inter-cluster distance. Correlation functions were computed for the first-order ring of the particle-particle scattering ($q = 2\pi/d$, where d is the particle-particle spacing). The signal-to-noise is fit to a universal scaling of ($S/N \sim 1/\sqrt{\rho N_c} = \frac{\sqrt{A}}{N_c}$), arising due to the increasing influence of inter-cluster interference at high density, corrupting the correlation function by introducing spurious correlations.

This trend was investigated through synthetic data of triangular mesoclusters (see SEM in figure 2). The details of the computation of these simulations are found in the supplementary materials. Figure 7 investigates in detail the scaling of signal-to-noise function of N_c scattering entities within the measurement volume, using simulations of triangular meso-clusters (each cluster is 3 nanoparticles arranged into an equilateral triangle). We find that, for a measurement cross-section area A , the signal-to-noise scales as $S/N \sim \frac{\sqrt{A}}{N_c}$ (obtained by fitting); that is, it is advantageous to distribute the N_c meso-objects over as wide an area A as possible.

Fig. 8. Same as Fig. 7, except that the scattering was smoothed using a kernel 8.3 times the meso-cluster r_{eff} before computing the correlation functions. The signal-to-noise is fit to a universal scaling of $(S/N \sim 1/\rho = \frac{A}{N_c})$. The stronger scaling versus A compared to Fig. 7 is due to the smoothing suppressing some of the noise introduced by inter-cluster interference.

Figure 8 investigates the same simulation except in the case of smoothing using an optimal smoothing kernel (found to be ~ 8.3 times the effective radius). In this case, the signal-to-noise is now more favorable, scaling as $S/N \sim \frac{1}{\rho} = \frac{A}{N_c}$. Since the principle contribution of inter-cluster interference is to add high-frequency oscillations (encoding the inter-cluster distances), an improvement in signal-to-noise is expected with the use of smoothing. Importantly, smoothing also improves the *scaling* of S/N with meso-cluster density. The difference in scaling ($\sqrt{A}$ versus A) is intuitively related to the difference between 1D smoothing (convolution along $\Delta\phi$) and 2D smoothing (using a Gaussian kernel), respectively.

In the limit of very few meso-clusters within the beam, the signal-to-noise converges to the incoherent scattering limit (single-entity measurement). Conversely, as the areal density of meso-clusters is increased, the robustness of the correlation function decreases, following $S/N \sim 1/\rho$. This trend again highlights the corrupting effect

of inter-cluster interference. The initial smoothing has the effect of suppressing the high-frequency interference terms (large spatial separation) so that their contribution is negligible. However, clusters whose separation is comparable to the length scales within the meso-cluster will contribute inter-cluster interference terms on the same length scale, and will thus not be suppressed by smoothing. Decreasing the density decreases the probability of this proximal separation and thus allows the higher-frequency terms to be suppressed by the smoothing. However, at high meso-cluster density, the average inter-cluster separation becomes similar to the length scale of the internal meso-cluster structures. In such cases, smoothing cannot eliminate the spurious inter-cluster correlations without also eliminating the structural correlation signal of interest.

Solution : Dilute Solutions

This analysis points towards simple methods for optimizing sample preparation for subsequent $(q, \Delta\phi)$ correlation analysis. In particular, low meso-object densities naturally avoid confounding inter-object interference. Where possible, sample density should be decreased, and beam cross-section correspondingly increased (i.e. defocusing) so that the number of clusters within the beam, and the beam photon flux, are conserved (even as average inter-cluster distance is increased).

Problem Aggregation of Meso-objects

Given the clear trend of signal-to-noise with respect to meso-cluster density, we now consider the problematic case where meso-clusters exhibit a propensity to attract one another, or even aggregate. These cases of course increase the ‘local density’ in the sense that inter-cluster attraction decreases the average separation distance and thus increases the inter-cluster interference effect. We explore this phenomenon by, again, simulating triangular meso-clusters ($N_c = 10$ within the beam), and computing S/N of the correlation function based on the first-order structural peak ($q = 2\pi/d$).

Aggregation of the triangular meso-objects is simulated as follows. Triangles are inserted one by one at a random position within the simulation area. Aggregation behavior is regulated by: (1) p_a , a probability for the inserted triangle to be added near to an existing triangle (forming a group/aggregate), where it is otherwise placed randomly in the simulation volume, independent of existing objects; (2) r_a , a characteristic distance used when adding triangles to a group (near an existing triangle), which is used as the width (standard deviation) of a normal distribution (centered on the selected triangle) for selecting the position of the added object.

Figure 9 shows the behavior of the correlation signal-to-noise as a function of the aggregation probability (p_a) and aggregation length-scale (r_a). Increasing aggregation probability, and/or decreasing aggregation distance, lead to a decrease in the S/N , since this increases the local density of the meso-clusters, introducing inter-cluster interference and thus spurious correlations. For highly overlapping meso-clusters, the computed correlation function exhibits substantial deviations from the ideal function.

This result is problematic in the sense that experimental systems of meso-clusters (e.g. self-assembled hybrids of inorganic nanoparticles assembled using soft linkers) frequently exhibit spontaneous aggregation behavior, depending on local environment (solvent composition, salt concentration, temperature, etc.). Our results indicate that aggregation strongly corrupts the correlation signal from the meso-cluster internal structure. Thus, even if care is taken to prepare a concentration where the average meso-cluster distance should be large enough to avoid inter-cluster scattering artifacts, even relatively weak aggregation behavior ($p_a \approx 0.3$) can strongly corrupt the correlation function. If the spurious correlations are interpreted as arising from meso-cluster internal geometries, then the meso-cluster structure will be incorrectly assigned. On the other hand, it is clear that weaker inter-cluster attraction is less problematic.

Solution : Larger Separation Distances

If the average separation distance of the clusters remains somewhat larger than the meso-cluster ($r_a > 5r_{eff}$), the spurious correlations are negligibly small. Overall, our results highlight the importance of understanding the aggregation/attractive behavior of meso-objects before applying correlation analysis to resolve their internal structure. Wherever possible, aggregation (or even attraction) should be avoided or suppressed.

Fig. 9. The signal-to-noise (S/N) for the simulated trimers in Fig. 2d while varying the aggregation probability (p_a) and a characteristic aggregation distance (r_a , noted in the graph as a multiple of r_{eff}). The spatial arrangement of the trimers is regulated by these two parameters. A propensity to aggregate (larger p_a and/or smaller r_a) corrupts the correlation functions (magenta; ideal curves in blue), and therefore decreases S/N .

3.6. Measurement Trade-off

We have shown that both background scattering and inter-cluster interference decrease the signal-to-noise of a correlation experiment. This exposes a conflict in experiment optimization. A standard procedure for increasing S/N (e.g. in a conventional SAXS experiment) is to increase the number of scattering entities within the beam, in order to bring their scattering signal well above background. Unfortunately, as we have shown, this correspondingly increases the corrupting effect of inter-object interference. Given

this conflict, we consider here the scaling of S/N in the presence of a measurement background, as a function of the number of meso-clusters within the beam. Figure 10 shows experimental data collected for model meso-clusters (lithographically-defined arrays of Au nanoparticles packed into a local hexagonal lattice). We systematically varied the number of meso-clusters within the beam from $N = 1$ to $N = 20$. The trade-off between the two effects can be clearly seen. While increasing the number of meso-clusters improves S/N by bringing the overall correlation signal higher above background scattering, the higher number-density of meso-clusters gives rise to more inter-cluster interference, which corrupts the correlation function. Thus, there is an optimal number of meso-clusters (for a given measurement volume) which maximizes the signal-to-noise (~ 5 meso-clusters for the example shown here). Thus, the S/N of a correlation analysis is in marked contrast to conventional SAXS, where increasing the number of scattering entities leads to a straightforward improvement in the data quality.

These results were also confirmed using numerical simulations of meso-clusters (Supplemental Information figure 1, where the trends of Fig. 10 appear across a wide range of different background intensities. The optimal number of meso-clusters within the measurement volume (the N_c that maximizes S/N) appears to be that which yields a diffraction pattern (barely) visible above background noise. This density could be empirically determined by probing the scattering versus concentration, and selecting conditions where the scattering peaks from the meso-objects are at the visibility limit.

Finally, a pristine example of an efficient use of the pipeline is the sample used in Fig 1. In this case, approximately 400 grains were in the beam. Their characteristic length was approximately 1/10th of their size (the lattice spanned approximately 10x10 elements). This enforced a separation well beyond their characteristic lengths, effectively reducing the contributions from inter-cluster interference at the characteristic length

in the scattering.

Fig. 10. Signal-to-noise, in the presence of a measurement background, as a function of the number of meso-clusters within the beam cross-section. The meso-clusters were small hexagonal arrays of Au nanoparticles (SEM in Fig. 2b). The number of meso-clusters within the beam was varied systematically from $N_c = 1$ to $N_c = 20$ (all clusters were localized to a $7 \mu\text{m}$ diameter region at the beam center; clusters were positioned and oriented randomly). Data for two different q -rings are shown (blue and green); correlation functions were computed from 16 instances (separate realizations with different random positions and orientations). Two competing effects arise: increasing the number of meso-clusters increases the correlation signal relative to the background noise. However, increasing the number of clusters (within the prescribed cross-sectional area) increases the corrupting effect of inter-cluster interference. An intermediate number of meso-clusters maximizes S/N .

The optimal number of meso-clusters for a given experimental setup will depend on the size of the individual meso-clusters, the beam size, and the background. Our results provide guidance with respect to optimizing an experimental protocol to yield robust correlation function analysis. The number-density of meso-clusters should be adjusted such that the average inter-cluster distance is larger than the meso-cluster internal length scales ($> 5r_{eff}$). Aggregation of meso-clusters should be avoided. Inter-cluster attraction should also be avoided, though weak attraction can be compensated by decreasing meso-cluster concentration so as to increase the average inter-cluster

distance. A particularly useful strategy may be to selectively adsorb meso-clusters to substrate patches that are carefully spaced so as to avoid inter-cluster interference. An optimized protocol would involve ‘docking’ single meso-clusters onto substrate patches that are separated by $\sim 6r_{eff}$, and arranged randomly (to avoid a strong scattering signal from patch-patch interference). Such a protocol would maximize the number of meso-clusters within the measurement volume, while avoiding any corruption from inter-cluster effects. Thus, an optimized S/N can be engineered into the sample itself.

4. Conclusion

Angular correlation analysis promises to be an extremely useful technique to characterize meso-objects under *in-situ* conditions, as it allows the internal, local geometry of meso-cluster components to be resolved, even (in principle) from extremely noisy data arising from weakly scattering samples and short exposures (necessary to probe dynamic meso-objects). We have investigated several challenges that arise when attempting to apply correlation analysis to realistic samples of meso-clusters. We propose a new metric to quantify the signal-to-noise of a correlation analysis, by computing the mismatch between an acquired correlation function and an idealized function obtained from simulations.

Correlation analysis is found to be highly sensitive to background scattering, which strongly decreases the signal-to-noise of the result. Thus, experiments targeting correlation analysis should scrupulously eliminate sources of background (air scattering, chamber windows, etc.). Of particular importance is to avoid a structured background (e.g. diffuse scattering shadowed by beamline components), as this will introduce unintended angular correlations into the data. The influence of the experimental background can be greatly minimized simply by subtracting the ensemble scattering (average of multiple measurements) from the data.

The mask applied to the experimental data is found to introduce undesired and unphysical correlations to the analysis. These artifacts can be suppressed by repeated sample measurements, if those measurements are orientation independent.

A crucial confounding aspect of correlation measurements is inter-object interference. The coherent interference between different meso-clusters leads to high-frequency oscillations in the data, modulating the structural peaks from the meso-cluster internal structure. These additional oscillations give rise to spurious angular correlations, thereby corrupting the correlation function. We demonstrate that smoothing of the data can eliminate these effects, by suppressing the high-frequency oscillations. However, when samples densities are such that inter-cluster distances approach the intra-cluster structural length scales, smoothing is no longer sufficient to suppress spurious inter-object effects. Our work suggests that experimental protocols should avoid high meso-object densities, and especially avoid aggregation of the meso-objects.

Combining our results allowed the design of an optimized workflow to yield robust correlation analysis of scientifically-interesting samples. For instance, in Figure 1, we show data for a dense two-dimensional array of hexagonally-packed nanoparticles, derived from a self-assembled block-copolymer template (Au nanoparticles 20 nm diameter, ~ 40 nm spacing). The high density of hexagonal grains within the beam gives rise to a seemingly isotropic scattering ring (approximately 400 domains within the $10\,\mu\text{m} \times 10\,\mu\text{m}$ beam cross-section). Despite this, correlation analysis is able to recover the physically-real 6-fold symmetry of the sample, which arises from the local particle packing. Conversely, we also demonstrate that our workflow can measure in the limit of very few meso-objects. Figure 10 demonstrates the ability to experimentally measure a countable number (1–10) of meso-clusters, extracting the local symmetry of particle arrangement. We have provided several experimental examples where the individual scattering frames are so weak and noisy that the sample structure cannot

be identified (Fig. 5, 6, & 10), and yet where appropriately-computed correlation functions unambiguously identify the local particle packing motifs.

With improved understanding of the experimental limitations of angular correlation analysis, a host of challenging *in-situ* experiments on dynamic finite-size meso-objects now become viable.

Appendix A

Comparison of signal-to-noise metric with previous literature

The signal-to-noise calculated in section 3.2 calculated using the new metric yields similar results to that calculated by (Kirian *et al.*, 2011). In the limit of low count rates $\langle \mu \rangle_\phi$, Kirian calculates (Eq (57), assuming $\vec{q}_1 = \vec{q}_2 = (q, \phi)$ and that $\langle \rangle_\omega = \langle \rangle_\phi$):

$$S/N_{\text{Kirian}} \approx \sqrt{M \frac{N_s^2 \langle \delta \bar{n}^s(\phi) \delta \bar{n}^s(\phi + \Delta\phi) \rangle_\phi^2}{N_{bg}^2 \langle \bar{n}^{bg}(\phi) \rangle_\phi^2 + N_s^2 \langle \bar{n}^s(\phi) \rangle_\phi^2}} \quad (41)$$

where some notation was changed to avoid confusion (s refers to sample and bg to background solvent molecules). Here, M is the number of measurements, N_s the number of clusters, N_{bg} the number of solvent molecules, $N_s \bar{n}^s(\phi)$ and $N_s \delta \bar{n}^s(\phi)$ the underlying intensity and background subtracted underlying intensity from the sample, $N_{bg} \bar{n}^{bg}(\phi)$, the underlying intensity of the solvent. We assume that our background is isotropic (no ϕ dependence), and make the following substitutions:

$$N_s^2 \langle \delta \bar{n}^s(\phi) \delta \bar{n}^s(\phi + \Delta\phi) \rangle_\phi \rightarrow I_s^2 c(\Delta\phi) \quad (42)$$

$$N_{bg} \langle \bar{n}^{bg}(\phi) \rangle_\phi \rightarrow \mu_{bg} \quad (43)$$

$$N_s \langle \bar{n}^s(\phi) \rangle_\phi \rightarrow \langle \mu_s \rangle_\phi, \quad (44)$$

we obtain:

$$S/N_{\text{Kirian}} \approx \sqrt{M \frac{\frac{1}{N_s^2} I_s^4 c^2(\Delta\phi)}{\mu_{bg}^2 + \langle \mu_s \rangle_\phi^2}} \quad (45)$$

$$= \sqrt{M \frac{\langle \mu \rangle_\phi^2 X^4}{N_s^2} c^2 (\Delta \phi)} \quad (46)$$

$$= \sqrt{M} \frac{1}{N_s} \langle \mu \rangle_\phi X^2. \quad (47)$$

Assuming the number of clusters is $N_s = 1$ and assuming one measurement $M = 1$, this agrees with results with our signal-to-noise metric.

Acknowledgments

This research used resources of the Center for Functional Nanomaterials, and the National Synchrotron Light Source II, which are U.S. DOE Office of Science Facilities, operated at Brookhaven National Laboratory under Contract No. DE-SC0012704.

References

- Als-Nielsen, J. & McMorrow, D. (2011). *Elements of modern X-ray physics*. John Wiley & Sons.
- Altarelli, M., Kurta, R. & Vartanyants, I. (2010). *Phys. Rev. B*, **82**(10), 104207.
- Chen, G., Modestino, M. A., Poon, B. K., Schirotzek, A., Marchesini, S., Segalman, R. A., Hexemer, A. & Zwart, P. H. (2012). *J. Synchrotron Radiat.* **19**(5), 695–700.
- Chubar, O., Fluerasu, A., Berman, L., Kaznatcheev, K. & Wiegart, L. (2013). In *Journal of Physics: Conference Series*, vol. 425, p. 162001. IOP Publishing.
- Donatelli, J. J., Zwart, P. H. & Sethian, J. A. (2015). *Proc. Natl. Acad. Sci.* **112**(33), 10286–10291.
- Elser, V. (2011). *New J. Phys.* **13**(12), 123014.
- Fienup, J. R. (1982). *Appl. Opt.* **21**(15), 2758–2769.
- Fluerasu, A., Chubar, O., Kaznatcheev, K., Baltser, J., Wiegart, L., Evans-Lutterodt, K., Carlucci-Dayton, M. & Berman, L. (2011). In *Proc. SPIE*, p. 81410J. International Society for Optics and Photonics.
- Förster, S., Timmann, A., Konrad, M., Schellbach, C., Meyer, A., Funari, S., Mulvaney, P. & Knott, R. (2005). *J. Phys. Chem. B*, **109**(4), 1347–1360.
- Hunter, J. D. *et al.* (2007). *Comput. Sci. Eng.* **9**(3), 90–95.
- Kam, Z. (1977a). *Macromolecules*, **10**(5), 927–934.
- Kam, Z. (1977b). *Macromolecules*, **10**(5), 927–934.
- URL:** <http://pubs.acs.org/doi/abs/10.1021/ma60059a009>
- Kirian, R. A., Schmidt, K. E., Wang, X., Doak, R. B. & Spence, J. C. (2011). *Phys. Rev. E*, **84**(1), 011921.
- Kurta, R., Altarelli, M., Weckert, E. & Vartanyants, I. (2012). *Phys. Rev. B*, **85**(18), 184204.
- Latychevskaia, T., Mancini, G. F. & Carbone, F. (2015). *Sci. Rep.* **5**.
- Lhermitte, J., (2017). Angular Correlations. [Online; accessed 02-Feb-2017].
- URL:** <https://github.com/ordirules/rdeltaphicorr>
- Lo, P. K., Karam, P., Aldaye, F. A., McLaughlin, C. K., Hamblin, G. D., Cosa, G. & Sleiman, H. F. (2010). *Nat. Chem.* **2**(4), 319–328.
- Mancini, G. F., Latychevskaia, T., Pennacchio, F., Reguera, J., Stellacci, F. & Carbone, F. (2016). *Nano Lett.* **16**(4), 2705–2713.

- Maye, M. M., Kumara, M. T., Nykypanchuk, D., Sherman, W. B. & Gang, O. (2010). *Nat. Nanotechnol.* **5**(2), 116–120.
- Mirkin, C. A., Letsinger, R. L., Mucic, R. C., Storhoff, J. J. *et al.* (1996). *Nature*, **382**(6592), 607–609.
- Murray, C. B., Kagan, C. & Bawendi, M. (2000). *Annu. Rev. Mater. Sci.* **30**(1), 545–610.
- Nykypanchuk, D., Maye, M. M., van der Lelie, D. & Gang, O. (2008). *Nature*, **451**(7178), 549–552.
- Oliphant, T. E. (2007). *Comput. Sci. Eng.* **9**(3), 10–20.
- Pal, S., Zhang, Y., Kumar, S. K. & Gang, O. (2015). *J. Am. Chem. Soc.* **137**(12), 4030–4033.
- Park, S. Y., Lytton-Jean, A. K., Lee, B., Weigand, S., Schatz, G. C. & Mirkin, C. A. (2008). *Nature*, **451**(7178), 553–556.
- Pedrin, B., Menzel, A., Guizar-Sicairos, M., Guzenko, V., Gorelick, S., David, C., Patterson, B. & Abela, R. (2013). *Nat. Commun.* **4**, 1647.
- Peng, Q., Tseng, Y.-C., Darling, S. B. & Elam, J. W. (2010). *Adv. Mater.* **22**(45), 5129–5133.
- Putkunz, C. T., Clark, J. N., Vine, D. J., Williams, G. J., Pfeifer, M. A., Balaur, E., McNulty, I., Nugent, K. A. & Peele, A. G. (2011). *Phys. Rev. Lett.* **106**(1), 013903.
- Rahman, A., Ashraf, A., Xin, H., Tong, X., Sutter, P., Eisaman, M. D. & Black, C. T. (2015). *Nat. Commun.* **6**.
- Redl, F. X., Cho, K.-S., Murray, C. B. & O’Brien, S. (2003). *Nature*, **423**(6943), 968–971.
- Saldin, D., Poon, H., Bogan, M., Marchesini, S., Shapiro, D., Kirian, R., Weierstall, U. & Spence, J. (2011a). *Phys. Rev. Lett.* **106**(11), 115501.
- Saldin, D., Poon, H., Shneerson, V., Howells, M., Chapman, H., Kirian, R., Schmidt, K. & Spence, J. (2010a). *Phys. Rev. B*, **81**(17), 174105.
- Saldin, D., Poon, H.-C., Schwander, P., Uddin, M. & Schmidt, M. (2011b). *Opt. Express*, **19**(18), 17318–17335.
- Saldin, D., Shneerson, V., Howells, M. R., Marchesini, S., Chapman, H. N., Bogan, M., Shapiro, D., Kirian, R., Weierstall, U., Schmidt, K. *et al.* (2010b). *New J. Phys.* **12**(3), 035014.
- Saldin, D., Shneerson, V., Starodub, D. & Spence, J. (2010c). *Acta Crystallogr. Sect. A*, **66**(1), 32–37.
- Senesi, A. J. & Lee, B. (2015). *J. Appl. Cryst.* **48**(4), 1172–1182.
- Sharma, J., Chhabra, R., Cheng, A., Brownell, J., Liu, Y. & Yan, H. (2009). *Science*, **323**(5910), 112–116.
- Starodub, D., Aquila, A., Bajt, S., Barthelmess, M., Barty, A., Bostedt, C., Bozek, J. D., Coppola, N., Doak, R. B., Epp, S. W. *et al.* (2012). *Nat. Commun.* **3**, 1276.
- Sun, T., Jiang, Z., Strzalka, J., Ocola, L. & Wang, J. (2012). *Nat. Photon.* **6**(9), 586–590.
- Sutton, M. (2008). *C. R. Phys.* **9**(5), 657–667.
- Takahashi, Y., Suzuki, A., Zetsu, N., Oroguchi, T., Takayama, Y., Sekiguchi, Y., Kobayashi, A., Yamamoto, M. & Nakasako, M. (2013). *Nano Lett.* **13**(12), 6028–6032.
- Tan, L. H., Xing, H. & Lu, Y. (2014). *Acc. Chem. Res.* **47**(6), 1881–1890.
- Tian, Y., Wang, T., Liu, W., Xin, H. L., Li, H., Ke, Y., Shih, W. M. & Gang, O. (2015). *Nat. Nanotechnol.* **10**(7), 637–644.
- Vartanyants, I. & Robinson, I. (2001). *J. Phys. Condens. Matter*, **13**(47), 10593.
- Vartanyants, I. & Robinson, I. (2003). *J. Synchrotron Radiat.* **10**(6), 409–415.
- Williams, G., Quiney, H., Dhal, B., Tran, C., Nugent, K. A., Peele, A., Paterson, D. & De Jonge, M. (2006). *Phys. Rev. Lett.* **97**(2), 025506.
- Yager, K. G., Zhang, Y., Lu, F. & Gang, O. (2014). *J. Appl. Cryst.* **47**(1), 118–129.
- Ye, X., Zhu, C., Ercius, P., Raja, S. N., He, B., Jones, M. R., Hauwiller, M. R., Liu, Y., Xu, T. & Alivisatos, A. P. (2015). *Nat. Commun.* **6**.
- Zaluzhnyy, I., Kurta, R., Sulyanova, E., Gorobtsov, O., Shabalin, A., Zozulya, A., Menushenkov, A., Sprung, M., Ostrovskii, B. & Vartanyants, I. (2015). *Phys. Rev. E*, **91**(4), 042506.

- Zhang, C., Li, X., Tian, C., Yu, G., Li, Y., Jiang, W. & Mao, C. (2014). *ACS Nano*, **8**(2), 1130–1135.
- Zhang, Y., Lu, F., Yager, K. G., van der Lelie, D. & Gang, O. (2013). *Nat. Nanotechnol.* **8**(11), 865–872.
- Zhang, Y., Pal, S., Srinivasan, B., Vo, T., Kumar, S. & Gang, O. (2015). *Nat. Mater.* **14**(8), 840–847.

Synopsis

The theoretical limitations and experimental challenges of x-ray angular correlation analysis methods are investigated. Using a combination of simulated data, and measurements on challenging meso-cluster samples—prepared using lithography and self-assembly—the primary experimental challenges are explored. Background scattering, data masking, and inter-object interference are identified as primary challenges. Strategies to overcome correlation artifacts are presented.