scientific reports

OPEN

An encryption–decryption framework to validating single-particle imaging

Zhou Shen[1,2], Colin Zhi Wei Teo[1,2], Kartik Ayyer[3,4] & N. Duane Loh[1,2,5✉]

We propose an encryption–decryption framework for validating diffraction intensity volumes reconstructed using single-particle imaging (SPI) with X-ray free-electron lasers (XFELs) when the ground truth volume is absent. This conceptual framework exploits each reconstructed volumes' ability to decipher latent variables (e.g. orientations) of unseen *sentinel* diffraction patterns. Using this framework, we quantify novel measures of orientation *disconcurrence*, inconsistency, and disagreement between the decryptions by two independently reconstructed volumes. We also study how these measures can be used to define data sufficiency and its relation to spatial resolution, and the practical consequences of focusing XFEL pulses to smaller foci. This conceptual framework overcomes critical ambiguities in using Fourier Shell Correlation (FSC) as a validation measure for SPI. Finally, we show how this encryption-decryption framework naturally leads to an information-theoretic reformulation of the resolving power of XFEL-SPI, which we hope will lead to principled frameworks for experiment and instrument design.

X-ray free-electron lasers (XFELs) are a promising tool for studying the three-dimensional (3D) structures of macromolecular assemblies[1,2]. The short and intense XFEL pulses make it possible to collect diffraction patterns of a macromolecule before the XFEL-damaged atomic nuclear motions become substantial[3–7].

XFEL pulses are sufficiently intense and coherent for single-particle imaging (SPI), where a single macromolecule can scatter enough photons for us to infer its 3D orientation, hence structure[8–11]. XFEL-SPI makes the difficult task of growing large, well-diffracting macromolecular crystals (even micrometer size ones[12]) unnecessary.

Instead, desiccated samples are randomly injected at unknown orientations into a regular train of XFEL pulses. To understand how orientations are defined in SPI, consider what happens when a scatterer, whose 3D diffraction volume is denoted W, is presented to the SPI laboratory reference frame (Fig. 1).

Collected diffraction patterns are identified and analyzed in various ways including: determining the 3D structures that most likely produced the ensemble of SPI patterns[13], or studying the range of 3D morphologies spanned by the XFEL scatterers[14–16].

Reconstructing a set of 3D structure from many SPI patterns comprises three sequential stages, each of which can be considered for validation[6]. These stages are: recovering a set of 3D diffraction intensities W from many two-dimensional (2D) SPI patterns; using phase-retrieval to reconstruct the 3D realspace scattering density from W; fitting atomic coordinates to the scattering density. Separate validation routines between these stages can help diagnose where resolution loss might have occurred.

This work focuses on validating the first stage, where we reconstruct W by inferring the latent 3D orientations of SPI diffraction patterns. This inference is challenging for small macromolecules that produce weak diffraction patterns. In these cases, the Fourier Shell Correlation (FSC)[17], which is typically used to validate 3D structures recovered using cryo-electron microscopy, has become increasingly popular for estimating spatial resolution[13,16,18–29].

However, the use of FSC, as well as other proposed measures of reconstruction errors[6,30], to characterize XFEL-SPI resolution suffers three main issues. First, and most importantly, Fig. 2 illustrates how the resolution reported using the popular half-bit FSC criterion actually improves with increased orientation blurring. This occurs because XFEL-SPI reconstructions approach the same *virtual powder average* as their input patterns

[1]Centre for Bio-imaging Sciences, National University of Singapore, 14 Science Drive 4, 117557 Singapore, Singapore. [2]Department of Physics, National University of Singapore, 2 Science Drive 3, 117551 Singapore, Singapore. [3]Max Planck Institute for the Structure and Dynamics of Matter, Luruper Chaussee 149, 22761 Hamburg, Germany. [4]Center for Free-Electron Laser Science, Luruper Chaussee 149, 22761 Hamburg, Germany. [5]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117557 Singapore, Singapore. ✉email: duaneloh@nus.edu.sg
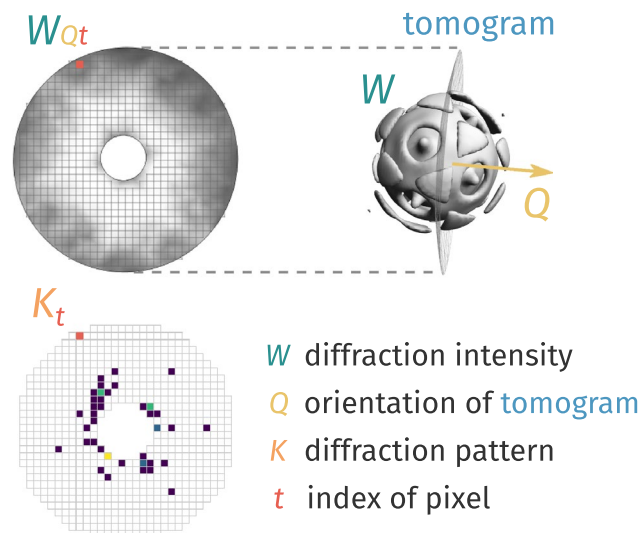
Figure 1. Schematic of how orientations are encoded in XFEL-SPI. A diffraction pattern collected on a detector (K_t where t labels the pixels on the detector) of a scatterer is an Ewald tomogram W_{Qt} through the 3D diffraction volume W. When this scatterer suffers an active random 3D rotation Ω about its own original reference frame, it is equivalent to a passive rotation of said Ewald tomogram in the opposite sense (i.e. Ω^{-1}). Throughout the rest of the paper, we parametrize this rotation with unit quaternions $Q \equiv \Omega(Q)$ (primer on unit quaternions in Supplementary Appendix).

become more misoriented. Consequently the 'noise terms' between two independently reconstructed volumes (see Eq. (3) in[31]) become correlated. Hence the FSC measure, which is invariant to isotropic filtering, can paradoxically report better resolutions when the orientation uncertainty of patterns increases. Second, the threshold criterion for determining resolution is controversial even in the cryo-electron microscopy community[31,32]. This criterion is demonstrably dependent on the speckle sampling ratio (i.e. size of realspace support), the symmetry of the particle, and assumes additive noise[31]. Unfortunately, there are still prominent violations of these criteria[33]. Third, to compute the FSC between two 3D volumes, their relative orientations must be accurately determined.

To circumvent some of these issues with FSC, we propose examining the source of correlations between two independently reconstructed volumes: the 'disconcurrence', inconsistency, and agreement between how these volumes orient individual patterns. A similar orientation-based approach to validation was explored by Tegze and Bortel[34], where they proposed using the fraction of patterns that are well-oriented to validate intensity reconstructions. However, the so called C-factor that they proposed for validation only considered orientation precision but not accuracy nor reproducibility.

It can be useful to recast the XFEL-SPI validation problem in information theoretic terms. Indeed, information theory has been insightful for SPI[35] as well as coherent diffraction imaging[36,37]. In fact, the half-bit criterion for FSC in cryo-electron microscopy[31] established a connection between spatial resolution and information theory. There, however, the half-bit criterion merely referred to when the signal-to-noise ratio of an idealized noisy channel attained a value of $\sqrt{2} - 1$. What this signal-to-noise ratio means for resolving spatial features within an object remains unclear.

Looking farther back, Shannon's original proof of the noisy channel theorem was based on a straightforward encoding–decoding scheme[38]. Below we show how Shannon's scheme can be explicitly constructed for the orientation determination problem in SPI. Doing so, allows us to validate reconstructions using an orientation resolution that can be directly related to the mutual information of the SPI experiment.

An SPI reconstruction is similar to probabilistic symmetric-key cryptography, where *plaintext messages* are encrypted into *ciphertexts* using a *correct key* plus a randomness scheme. Because of this randomness, the same plaintext message can produce different ciphertexts.

The analogous messages in an XFEL-SPI experiment are the hidden orientations of illuminated single particles[39]. The experimental setup itself can be viewed as a cipher algorithm that encrypts these messages as noisy two-dimensional (2D) diffraction patterns. When these orientations (messages) are properly decrypted, the full three-dimensional (3D) diffraction volume of the target particle can be recovered.

The conundrum for SPI, however, is that these orientations are best decrypted using the ground truth 3D diffraction volume. Hence, reconstructing this diffraction volume can be viewed as 'cracking' (i.e. guessing) the correct symmetric key in probabilistic cryptography. Figure 3 shows the similarities between SPI-validation and key-cracking in cryptography, which has the following correspondence:

- *correct key* ↔ ground truth 3D diffraction intensities;
- *encryption cipher* ↔ SPI experiment;
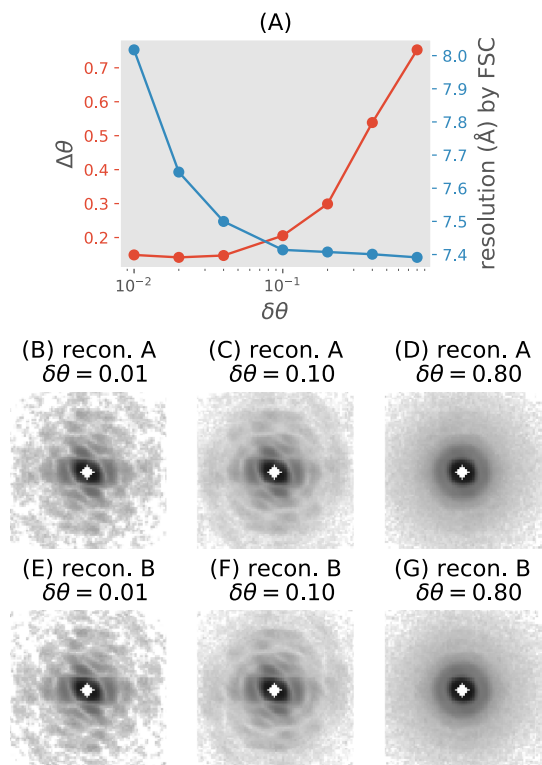- *decryption cipher* ↔ orientation inference scheme;

Figure 2. Fourier shell correlation (FSC) reports improved resolution despite increased orientational blurring. Two disjoint SPI datasets were simulated, *A* and *B*, each with 5000 patterns. (**A**) The FSC was calculated for all pairs of reconstructions from the same dataset and with the same orientation blurring $\delta\theta$ (blue curve). Diffraction volumes were reconstructed from each dataset by interpolating each pattern back into ten random orientations near the true one. The true variance of these orientations is denoted $\delta\theta^2$, which is proportional to the degree of deliberate orientation blurring. The orientation disconcurrence proposed in this paper, $\Delta\theta$ (red curve), was computed using a third smaller sentinel dataset (1000 patterns) not used in the reconstructions. For each dataset, seven 3D volumes were reconstructed by interpolating all patterns back into the 3D diffraction volume with $\delta\theta = \{0.01, 0.02, 0.04, 0.1, 0.2, 0.4, 0.8\}$. (**B–D**) The central slices of one of the seven volumes for each $\delta\theta$ from dataset *A*, (**E–G**) and those from dataset *B*.
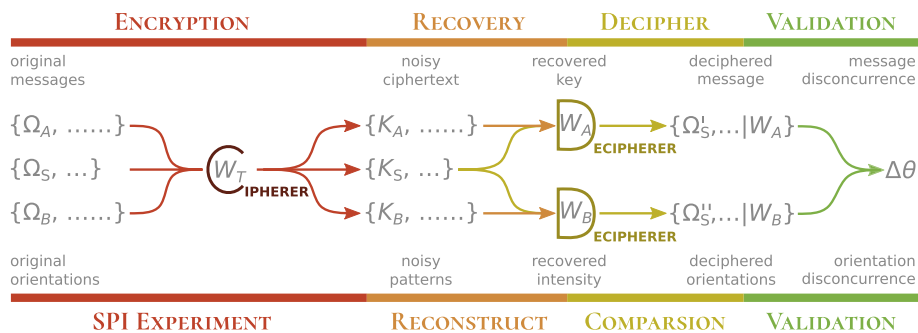

Figure 3. Analogy between 'key-cracking' in cryptography (text in upper rows) and validation for single particle imaging (text in lower rows).

- *ciphertexts* ↔ photon patterns collected in experiment;
- *messages* ↔ orientations of individual photon patterns.

Algorithms that discover the orientations of SPI patterns[8,10,40,41], analogously, try to recover the unknown key (i.e. 3D diffraction intensities) given many ciphertexts (i.e. photon patterns).

Now let us consider how one can check/validate the accuracy/correctness of a recovered key, absent the ground truth. An obvious method is to determine whether the recovered key is consistent with known prior

constraints or independent measurements. Such external validations, however, are not always possible in SPI especially when resolving novel structural forms.

We know that a correct key must decipher each ciphertext into a unique message. However, this *uniqueness alone is insufficient to determine correctness*, since wrong keys given to a deterministic cipher can yield unique but wrong decipherments. An example of this occurs when a recovered key overfits to a set of ciphertexts. Nevertheless, we can exploit this uniqueness requirement to design a scheme that detects if at least one of two candidate keys is incorrect.

Suppose we are given two disjoint sets of ciphertexts ($\{K_A\}, \{K_B\}$) that are encrypted by the same solution key W_T. We can independently recover two keys (W_A, W_B), one from each set of ciphertexts. Disagreements between how these two keys decipher a third hidden set of ciphertexts $\{K_S\}$ betrays the incorrectness of at least one of these two keys. If the first two sets of ciphertexts are sufficiently large and randomly chosen then both candidate keys are likely incorrect.

Owing to the randomness in probabilistic encryption, it is practically impossible to guarantee a perfectly accurate key given only a finite number of noisy ciphertexts. Analogously, we cannot perfectly recover the ground truth SPI diffraction volume only from a finite number of noisy, incomplete photon patterns. Consequently, any pair of recovered keys must differ measurably from each other. This difference quantifies the *decryption precision* of these keys, which is the lower bound of their *decryption accuracies*.

Back to the SPI data analysis, we wish to find the difference in how two independently reconstructed volumes W_A and W_B decrypt the orientations of a third disjoint set of *sentinel photon patterns*, $\{K_S\}$. This difference in decryption increases if the disagreement between W_A and W_B increases. More importantly, it also increases as either volume departs farther from the hidden ground truth volume W_T. We refer to this difference as the *orientation disconcurrence* between these two volumes.

To define this framework in Fig. 3 requires well-defined encryption and decryption procedures. In an XFEL-SPI experiment, this encryption is described by how an illuminated scatterer at a certain orientation generates a noisy photon pattern (Fig. 1). In a Bayesian framework, the probability that a scatterer's specific orientation (Q) is encrypted as a particular photon pattern (K) is termed the *data likelihood*. Inversely, the probability that a pattern K will be decrypted as a particular orientation Q is its equivalent orientation posterior distribution (OPD).

This encryption of orientation information into a photon pattern is governed by the physics of photon–particle interaction, wavefront propagation, and photon measurement on the detector. Under ideal XFEL-SPI experimental conditions the photon pattern K_t is a Poisson sample from an Ewald tomogram, W_{Qt}, of a particle at orientation Q (Fig. 1). This idealization allows an explicit formulation of the likelihood (see Eq. (10)), and hence OPD. Additionally, one might consider factors such as extraneous photon scattering sources, non-linear detector artefacts, and the local fluence of the XFEL pulses each particle randomly encounters. Such non-Poissonian OPDs were shown to be effective in different XFEL-SPI experiments[13,19,39]. More generally, there is an infinite number of alternatives to the Poissonian OPD that could be used to decrypt particle orientation from photon patterns. Exploring the efficacy of these myriad alternatives is clearly beyond the scope of this paper.

The encryption–decryption framework that validates two intensity reconstructions (W_A, W_B) in Fig. 3 is indifferent to the algorithms that were used to reconstruct W_A and W_B. And while the Poissonian OPD chosen in this paper was also used in the original EMC algorithm to infer the orientations of photon patterns[8], here this OPD is used to decrypt orientations for validating 3D intensity volumes W_A, W_B, which could be reconstructed with algorithms other than EMC. Since our validation occurs after W_A and W_B are separately reconstructed, it does not add any computational overhead during their reconstructions.

The OPD that most accurately describes the experiment should be used both to reconstruct and validate reconstructions. Hence it is unsurprising that the OPD used in both situations are identical.

Finally, since the validation framework in Fig. 3 compares the ability of two volumes W_A and W_B to decrypt orientations, we are essentially comparing their OPDs from decrypting the orientations of a set of sentinel patterns. To compare these OPDs, we evaluate their convolutions in orientation space to produce what we call *angular displacement distributions (ADD)*. The orientation disconcurrence between W_A and W_B are then extracted from this ADD. The procedure to compute the orientation disconcurrence given W_A and W_B is outlined below.

1. Partition the XFEL-SPI photon patterns $\{K\}$ into three disjoint sets: two larger and equally sized sets, $\{K_A\}$ and $\{K_B\}$, for reconstructions; and a third, smaller set of unseen *sentinel patterns* $\{K_S\}$ to measure orientation disconcurrence.

2. Using any algorithm you desire, reconstruct two 3D intensities from the two larger sets of patterns: $\{K_A\} \rightarrow W_A$, and $\{K_B\} \rightarrow W_B$.

3. For each sentinel pattern K_S, compute the OPD of the reconstructed volumes W_A and W_B. This is the probability that K_S corresponds to the Ewald sphere section of orientation Ω in each reconstructed volume (i.e. $P(\Omega_A|K_S, W_A)$ and $P(\Omega_B|K_S, W_B)$). This step creates $2|\{K_S\}|$ distributions, two for each sentinel pattern, where $|\{K_S\}|$ is the number of sentinel patterns used.

4. Next, we compute the angular displacement distribution (ADD, defined in Eq. (13)) of the sentinel patterns from the OPD of W_A and W_B. The ADD for each sentinel pattern K_S (the red or blue distribution in Fig. 4) is essentially a convolution of OPD$_A$ and OPD$_B$ over the space of relative orientations between W_A and W_B. If OPD$_A$ and OPD$_B$ were delta functions, then this convolution peaks at the relative orientation between W_A and W_B. The ADD$_{AB}$ (the grey distribution in Fig. 4), which is the normalized sum of these convolutions for all sentinel patterns (Eq. (14)), is the distribution of relative orientations between W_A and W_B as 'measured by' $\{K_S\}$.

5. Finally, from the ADD of all the sentinel patterns between the volumes W_A and W_B, estimate their orientation disconcurrence.
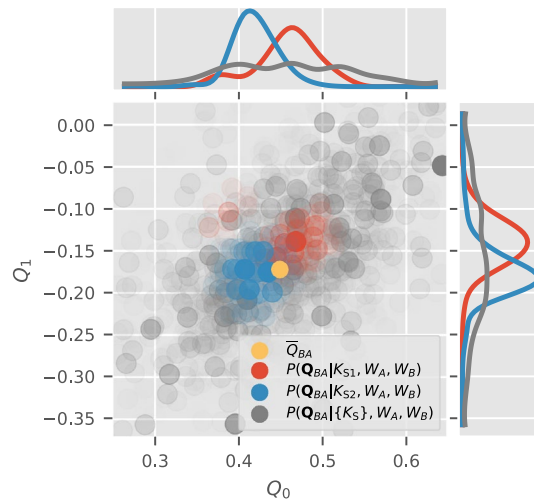
Figure 4. Clustering of the angular displacement distribution (ADD) for 1000 sentinel patterns given two independently reconstructed volumes W_A and W_B, in the space of possible unit quaternions. Only the first two components of these quaternions (Q_0, Q_1) are shown. The disks represent the set of most significant relative quaternions given each sentinel pattern, $\{Q_{BA} \mid K_S\}$, as defined by all possible pairs of those in Eq. (12). The opacities of these disks are proportional to the value of the ADD at these quaternions. The blue and red disks represent the ADDs for two specific sentinel patterns respectively. The yellow disk shows the average overall rotation $\overline{Q}_{BA}$ as defined in Eq. (16).
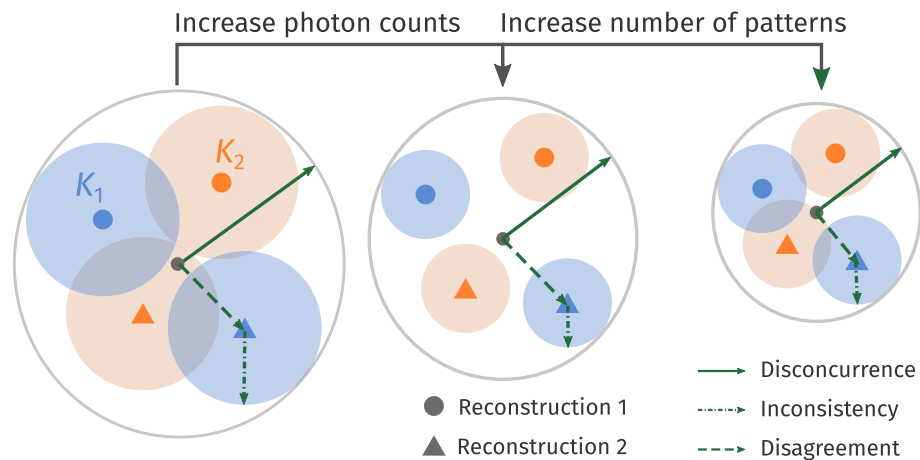

Figure 5. The orientation disconcurrence for two sentinel patterns (K_1 in blue, and K_2 in orange) consists of two parts: the inconsistency that each model orients sentinel patterns (disk spanned by dashed-dotted radii), and the disagreement between how different models orient these patterns (disk spanned by dashed radii). These aspects are affected by the photon counts per pattern (N) and the number of patterns (M_{data}) respectively.

Results

Measures of orientation uncertainties. The orientation *disconcurrence* between two independently reconstructed volumes comprises two aspects: *inconsistency* and *disagreement*. By the cryptographic analogy, the first aspect characterizes how *consistently* each volume separately decrypts the orientations of sentinel patterns; the second aspect describes how often the decryptions of two (or more) volumes *mutually agree*. These concepts are illustrated in Fig. 5, and defined below.

In the following numerical simulations, we use the disconcurrence between independent reconstructions from the same scatterer to estimate the lower bound of their correctness. Recall that this procedure requires partitioning a set of photon patterns into three disjoint sets ($\{K_A\}, \{K_B\}, \{K_S\}$). We reconstruct two 3D intensities from the first two sets (W_A and W_B respectively), while the last sentinel set is reserved for validation. Unlike an actual experiment, the true solution intensities W_T that generated these patterns are known in these simulations, and will provide useful insights. Given these definitions, let us consider different orientation measures at the end of the procedure outlined at the end introduction section.

1. Measure of orientation disconcurrence: $\Delta\theta_c(W_A, W_B)$ (Eq. (17)) is computed from the width of the angular displacement distribution (ADD) between intensities W_A and W_B that are independently reconstructed from two disjoint sets of patterns. $\Delta\theta_c$ measures the difference between the orientations of specific sentinel patterns within W_A and W_B, despite having aligned the centroids of these two distributions (i.e. overall orientations of W_A and W_B).

2. Measure of average orientation inconsistency:

$$\Delta\theta_i(W_A, W_B) = \sqrt{\frac{1}{2} \sum_{i \in \{A,B\}} \Delta\theta_c^2(W_i, W_i)} \, . \tag{1}$$

This is the root-mean-squared (RMS) angular width of the autocorrelation of W_A's and W_B's orientation posterior distribution (OPD), which is equivalent to repeating the intensity model labels in Eq. (18). In Fig. 4, the angular width of the blue and red points show the orientation inconsistency for decryption the orientations of two sentinel patterns (K_1 and K_2). The RMS of $\Delta\theta_c^2(W_A, W_A)$ and $\Delta\theta^2(W_B, W_B)$ is used to approximate the angular width (red or blue distribution) in Fig. 4, because it is expensive to calculate the inconsistency between W_A and W_B for each sentinel patterns and it is a good approximation when the OPD is assumed to be a Gaussian distribution (see more details in "A one-dimensional (1D) model" section). Thus $\Delta\theta_i$ simply averages this width over all sentinel patterns and both reconstructions W_A and W_B.

3. Measure of orientation disagreement:

$$\Delta\theta_a(W_A, W_B) = \sqrt{(\Delta\theta_c(W_A, W_B))^2 - (\Delta\theta_i(W_A, W_B))^2} \, , \tag{2}$$

which is the angular displacement between reconstructions W_A and W_B that is not due to an overall rotation between the two volumes, nor from the angular width $\Delta\theta_i$ of the OPD. In "A one-dimensional (1D) model" section, this relation is illustrated with a 1D model in more detail.

4. Measure of orientation inconsistency given the ground truth:

$$\Delta\theta_i^* = \Delta\theta_c(W_T, W_T) \, , \tag{3}$$

which measures the angular width of the OPD in determining the patterns' orientations given the ground truth W_T. With enough patterns in $\{K_A\}$ and $\{K_B\}$, such that W_A and W_B do not over-fit to their respective photon patterns, we expect $\Delta\theta_i \geq \Delta\theta_i^*$.

5. Measure of orientation disconcurrence with ground truth:

$$\Delta\theta_c^*(W_A) = \Delta\theta_c(W_A, W_T) \, , \tag{4}$$

which is the angular width of the ADD between the reconstructed and ground truth intensity volumes (W_A vs W_T respectively). Notice that $\Delta\theta_c$ is identical to $\Delta\theta_c^*$ above if we replaced $W_B \rightarrow W_T$. Hence, $\Delta\theta_c^*$ is essentially the orientation disconcurrence between W_A and the ground truth.

6. Measure of average orientation disconcurrence with ground truth:

$$\langle \Delta\theta_c^* \rangle = \sqrt{\frac{1}{2} \sum_{i \in \{A,B\}} \left(\Delta\theta_c^*(W_i) \right)^2} \, , \tag{5}$$

which is the average angular width of the ADDs between the reconstructed versus the ground truth intensity volumes (W_A, W_B vs W_T respectively). If only two volumes were reconstructed, W_A and W_B, then $\langle \Delta\theta_c^* \rangle$ represents the average orientation disconcurrence against the ground truth.

Factors that influence disconcurrence. Many experimental factors influence the orientation disconcurrence of an SPI intensity reconstruction including: incident photon fluence, number of photon patterns from single particles, resolution and sampling of each pattern, amount of missing detector data (i.e. beamstop, gaps in compound detectors, inactive pixels), extent of photon background (i.e. from particles' incoherent scattering or stray light sources), degree of structural heterogeneity between particles in the ensemble. The choice of algorithms and their parameters used to reconstruct the intensities also play important roles. Furthermore, the symmetries of the scatterer itself can also affect how the ADD is interpreted (see Fig. 9 and "Methods" section).

In this section, we focus on three of these factors: the average number of photons per pattern N, the fineness of orientation space sampling by reconstruction algorithms, and the number of patterns M_{data}. In each scenario studied below, we simulated diffraction patterns with a small 105 kDa protein (PDB code, 4ZW6[42]) under experimental conditions that were modeled after those at the Tender X-ray endstation at the Linac Coherent Light Source (see Table 1). We then used the EMC algorithm to reconstruct two independent 3D volumes each from disjoint sets $\{K_A\}$, $\{K_B\}$, each with M_{data} patterns. For each test condition, a single set of 1000 sentinel patterns was reserved $\{K_S\}$ to evaluate the six types of $\Delta\theta$ listed above. The user should choose the number of sentinel patterns such that the uncertainties of their orientation disconcurrence is acceptably small. Another consideration is whether the range of SO(3) orientations is adequately covered by randomly oriented sentinel patterns (see "Sentinel pattern coverage in the SO(3) orientation space" section).

The average number of photons per diffraction pattern (N) is directly related to the mutual information for inferring latent parameters (e.g. orientations) as well as the particle's structure[8]. N depends on the brightness of the X-ray beam, the size of the X-ray focus (i.e. beam intensity), as well as the relative alignment between particle

Parameter	Value
Photon wavelength (Å)	3.4
Detector distance (mm)	300
Detector pixel size (mm)	1.2
Detector size (pixel)	100×100
Beamstop radius (pixel)	10
Photon fluence (photons μm^{-2})	10^{13} to $5\,5 \times 10^{13}$
Focal area, μm^2	0.33^2 to 0.15^2

Table 1. Range of parameters used to simulate XFEL-SPI photon patterns of a 105 kDa protein (PDB code, 4ZW6) in this paper. Here we assume that the incident beam energy 3 mJ, transmission efficiency 20%, and a binned detector is used here for computational efficiency.
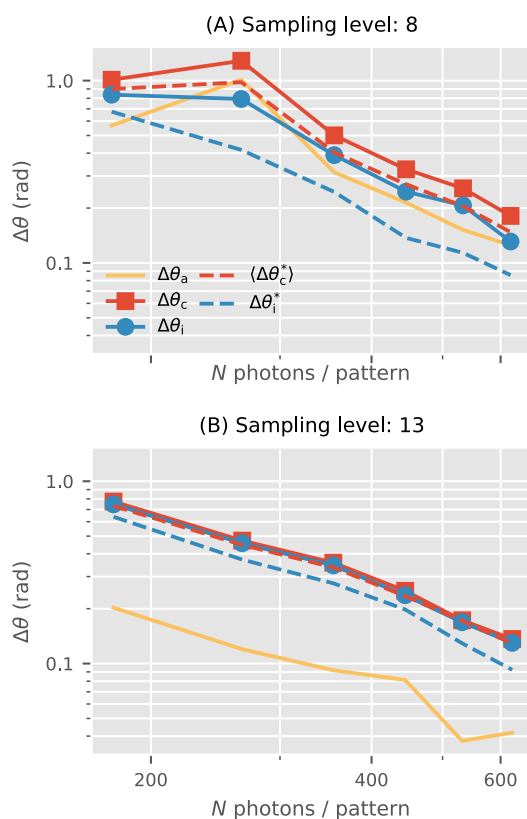

Figure 6. Effects of incident photon counts per pattern and sampling fineness of the latent orientation space. Each data point compares two 3D intensity reconstructions with 5000 photon patterns (solid lines), or each one of them with a ground truth 3D intensity volume (dashed lines). The rotation group is sampled with refinement levels $n = 8$ or $n = 13$. As the average photon counts per pattern increases, all varieties of angular uncertainties specified in "Measures of orientation uncertainties" section decrease. The uncertainties involving the ground truth (*-superscript, dashed lines here) are typically lower than those with only the reconstructed volumes (solid lines). Finer orientation sampling reduces all orientation uncertainties. Furthermore, orientation disconcurrence ($\Delta\theta_c$, red) is dominated by inconsistency ($\Delta\theta_i$, blue) as orientation disagreement ($\Delta\theta_a$, yellow) is suppressed.

and X-ray beams. In general, all six types of $\Delta\theta$ fall when N increases in Fig. 6. Simply put, more photons per pattern reduces orientation disagreement and inconsistency, hence disconcurrence. Additionally, the orientation disconcurrence between W_A and W_B falls with their respective disconcurrences with the ground truth W_T. This correspondence is consistent with the fact that uniqueness is a necessary condition for correctness (i.e. 'precision ≤ accuracy').

How finely orientations are sampled in XFEL-SPI reconstruction algorithms impacts the quality of reconstructed results[8]. Recall, this sampling fineness is different from the adaptive refinement scheme for OPD and ADD Eq. (12): the former pertains to the reconstruction algorithm, while the latter evaluates the reconstructed results. Fig. 6 shows that a higher sampling level in the EMC reconstruction algorithm generally reduces all
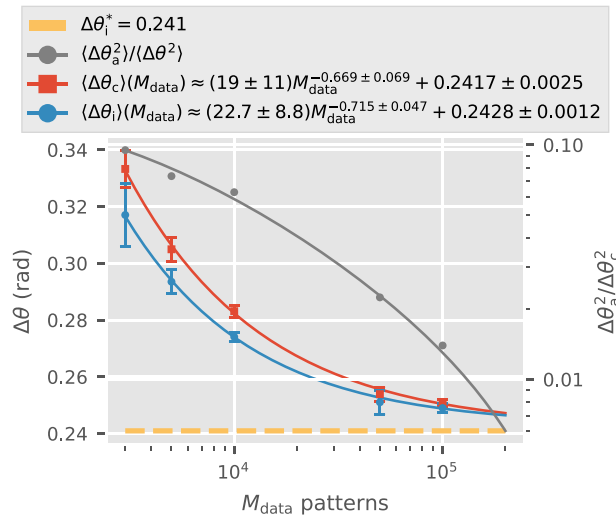
Figure 7. Orientation disconcurrence ($\Delta\theta_c$) and inconsistency ($\Delta\theta_i$) converge to $\Delta\theta_i^*$ as the number of patterns (M_{data}) increase. Each dot and its error bars represent the average and standard deviation of $\Delta\theta$ of all pairs among five reconstructions from four different disjoint datasets (average of 355 photons/pattern, rotation group sampling $n = 13$). The same 1000 sentinel patterns are used in all four instances. The ratio of orientation disagreement $\Delta\theta_a$ to disconcurrence $\Delta\theta_c$, which is represented by the grey curve (labeled on right vertical axis), decreases with increasing M_{data}.

alignment uncertainties $\Delta\theta$. While the various forms of $\Delta\theta$ have a noticeable spread at $n = 8$ orientation sampling, this spread significantly reduces when this sampling fineness is increased to $n = 13$. Numerically, we found the average angular separation between the quasi-uniform unit quaternions samples to be 0.161 and 0.099 radians respectively. This figure complements the information-theoretic heuristic for deciding sampling sufficiency in[8]. With sufficient sampling, Fig. 6 shows that the orientation disconcurrence is dominated by the orientation inconsistency rather than orientation disagreement: $\Delta\theta_c(W_A, W_B) \approx \Delta\theta_i(W_A, W_B) > \Delta\theta_a(W_A, W_B)$.

In an SPI experiment the number of SPI patterns, M_{data}, is a product of the fraction of particles that are illuminated by x-ray pulses (i.e. hit-rate), the pulse repetition rate, and the total experiment time. One intuitively expects that reconstructions improve with larger M_{data}, which Fig. 7 confirms. The intrinsic orientation inconsistency of each reconstruction, $\Delta\theta_i$, falls with more patterns (blue curve). The orientation disconcurrence $\Delta\theta_c$, likewise, also falls with more patterns.

We found that in Fig. 7 that $\Delta\theta_c$ and $\Delta\theta_i$ both decrease numerically with the number of patterns as $\alpha M_{data}^{-\beta} + \Delta\theta_i^*$, where α is a multiplicative constant, β is a real positive number, and $\Delta\theta_i^*$ is the angular width of the OPD given the patterns $\{K_S\}$ and ground truth model. Although $\Delta\theta_c \rightarrow \Delta\theta_i^*$ as $M_{data} \rightarrow \infty$, we can only assert that the reconstructed pairs of models (W_A and W_B) are closer to each other, but not whether either are close to the ground truth W_T. The former is evident from the ratio of orientation disagreement against disconcurrence, $\Delta\theta_a^2/\Delta\theta_c^2$ (gray dots in Fig. 7): increasing M_{data} eliminates orientation disagreements ($\Delta\theta_a$) between two independent reconstructions faster than intrinsic inconsistency ($\Delta\theta_i$). Using Eq. (2) and the fitted forms in Fig. 7, this vanishing of the orientation disagreement becomes clear:

$$
\begin{aligned}
\Delta\theta_a &= \sqrt{\Delta\theta_c^2 - \Delta\theta_i^2} \\
&= \sqrt{\left(\alpha_c M_{data}^{-\beta_c} + \gamma_c\right)^2 - \left(\alpha_i M_{data}^{-\beta_i} + \gamma_i\right)^2} \\
&\approx M_{data}^{-\beta_c/2} \sqrt{(\alpha_c + 2\gamma)\alpha_c} ,
\end{aligned}
\tag{6}
$$

where we assumed $\beta_c < \beta_i$, and $\gamma_c \approx \gamma_i = \gamma$. Obviously, when M_{data} approaches infinity, $\Delta\theta_a$ gets close to 0. Simply put, as M_{data} increases independently reconstructed volumes become more unique but not necessarily more correct.

Relating $\Delta\theta$ to spatial resolution.

The 3D speckles in the reconstructed diffraction volume whose angular width are smaller or comparable to $\Delta\theta_c$ will lose contrast, hence spatial resolution. Let us denote the full angular width of these 3D speckles as $2\Delta\theta_{sp}(\mathbf{q})$ at spatial frequency $\mathbf{q}$. Naturally, the resolutions of reconstructions become orientation-limited at the frequencies where $\Delta\theta_{sp}(\mathbf{q})$ approaches the width of OPD which is about $\Delta\theta_c/\sqrt{2}$ ("A one-dimensional (1D) model" section).

We caution that the previous paragraph suggests an inequality rather than strict equality between spatial resolution and orientation disconcurrence. To understand why, consider how Fig. 8 shows that it is possible for reconstructions whose orientation disconcurrence is smaller than the angular width of a single pixel at the edge of the detector $\Delta\theta_{pix}$. This situation occurs with very high average number of photons per pattern ($N \gg 1$), abundant patterns ($M_{data} \gg 1$), and sufficiently fine sampling of the rotation group during reconstructions
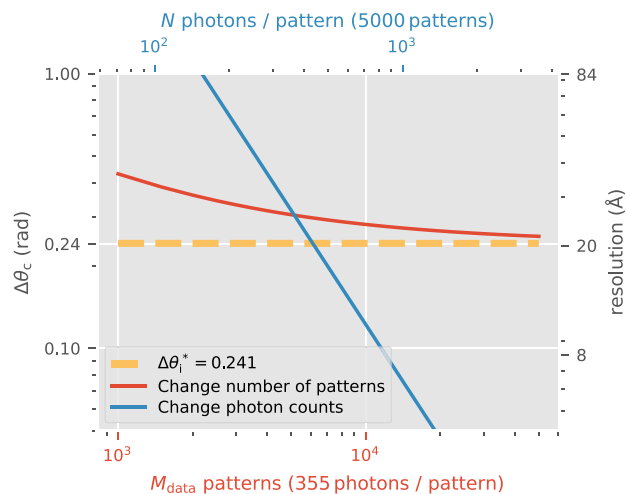
Figure 8. This figure shows how $\Delta\theta_c$ changes by increasing number of patterns (red curve, with $N \approx 355$) or number of photons per pattern (blue curve, with $M_{data} = 5000$). The measure of orientation inconsistency given the ground truth, $\Delta\theta_i^*$ (yellow), is computed for $N \approx 355$. The right axis shows the relation to spatial resolution according to Eq. (7) (2σ criterion).

(Fig. 6). Thus, the dynamic range and contrast of the reconstructed 3D diffraction speckles are high up to the detector's maximum captured resolution ($\mathbf{q}_{max}$), which allows us to distinguish arbitrarily small angular variations between actual diffraction patterns.

We must remember that the reconstructed diffraction volume W does not explicitly contain spatial information beyond the maximum spatial resolution $\mathbf{q}_{max}$. So even if $\Delta\theta_c \ll \Delta\theta_{pix}$, we can only say that spatial resolution is not orientation limited. Perhaps with additional priors about the structure of the particle (e.g. know sequence, similar structure known, atomicity, etc) is might be possible to extend the resolution beyond $\mathbf{q}_{max}$. But such extensions are beyond the scope of this discussion.

It should now be clear that orientation disconcurrence relates to how effectively one can resolve the orientation of an average SPI photon pattern. From this section, it should also be clear that spatial resolution can be limited by large orientation disconcurrences. More concretely, consider Fig. 8, which simulates an XFEL-SPI experiment of a 105 kDa protein at the Tender X-ray endstation at LCLS (Table 1). To resolve this protein to 10nm-resolution without significant orientation blurring requires more than 5000 patterns each with more than 600 photons. However, it is premature to define spatial resolution only in terms of orientation concurrence, especially since a decryption scheme for the spatial resolution (similar to Fig. 3) is absent. Such detailed discussions, however, are deferred to future studies.

Data sufficiency and mutual information. The question 'how many patterns are sufficient?' frequently occur in an XFEL-SPI experiment. The answer to this hypothetical question determines if a proposed experiment is 'feasible', as well as how many different samples to inject during the precious dozens of hours of XFEL beamtime allocated to each user group. Orientation disconcurrence can be used to define data sufficiency: when the number of patterns gives a disconcurrence smaller than the angular width of speckles at a target resolution $\mathbf{q}_{target}$:

$$2 \cdot \frac{\Delta\theta_c}{\sqrt{2}} \leq \theta_{sp}(\mathbf{q}_{target}) \,. \tag{7}$$

If the ADD peak in Fig. 4 were compact and locally Gaussian ("A one-dimensional (1D) model" section), this last condition means that approximately 74% (2σ criterion) of the oriented sentinel patterns should intersect their target 3D speckle at resolution $\mathbf{q}_{target}$.

With the disconcurrence target defined, we can extrapolate data sufficiency with bootstrapping. Given M_{data} total patterns, one can compute $\Delta\theta_c(M_{data})$ for pairs of models reconstructed from random, non-overlapping, equal subsets from the full M_{data} dataset similar to the data points in Fig. 7. Repeating this procedure via a simple bootstrapping scheme gives the orientation disconcurrence curves in Fig. 7. These curves fit reasonably well to a lifted power law, $\Delta\theta_c = \alpha_c M_{data}^{-\beta_c} + \gamma_c$. The shrinking error bars on $\Delta\theta_c$ from bootstrapping with increasing M_{data} in Fig. 7 suggests that this fit requires sufficiently many patterns to be robust.

Owing to various constraints, only a finite number of XFEL-SPI patterns are collected each time (say M_{exp}). To maximize signal-averaging in a reconstruction logically requires the input from all collected patterns. Yet the two independent reconstructions in this framework (Fig. 3) only sees only a little less than half of the full dataset ($< M_{exp}/2$). Fortunately, the lifted power law fit in Fig. 7 allows us to extrapolate the orientation disconcurrence between a pair of hypothetical independent 3D reconstructions that each used all patterns in an XFEL-SPI dataset. Specifically, if $\Delta\theta_c(M_{data} \leq M_{exp}/2)$ were computed between pairs of reconstructed volumes each using up

to $M_{\text{exp}}/2$ bootstrapped photon patterns, then the angular uncertainty of a single volume with all M_{exp} patterns can be extrapolated using the fit: $\Delta\theta_c(M_{\text{data}} = M_{\text{exp}}) = \alpha_c M_{\text{exp}}^{\beta_c} + \gamma_c$. A similar extrapolation from bootstrapped reconstructions was proposed to define spatial resolution in cryo-electron microscopy[43].

This lifted power law also helps us extrapolate to a second scenario. Should the target orientation disconcurrence be the angular width of a single pixel at the edge of the detector, $\Delta\theta_c = \Delta\theta_{\text{pix}}(\mathbf{q}_{\text{max}})$, then $\gamma_c < \Delta\theta_{\text{pix}}(\mathbf{q}_{\text{max}})$ is required. If this requirement is satisfied, then $\frac{1}{\beta_c}\log\left[\alpha_c/(\Delta\theta_{\text{pix}}(\mathbf{q}_{\text{max}}) - \gamma_c)\right]$ patterns are needed to reach this target.

The lifted power law form of $\Delta\theta_c = \alpha_c M_{\text{data}}^{-\beta_c} + \gamma_c$ in Fig. 7 allows us to parametrize data sufficiency in an information-theoretic sense. Essentially, the mutual information here can be defined as the reduction in the entropy of orienting an average sentinel pattern give a set of M_{data} photon patterns $\{K\}$. Ignoring factors of order unity, this mutual information, is approximately

$$I(\Omega_S, \{K\}) \approx \log\left(\frac{2\pi^2}{\Delta\theta_c^3}\right)$$
$$\approx \log\left(\frac{2\pi^2}{\Delta\theta_i^{*3}}\right) - \frac{3\alpha_c}{\Delta\theta_i^*}M_{\text{data}}^{-\beta_c}, \tag{8}$$

assuming $M_{\text{data}} \gg 1$.

Equation (8) contains two intuitive results. First, this mutual information is bounded from above by that when the solution intensities are known: $\log\left(2\pi^2/(\Delta\theta_i^*)^3\right)$. This upper bound can be viewed as the SPI channel capacity for decryption orientations, and is computed in the same manner as the mutual information $I(K, \Omega)|_W$ in[8]. Second, the mutual information for decryption orientations increases with the number of patterns. This assumes that $\alpha_c/\Delta\theta_i^* > 0$ and $\beta_c > 0$, which are manifest in Fig. 7. Furthermore, $\beta_c > 0.5$ in Fig. 7, which is better than one would expect if patterns were mutually independent (i.e. $\beta_c = 0$). This 'co-dependence' arises because additional patterns can improve the reconstructed volumes, which in turn help earlier patterns distribute their photons more precisely into orientation classes.

Focal spot size affects hit rate and orientation disconcurrence.

The linear size of the XFEL focus L_{focus} is a critical parameter in an SPI experiment (see Table 1). This choice of focus size can be paraphrased simply: given a fixed total number of photons per XFEL pulse, would it be better to 'distribute' them into more patterns with fewer photons each, or fewer patterns with more photons each? Whereas a larger focus can dramatically increase the odds of illuminating randomly injected particles, it also drastically decreases the number of scattered photons should a particle be illuminated (N). These odds, also known as the 'hit-rate', is effectively M_{data} per time. In fact, $N \propto L_{\text{focus}}^{-2}$ while $M_{\text{data}}/\text{time} \propto L_{\text{focus}}^2$. In this hypothetical scenario, the total number of photons measured per time ($NM_{\text{data}}/\text{time}$) remains constant despite L_{focus}. Suppose that in either case, you had enough patterns to adequately sample different views of the scatterer, and were perfectly able to detect particle hits against background scatter/noise. This same ambivalence to the focus size appears again in the simple signal-to-noise ratio (SNR) described in[8]:

$$\text{SNR} = \left(\frac{NM_{\text{data}}}{M_{\text{rot}}}\right)^{1/2}, \tag{9}$$

where M_{rot} is the number of rotation samples used to reconstruct the intensity volumes W_A and W_B. This SNR is motivated by a simple distribution of photons across a limited number of Ewald tomograms, and has been used to indicate data sufficiency in the orientation space[9].

The discussion above may lead one to believe that there is no ideal focus size. However, if we again used a smaller orientation disconcurrence $\Delta\theta_c$ to quantify when things are 'better', the preference is to reduce L_{focus}. Notice that nearly doubling the average number of photons per pattern ($N = 355$ to $N = 622$ given $M_{\text{data}} = 5000$) in Fig. 6 reduces both $\Delta\theta_c$ and $\Delta\theta_i$ more than if we doubled the number of patterns ($M_{\text{data}} = 5000$ to $M_{\text{data}} = 10000$ given $N = 355$) in Fig. 7. The total number of photons in all patterns is approximately equal in both cases. Yet doubling the average number of photons per pattern substantially improves the asymptotic orientation inconsistency (i.e. $\Delta\theta_i^*$ falls).

Discussion

In summary, we propose an encryption–decryption approach to validate 3D intensity volumes reconstructed in XFEL-SPI. This validation is based on the volumes' ability to decrypt the orientations of sentinel patterns unused in these reconstructions. While these volumes can be reconstructed from any algorithmic means, they must strictly adhere to the data independence scheme laid out in Fig. 3. This scheme can be generalized to validate other latent information inferred within the full dataset (e.g. unmeasured local photon fluence, structural class, etc).

From realistic simulations of SPI experiments this approach can validate reconstructions in a principled information-theoretic manner. Our approach relates the challenging question of data sufficiency intuitively to key experimental variables such as the number of measured photon patterns, and nominal incident photon intensity. Furthermore, the various forms of decrypting (orientation) uncertainties shown here can be interpreted as disconcurrence, disagreement, and inconsistencies in how confidently the latent variables are inferred. These interpretations give a more informative and comprehensive view of the validation exercise.

Whereas there were studies about the expected scattered photon signals from biomolecules in idealized XFEL-SPI scenarios[44,45], systematic studies of how well these signals can be integrated into a 3D diffraction

volume despite missing information is still sorely lacking. Our results show that the complex considerations that contribute to data sufficiency in XFEL-SPI can be fitted as simple parameters (e.g. α, β, γ). Relating these parameters to basic properties of the target scatterer (e.g. mass, radius of gyration, etc), experimental conditions (e.g. beam intensity, photon wavelength, background scattering, etc), and choice of reconstruction algorithms, will be useful for experiment design and planning.

An extension of our encryption–decryption approach can be used to define and validate the spatial resolution of XFEL-SPI and cryo-electron microscopy reconstructions. In principle, the resolving power of an imaging instrument should be the reduction in uncertainty of locating spatial features within the sample. Re-framing this uncertainty reduction in the encryption–decryption framework of Fig. 3 may give rise to more interpretable notions of spatial resolution. This information theoretic formulation of this conceptual framework, similar to Eq. (8), also naturally accounts for external priors for localizing spatial features.

Ultimately, our encryption-decryption approach demonstrably overcomes the difficulties of using FSC as a validation measure for XFEL-SPI, in spite of FSC's popularity[13,16,18–29]. The data throughput from XFELS will rapidly increase because of higher pulse repetition rates[46], and more efficient sample injection techniques. This trend inevitably creates a larger data load, which in turn increases our reliance on statistical techniques to assign confidence to de novo structural reconstructions. Such confidence is especially important when imaging structural ensembles with considerable flexibilities, or other structural variations. Despite the specificity of our validation routine to orientations, the encryption–decryption framework proposed in Fig. 3 can be readily generalized to test the reproducibility of claims of novel reconstructed structures. Such tests, we believe, are central to illuminating our path towards novel structural insights as we navigate through the photon-limited world of XFEL-SPI.

Methods

Sampling orientations.
A scatterer can take on an infinite number of possible 3D orientations. In practice these orientations Q are discretely sampled to angular divisions smaller than the intrinsic angular precision of the patterns (see "Relating $\Delta\theta$ to spatial resolution" section). We adopt a quasi-uniform sampling scheme based on[8], which adaptively refines the 600-cell polytope with refinement parameter n. In this scheme the number orientation samples scales like n^3, while their angular resolution increases like $1/n$.

Orientation posterior distribution (OPD) of sentinel patterns.
The orientation posterior distribution (OPD) of a particular sentinel pattern K_S defines the probability of orienting it within a specific 3D diffraction volume W. This OPD, written here as $P(Q \mid K_S, W)$, can be inferred from the likelihood $P(K_S \mid Q, W)$ using Bayes' theorem,

$$P(Q \mid K_S, W) \propto P(K_S \mid Q, W)\, P(Q), \tag{10}$$

where the prior distribution of orientations, $P(Q)$, is uniformly distributed unless the specimens have a known orientation bias. Because the space of orientations is only quasi-uniformly sampled by unit quaternions in our discretization scheme, we replace $P(Q)$ with the numerically computed non-uniform weights $w(Q)$[9]. Note that this OPD can be computed even if K_S did not in fact originate from W: such a computation will naturally yield highly uncertain orientations of K_S.

We presume the likelihood of detecting a sentinel pattern K_S (comprising pixels indexed by t) from the Ewald tomogram at orientation Q of volume W (see Fig. 1) assuming perfect detection absent background photon sources is

$$P(K_S \mid Q, W) = \prod_{t \in \text{detector}} \frac{e^{-W_{Qi}}\, W_{Qt}^{K_{St}}}{K_{St}!}. \tag{11}$$

This likelihood can be replaced if the true detection statistics departs from this Poissonian form.

Often the posterior and likelihood in Eqs. (10) and (11) of a converged intensity volume is significant only for a relatively small set of orientations. For a given pattern K_S, we represent this set of *important orientations* by their corresponding important unit quaternions $\{\boldsymbol{Q} \mid K_S\}$ (written in boldface). For computation efficiency, only the probability at $\{\boldsymbol{Q} \mid K_S\}$ is recorded; those at other quaternions are safely set to zero.

For sufficient orientation coverage, we require these important quaternions to capture at least 99% of the total posterior distribution. To implement this, all patterns' posterior distributions are first sampled by a unit quaternion set $\{Q \mid n\}$ with 600-cell quaternion sampling strategy[8] where n is the sampling refinement level. Then we increase n until the smallest set of important quaternions $\{\boldsymbol{Q} \mid K_S, n\}_{\min} \subset \{Q \mid n\}$ that captures this total posterior distribution comprises at least 100 important quaternions:

$$\left\langle \sum_{Q \in \{\boldsymbol{Q} \mid K_S, n\}_{\min}} P(Q \mid K_S, W) \right\rangle_{K_S} \geq 0.99 \,, \tag{12}$$

and the size of every $K_S, |\{\boldsymbol{Q} \mid K_S, n\}_{\min}| \geq 100$. To be concise, we omit the subscript $\cdot_{\min}$ in subsequent formulae.

Angular displacement distribution (ADD) between two reconstructed volumes.
Returning to our cryptography analogy, our next step is to compare how two diffraction volumes decrypt the orientations of a set of sentinel patterns. Three key considerations stand out here. First, the orientation of a noisy sentinel pattern is described by a probability distribution (i.e. OPD) rather than a point estimate. Second, W_A and W_B would almost always differ by an overall mutual 3D rotation Q_{BA} because each volume is typically randomly initialized

to avoid reconstruction biases. Hence, the sentinel OPDs for W_A and W_B would also be displaced by Q_{BA}. Third, we must average the OPDs for different sentinel patterns to obtain a robust estimate of the orientation disconcurrence between W_A and W_B. These considerations are captured in the *angular displacement distribution* (ADD) between W_A and W_B. The ADD allows us to compare the OPD of a single sentinel pattern (K_S) given W_A and W_B without having to pre-align them in the space of possible orientations.

Mathematically, the ADD for a single sentinel pattern K_S is the outer product (or convolution) of its two OPDs given W_A and W_B on their respective important quaternions,

$$
\begin{aligned}
P(\boldsymbol{Q}_{BA}|K_S, W_A, W_B) &= \sum_{\boldsymbol{Q}_A} P(\boldsymbol{Q}_A|K_S, W_A) P(\boldsymbol{Q}_B|K_S, W_B) \\
&= \sum_{\boldsymbol{Q}_A} P(\boldsymbol{Q}_A|K_S, W_A) P(\boldsymbol{Q}_{BA}\boldsymbol{Q}_A|K_S, W_B) \,,
\end{aligned}
\tag{13}
$$

which is computed over the set of important unit quaternions. Here $\boldsymbol{Q}_{BA} = \boldsymbol{Q}_B \boldsymbol{Q}_A^{-1}$ represents the possible relative orientations between the reconstructed volumes W_A and W_B over the two sets of important quaternions $\{\boldsymbol{Q}_A|K_S\}$ and $\{\boldsymbol{Q}_B|K_S\}$ as defined in Eq. (12). Since $\boldsymbol{Q}_{BA}$ depends on the sentinel pattern K_S, the ADD in Eq. (13) may be different for different K_S. Averaging the ADD over all the set of sentinel patterns $\{K_S\}$ we get

$$
P(\boldsymbol{Q}_{BA}|\{K_S\}, W_A, W_B) \equiv \Big\langle P(\boldsymbol{Q}_{BA}|K_S, W_A, W_B) \Big\rangle_{\{K_S\}} .
\tag{14}
$$

Given the noise in the diffraction patterns, we expect variations in the decrypted orientations of sentinel patterns. To compute this variation, an average of an ADD must be established. When the reconstructed volumes W_A and W_B are similar, the ADD of their many sentinel patterns tend to cluster around the average unit quaternion $\overline{Q}_{AB}$ in orientation space. This overall rotation $\overline{Q}_{AB}$ is not a mere linear average of the unit quaternions that sample the ADD since this average may not have unit length and hence not correspond to a 3D spatial rotation. To define $\overline{Q}_{AB}$, let us first consider the relative rotation between $\boldsymbol{Q}_{BA}$ and a presumptive average overall rotation $\widetilde{Q}$. This relative rotation can be written as a quaternion multiplication

$$
\boldsymbol{Q}_{BA}^{-1} \widetilde{Q} = \left\{ \cos\left(\frac{\theta}{2}\right), \sin\left(\frac{\theta}{2}\right)\hat{\boldsymbol{n}} \right\},
\tag{15}
$$

which is written here as a four-component vector; $\hat{\boldsymbol{n}}$ and θ are respectively the axis and magnitude of this relative rotation. The magnitude of this relative rotation, $\theta(\boldsymbol{Q}_{BA}, \widetilde{Q})$, vanishes as $\widetilde{Q}$ approaches $\boldsymbol{Q}_{BA}$.

We define the average overall rotation $\overline{Q}_{BA}$ of an ADD between W_A and W_B as that which minimizes the average θ against all the rotation samples of the ADDs for the set of sentinel patterns. Specifically, the average overall rotation is defined as the unit quaternion that minimizes the angular variance Θ^2:

$$
\overline{Q}_{BA} \equiv \underset{\widetilde{Q}}{\arg\min}\ \Theta^2\big(\widetilde{Q} \,\big|\, \{K_S\}, W_A, W_B\big),
\tag{16}
$$

and the orientation disconcurrence is the minimum value of $\sqrt{\Theta^2}$:

$$
\begin{aligned}
\Delta\theta_c(W_A, W_B) &\equiv \min_{\widetilde{Q}} \sqrt{\Theta^2\big(\widetilde{Q} \,\big|\, \{K_S\}, W_A, W_B\big)} \\
&= \sqrt{\Theta^2\big(\overline{Q}_{BA} \,\big|\, \{K_S\}, W_A, W_B\big)} \,,
\end{aligned}
\tag{17}
$$

where the angular variance is defined as

$$
\begin{aligned}
&\Theta^2\big(\widetilde{Q} \,\big|\, \{K_S\}, W_A, W_B\big) = \\
&\left\langle \sum_{\{\boldsymbol{Q}_{BA}|K_S\}} P(\boldsymbol{Q}_{BA} \,|\, K_S, W_A, W_B)\, \theta^2(\boldsymbol{Q}_{BA}, \widetilde{Q}) \right\rangle_{\{K_S\}} .
\end{aligned}
\tag{18}
$$

A special case here is when W_A and W_B are identical. In this case, $\overline{Q}_{BA} = (1, 0, 0, 0)$ which is the identity quaternion.

Resolving ambiguities from centro-symmetric diffraction volumes.

To obtain the most compact ADD (Eq. (14)), we must eliminate trivial symmetries in the diffraction patterns that broaden the ADD. One such example is the centro-symmetry of 3D diffraction intensities from optically thin samples, whose scattering density distribution is effectively real-valued. Consequently, at sufficiently low resolutions any two-dimensional diffraction pattern is similar to itself after a 180° in-plane rotation about the scattering experiment's optical axis ($\hat{z}$). Each such photon pattern K should have similar posterior probabilities to occur at either rotation Q or QQ_z:

$$
P(Q \,|\, K, W) \approx P(QQ_z \,|\, K, W) \,,
\tag{19}
$$

where the in-plane rotation about the z-axis is $Q_z = (0, 0, 0, 1)$. This two-fold ambiguity plus the fact that Q_z is its own inverse, means that in ADD, the relative rotation Q_{BA} or $Q'_{BA} = Q_B Q_z (Q_A)^{-1}$ could occur in Eq. (14). Hence, for each ADD sample we check the angular closeness of both Q_{BA} and Q'_{BA} to the ADD's average unit quaternion $\overline{Q}_{BA}$, and keep the one that is closer. This essentially replaces the θ expression in Eq. (18):
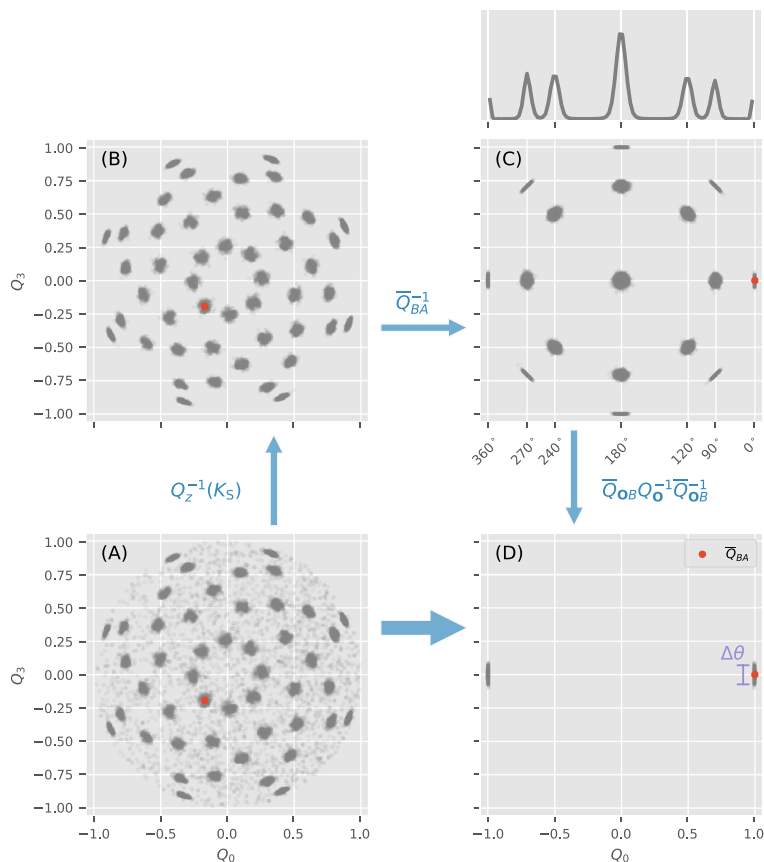
Figure 9. Collapsing the ADD of 500 sentinel patterns for a scatterer, whose diffraction volumes is centro-symmetric and has octahedral symmetry, into the fundamental domain: (**A**–**D**). Starting clockwise from (**A**), which shows a projection of the ADD onto two components of each quaternion ($Q = (Q_0, Q_1, Q_2, Q_3)$), we collapsed the points related by centro-symmetry (since 2D patterns have sufficiently low resolution) to obtain a sharper distribution in (**B**). The red disk throughout the panels represent the average quaternion $\overline{Q}_{AB}$ of the ADD. In (**C**), we rotate the ADD such that $\overline{Q}_{AB} = (1, 0, 0, 0)$ for clarity. The histogram of the ADD vs Q_0 is shown above panel (**C**), can sometimes reveal the flavor of symmetry in W. Finally, using the particle's known symmetry group operations we can fold the ADD into the fundamental domain in (**D**).

$$\theta^2(\mathbf{Q}_{BA}, \widetilde{Q}) \rightarrow \min\{\theta^2(\mathbf{Q}_B\mathbf{Q}_A^{-1}, \widetilde{Q}), \theta^2(\mathbf{Q}_B Q_z\mathbf{Q}_A^{-1}, \widetilde{Q})\} . \tag{20}$$

Discrete symmetries in the diffraction volume. Discrete symmetries in the diffraction volume can create multiple clusters in the ADD (Fig. 9). Examples of such symmetries include icosahedral viral capsids[13] and octahedral nanoparticles[18]. The multiplicity of these clusters arise because each pattern could be oriented at different and/or multiple locations of the symmetry orbit within the diffraction volume. As Fig. 9 shows, should this symmetry be known we can compute a single orientation disconcurrence by first folding these multiple symmetry-related peaks in ADD into its fundamental domain. We emphasize that this folding can be done even if this symmetry were not imposed during the reconstructions of W_A and W_B.

Figure 9 illustrates ADD folding for a particle with chiral octahedral symmetry (**O**). The reconstructed diffraction intensities of this particle (W_A and W_B) has 24 rotational symmetries (of order 24). Once W_A's body axes are canonically aligned, then each of these symmetry rotations can be represented by a canonical set of unit quaternions $\{Q_\mathbf{O} \mid [Q_\mathbf{O}] \in \mathbf{O}\}$ ($[Q_\mathbf{O}]$ is the equivalence class $Q_\mathbf{O} \sim -Q_\mathbf{O}$ owing to unit quaternions double covering SO(3).

To see how this symmetry manifests in an ADD, consider orienting a particular sentinel pattern K_S within W_A and W_B. Note that even though W_A and W_B have **O** symmetry, they are not canonically aligned by default. First, we focus on a tomogram of W_B at $\mathbf{Q}_B$, $T(\mathbf{Q}_B, W_B)$. Here, the symbol for tomogram is changed from the W_Q in the main text to avoid multiple level subscript. When we align W_B canonically by actively rotating it to $\widetilde{Q}_{\mathbf{O}B}[W_B]$, the tomogram should be rotated together to maintain unchanged, where $\widetilde{Q}_{\mathbf{O}B}$ actively rotates W_B to $\widetilde{Q}_{\mathbf{O}B}[W_B]$ into the canonical axes for the symmetry operations in $\{Q_\mathbf{O}\}$. In other words, we have

$$T(\mathbf{Q}_B, W_B) = T(\widetilde{Q}_{\mathbf{O}B}\mathbf{Q}_B, \widetilde{Q}_{\mathbf{O}B}[W_B]) \tag{21}$$

$$= T\big(\widetilde{Q}_{\mathbf{O}B}\boldsymbol{Q}_B, (Q_{\mathbf{O}}\widetilde{Q}_{\mathbf{O}B})[W_B]\big) \tag{22}$$

$$= T\big(\widetilde{Q}_{\mathbf{O}B}^{-1}Q_{\mathbf{O}}^{-1}\widetilde{Q}_{\mathbf{O}B}\boldsymbol{Q}_B, W_B\big). \tag{23}$$

The 24 elements in $\{Q_{\mathbf{O}}\}$ give 24 same tomograms at $\widetilde{Q}_{\mathbf{O}B}^{-1}Q_{\mathbf{O}}\widetilde{Q}_{\mathbf{O}B}\boldsymbol{Q}_B$ (all $Q_{\mathbf{O}}^{-1} \in \{Q_{\mathbf{O}}\}$ also), hence the same orientation posterior probability at these orientations. Recalling the ADD comprises the joint product of OPDs for K_S to be oriented at $\boldsymbol{Q}_A$ and $\boldsymbol{Q}_B$ within W_A and W_B respectively. We see this multiplicity of ADD in Fig. 9b (main text), which contains 48 clusters owing to the unit quaternion double covering SO(3). The number of clusters does not increase even if we include the symmetry operations of W_A by assuming W_A and W_B are similar, for the same reason that randomly oriented sentinel patterns in an asymmetric volume still produce a 2-clustered ADD (only one branch is plotted in Fig. 4).

For each sentinel pattern K_S, we can fold each important unit quaternion $\boldsymbol{Q}_{BA}$ in its ADD into the fundamental domain by exhaustively searching the symmetry operation in $\big\{ \widetilde{Q}_{\mathbf{O}B}^{-1}Q_{\mathbf{O}}\widetilde{Q}_{\mathbf{O}B}\boldsymbol{Q}_B \,\big|\, Q_{\mathbf{O}} \in \{Q_{\mathbf{O}}\}\big\}$ and in-plane inversion Q_z (either $\{1, 0, 0, 0\}$ or $\{0, 0, 0, 1\}$) that minimizes the angular variance

$$\theta_{\min}^2\big(\widetilde{Q}_{\mathbf{O}B}, \widetilde{Q} \,|\, K_S, \boldsymbol{Q}_{BA}\big) =$$
$$\min_{\{Q_{\mathbf{O}}\}\times\{Q_z\}} \theta^2\big(\widetilde{Q}_{\mathbf{O}B}^{-1}Q_{\mathbf{O}}\, \widetilde{Q}_{\mathbf{O}B}\boldsymbol{Q}_B Q_z \boldsymbol{Q}_A^{-1}, \widetilde{Q} \,|\, K_S\big). \tag{24}$$

Here, $\widetilde{Q}$ is the presumptive average relative rotation between W_A and W_B similar to that in Eq. (16). Like Eq. (20), we also minimize over each pattern's in-plane inversion. Therefore, the optimal relative rotation ($\overline{Q}_{BA}$) and canonical realignment ($\overline{Q}_{\mathbf{O}B}$) are found by minimizing the total angular variance weighted over all important unit quaternions for all sentinel patterns in the ADD:

$$(\overline{Q}_{\mathbf{O}B}, \, \overline{Q}_{BA}) = \underset{(\widetilde{Q}_{\mathbf{O}B}, \widetilde{Q})}{\arg\min} \, \Theta^2\big(\widetilde{Q}_{\mathbf{O}B}, \widetilde{Q} \,|\, \{K_S\}, W_A, W_B\big)$$

where

$$\Theta^2\big(\widetilde{Q}_{\mathbf{O}B}, \widetilde{Q} \,|\, \{K_S\}, W_A, W_B\big) = \left\langle \sum_{\{\boldsymbol{Q}_{BA} \,|\, K_S\}} P(\boldsymbol{Q}_{BA}|K_S, W_A, W_B)\, \theta_{\min}^2\big(\widetilde{Q}_{\mathbf{O}B}, \widetilde{Q} \,|\, K_S, \boldsymbol{Q}_{BA}\big) \right\rangle_{\{K_S\}}. \tag{25}$$

To recapitulate, the orientation disconcurrence between two symmetric volumes W_A and W_B is defined by Eq. (25) as

$$\Delta\theta_c^2 = \Theta^2\big(\overline{Q}_{\mathbf{O}B}, \overline{Q}_{BA} \,|\, \{K_S\}, W_A, W_B\big). \tag{26}$$

This computation involves separate optimizations: we iteratively refine $\widetilde{Q}_{BA} \to \overline{Q}_{BA}$ and $\widetilde{Q}_{\mathbf{O}B} \to \overline{Q}_{\mathbf{O}B}$ by minimizing Eq. (25); for each presumptive $\widetilde{Q}_{BA}$ and $\widetilde{Q}_{\mathbf{O}B}$, find the symmetry operation in $\{Q_{\mathbf{O}}\}$ for each sentinel pattern that minimizes the quantity in Eq. (24) as well as the most compatible in-plane rotations for each sentinel pattern ("Resolving ambiguities from centro-symmetric diffraction volumes" section). The results of these completed optimizations are used to fold the ADD into the fundamental domain in Fig. 9.

We note that one can discover the symmetry of W_A using a special case of ADD with itself (i.e. $W_A = W_B$). This 'self-ADD' will be similar to Fig. 9c (main text) since there is no relative rotation between W_A and itself. Because the first component of every unit quaternions in a symmetry group is independent on the choice of canonical axis, we may deduce W_A's symmetry group from number and positions of their clusters in their Q_0 histograms of its 'self-ADD' (panel above Fig. 9c (main text)).

A one-dimensional (1D) model. Here, we show the relation between the orientation disconcurrence and the disagreement (misalignment of the centers of ADDs) and the inconsistency (the size of each ADDs) with a one-dimensional (1D) rotation analogy as opposed to the full 3D rotation version in Fig. 4.

The unit quaternion $\boldsymbol{Q}$ that describes rotation about a 1D ring is a real number $\theta \in [0, 2\pi)$. Suppose that the two OPDs (of reconstructed models W_A and W_B) that comprise the ADDs for a set of sentinel patterns $\{K_S\}$ are mostly constrained within a small segment of this 1D ring. Let us further suppose that their ADD over $\{K_S\}$ can be approximated by local Gaussian distribution within this angular segment. We denote the 1D ADD averaged over all sentinel patterns $\{K_S\}$ as $P(\boldsymbol{Q} \,|\, \{K_S\}) \equiv P(\boldsymbol{Q} \,|\, \{K_S\}, W_A, W_B)$. For a single sentinel pattern K_S its ADD, $P(\boldsymbol{Q} \,|\, K_S)$ (blue or red distribution in Fig. 4), we denote its mean as $\overline{Q}(K_S)$, and variance as $\Delta\theta^2(K_S)$. Hence the mean and variance of this ADD for the entire set of sentinel patterns $\{K_S\}$ are equivalent to the overall orientation, $\overline{Q}(\{K_S\})$, and the square of orientation disconcurrence, $\Delta\theta_c^2(\{K_S\})$, defined in Eqs. (17) and (18) respectively. The square difference between the disconcurrence, $\Delta\theta_c(\{K_S\})$, and the inconsistency, $\sqrt{\langle\Delta\theta^2(K)\rangle_{K\in\{K_S\}}}$, is equivalent to the RMS distance between $\overline{Q}(K_S), K_S \in \{K_S\}$ and $\overline{Q}(\{K_S\})$, can be thought of as the disagreement, $\Delta\theta_a(W_A, W_B)$, between reconstructions W_A and W_B. This relation can be shown by

$$|\{K_S\}|\Delta\theta_c{}^2(\{K_S\}) - \sum_{K_S}\Delta\theta^2(K_S)$$

$$= \sum_{K_S}\sum_{\boldsymbol{Q}}P(\boldsymbol{Q}\,|\,K_S)\big(\boldsymbol{Q} - \overline{Q}(\{K_S\})\big)^2$$

$$\quad - \sum_{K_S}\sum_{\boldsymbol{Q}}P(\boldsymbol{Q}\,|\,K_S)\big(\boldsymbol{Q} - \overline{Q}(K_S)\big)^2$$

$$= \sum_{K_S}\sum_{\boldsymbol{Q}}P(\boldsymbol{Q}\,|\,K_S)\big(\boldsymbol{Q}^2 - 2\boldsymbol{Q}\overline{Q}(\{K_S\})+$$

$$\quad \overline{Q}^2(\{K_S\}) - \boldsymbol{Q}^2 + 2\boldsymbol{Q}\overline{Q}(K_S) - \overline{Q}^2(K_S)\big) \tag{27}$$

$$= \sum_{K_S}\sum_{\boldsymbol{Q}}P(\boldsymbol{Q}\,|\,K_S)\big(-2\overline{Q}(K_S)\overline{Q}(\{K_S\})+$$

$$\quad \overline{Q}^2(\{K_S\}) + 2\overline{Q}(K_S)\overline{Q}(K_S) - \overline{Q}^2(K_S)\big)$$

$$= \sum_{K_S}\sum_{\boldsymbol{Q}}P(\boldsymbol{Q}\,|\,K)\big(\overline{Q}(K_S) - \overline{Q}(\{K_S\})\big)^2$$

$$= \sum_{K_S}\big(\overline{Q}(K_S) - \overline{Q}(\{K_S\})\big)^2$$

$$\equiv \Delta\theta_a(W_A, W_B).$$

Above we use $\sqrt{\langle\Delta\theta^2(K)\rangle_{K\in\{K_S\}}}$ as the inconsistency in Eq. (27) instead of the definition in Eq. (1), because these two definitions are approximately the same if Gaussian distributions are assumed for OPDs, $P(\boldsymbol{Q}_i\,|\,K_S, W_i)$, $i = A, B$. As $P(\boldsymbol{Q}\,|\,K_S)$ is a convolution of these two Gaussian OPDs, its variance is $\Delta\theta^2(K_S) = \delta_A^2 + \delta_B^2$, where δ_A^2 and δ_B^2 are the variances of OPD_A and OPD_B. Meanwhile, the variances of auto-convolution of two OPDs are $\Theta^2(\overline{Q}_{ii} = 0\,|\,K_S, W_i) = 2\delta_i^2, i = A, B$, which gives us

$$\Delta\theta^2(K_S\,|\,W_A, W_B) \approx \frac{1}{2}\Theta^2(0\,|\,K_S, W_A) + \frac{1}{2}\Theta^2(0\,|\,K_S, W_B) = \Delta\theta_i^2(W_A, W_B). \tag{28}$$

The average of right hand side (RHS) of Eq. (28) over $\{K_S\}$ is consistent with RHS of Eq. (1).

The width of OPD, δ^2, quantifies how well we can identify the orientation for a given pattern. For a pixel at $\boldsymbol{q}$ in this pattern, we cannot decide whether this pixel belongs to a diffraction speckle near its most likely orientation if the speckle's radii $\theta_{sp}(\boldsymbol{q})$ is larger than δ. Strictly, if we want a 74% confidence interval, then we should have $\theta_{sp}(\boldsymbol{q}) \le 2\delta$. It should be noted that the confidence interval for 2σ is 74% instead of 95% since OPD is a 3D Gaussian distribution even though we simplified the derivation above with a 1D Gaussian distribution. The δ is computational expensive, but it can be easily inferred from $\Delta\theta_i$ by $\delta \approx \Delta\theta_i/\sqrt{2}$ if the Gaussian assumption discussed above is utilized. Moreover, being more cautious about the conclusion, we replace the $\Delta\theta_c$ instead of $\Delta\theta_i$ in Eq. (7).

Sentinel pattern coverage in the SO(3) orientation space.

Comparing a sentinel pattern to a diffraction intensity results in the former's OPD. This OPD covers a certain region in the SO(3) orientation space. The volume of this region should be proportional to the width of the OPD which could be estimated by $\Delta\theta_i/\sqrt{2}$ as mentioned in Eq. (28). If we crudely partitioned these OPDs with boxes whose average edge length is twice the average OPD width then the average volume covered by an OPD is $(2\Delta\theta_i/\sqrt{2})^3$. Given when the number of patterns diverges (the yellow asymptote) in Fig. 7, $\Delta\theta_i = 0.24$, then at least we need

$$\frac{\pi^2}{(2\times 0.24/\sqrt{2})^3} \approx 250 \tag{29}$$

OPDs to cover the whole SO(3) space, where π^2 is the total volume of SO(3).

References
 1. Spence, J. C. H. XFELs for structure and dynamics in biology. *IUCrJ* **4**(4), 322 (2017).
 2. Chapman, H. N. X-ray free-electron lasers for the structure and dynamics of macromolecules. *Annu. Rev. Biochem.* **88**, 35 (2019).
 3. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**(6797), 752–757 (2000).
 4. Jurek, Z., Faigel, G. & Tegze, M. Dynamics in a cluster under the influence of intense femtosecond hard X-ray pulses. *Eur. Phys. J. D* **29**(2), 217–229 (2004).
 5. Chapman, H. N. *et al.* Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nat. Phys.* **2**(12), 839–843 (2006).
 6. Yoon, C. H. *et al.* A comprehensive simulation framework for imaging single particles and biomolecules at the european X-ray free-electron laser. *Sci. Rep.* **6**, 24791 (2016).
 7. Fortmann-Grote, C. *et al.* SIMEX: Simulation of experiments at advanced light sources. *IUCrJ* **4**, 560–568 (2017).

8. Duane-Loh, N. T., Elser, V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. Stat. Nonlinear Soft Matter Phys.*, **80**, 26705 (2009).
9. Ayyer, K., Lan, T.-Y. & Elser, V. Dragonfly: An implementation of the expand–maximize–compress algorithm for single-particle imaging. *J. Appl. Crystallogr.* **49**(4), 1320–1335 (2016).
10. Kassemeyer, S. *et al.* Optimal mapping of X-ray laser diffraction patterns into three dimensions using routing algorithms. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **88**(4), 042710 (2013).
11. Yoon, C. H. *et al.* Unsupervised classification of single-particle X-ray diffraction snapshots by spectral clustering. *Opt. Express* **19**(17), 16542–16549 (2011).
12. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**(7332), 73–77 (2011).
13. Ekeberg, T. *et al.* Three-dimensional reconstruction of the giant mimivirus particle with an X-ray free-electron laser. *Phys. Rev. Lett.* **114**(9), 098102 (2015).
14. Loh, N. D. *et al.* Fractal morphology, imaging and mass spectrometry of single aerosol particles in flight. *Nature* **486**(7404), 513–517 (2012).
15. van der Schot, G. *et al.* Imaging single cells in a beam of live cyanobacteria with an X-ray laser. *Nat. Commun* **6**, 5704 (2015).
16. Hantke, M. F. *et al.* High-throughput imaging of heterogeneous cell organelles with an X-ray laser. *Nat. Photonics* **8**(12), 943–949 (2014).
17. Harauz, G. & van Heel, M. Exact filters for general geometry three dimensional reconstruction. *Optik* **73**(4), 146–156 (1986).
18. Rui, X. *et al.* Single-shot three-dimensional structure determination of nanocrystals with femtosecond X-ray free-electron laser pulses. *Nat. Commun.* **5**(1), 1–9 (2014).
19. Ayyer, K. *et al.* Low-signal limit of X-ray single particle diffractive imaging. *Opt. Express* **27**(26), 37816–37833 (2019).
20. Giewekemeyer, K. *et al.* Experimental 3D coherent diffractive imaging from photon-sparse random projections. *IUCrJ* **6**(3), 357–365 (2019).
21. Hosseinizadeh, A. . & Mashayekhi, G. . Conformational landscape of a virus by single-particle X-ray scattering. *Nat. Methods* **14**(9), 877–881 (2017).
22. Ikonnikova, K. A., Teslyuk, A. B., Bobkov, S. A., Zolotarev, S. I. & Ilyin, V. A. Reconstruction of 3D structure for nanoscale biological objects from experiments data on super-bright X-ray free electron lasers (XFELs): Dependence of the 3D resolution on the experiment parameters. *Procedia Comput. Sci.* **156**, 49–58 (2019).
23. Kim, S. S., Nepal, P., Saldin, D. K. & Yoon, C. H. Reconstruction of 3D Image of Nanorice Particle from Randomly Oriented Single-Shot Experimental Diffraction Patterns Using Angular Correlation Method. arXiv (2020). preprinted http://arXiv.org/10.1101/224402.
24. Nakano, M., Miyashita, O., Jonic, S., Tokuhisa, A. & Tama, F. Single-particle XFEL 3D reconstruction of ribosome-size particles based on Fourier slice matching: Requirements to reach subnanometer resolution. *J. Synchrot. Radiat.* **25**(4), 1010–1021 (2018).
25. Poudyal, I., Schmidt, M. & Schwander, P. Single-particle imaging by X-ray free-electron lasers—How many snapshots are needed?. *Struct. Dyn.* **7**(2), 024102 (2020).
26. Pryor, A. *et al.* Single-shot 3D coherent diffractive imaging of core-shell nanoparticles with elemental specificity. *Sci. Rep.* **8**(1), 8284 (2018).
27. Rose, M. *et al.* Single-particle imaging without symmetry constraints at an X-ray free-electron laser. *IUCrJ* **5**(6), 727–736 (2018).
28. Shi, Y. *et al.* Evaluation of the performance of classification algorithms for XFEL single-particle imaging data. *IUCrJ* **6**(2), 331–340 (2019).
29. von Ardenne, B., Mechelke, M. & Grubmüller, H. Structure determination from single molecule X-ray scattering with three photons per image. *Nat. Commun.* **9**(1), 9 (2018).
30. Liu, J., Engblom, S. & Nettelblad, C. Assessing uncertainties in X-ray single-particle three-dimensional reconstruction. *Phys. Rev. E* **98**, 013303 (2018).
31. van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *J. Struct. Biol.* **151**(3), 250–262 (2005).
32. Liao, H. Y. & Frank, J. Definition and estimation of resolution in single-particle reconstructions. *Structure* **18**(7), 768–775 (2010).
33. van Heel, M. & Schatz, M. Reassessing the revolution's resolutions. *bioRxiv* https://doi.org/10.1101/224402 (2017).
34. Tegze, M. & Bortel, G. Coherent diffraction imaging: Consistency of the assembled three-dimensional distribution. *Acta Crystallogr. A Found. Adv.* **72**(Pt 4), 459–464 (2016).
35. Elser, V. Noise limits on reconstructing diffraction signals from random tomographs. *IEEE Trans. Inf. Theory* **55**(10), 4715–4722 (2009).
36. Elser, V. & Eisebitt, S. Uniqueness transition in noisy phase retrieval. *New J. Phys.* **13**(2), 023001 (2011).
37. Jahn, T., Wilke, R. N., Chushkin, Y. & Salditt, T. How many photons are needed to reconstruct random objects in coherent X-ray diffractive imaging?. *Acta Crystallogr. A Found. Adv.* **73**(Pt 1), 19–29 (2017).
38. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Techn. J.* **27**(3), 379–423 (1948).
39. Loh, N. D. *et al.* Cryptotomography: Reconstructing 3D fourier intensities from randomly oriented single-shot diffraction patterns. *Phys. Rev. Lett.* **104**(22), 225501 (2010).
40. Bortel, G. & Tegze, M. Common arc method for diffraction pattern orientation. *Acta Crystallogr. A* **67**(6), 533–543 (2011).
41. Tegze, M. & Bortel, G. Selection and orientation of different particles in single particle imaging. *J. Struct. Biol.* **183**(3), 389–393 (2013).
42. Drinkwater, N. *et al.* Potent dual inhibitors of plasmodium falciparum m1 and m17 aminopeptidases through optimization of s1 pocket interactions. *Eur. J. Med. Chem.* **110**, 43–64 (2016).
43. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**(4), 721–745 (2003).
44. Shen, Q., Bazarov, I. & Thibault, P. Diffractive imaging of nonperiodic materials with future coherent X-ray sources. *J. Synchrotron. Radiat.* **11**(Pt 5), 432–438 (2004).
45. Giewekemeyer, K. *et al.* Experimental 3D coherent diffractive imaging from photon-sparse random projections. *IUCrJ* **6**(Pt 3), 357–365 (2019).
46. Sobolev, E. *et al.* Megahertz single-particle imaging at the European xfel. *Commun. Phys.* **3**(1), 97 (2020).

Acknowledgements

Author contributions

N.D.L. and Z.S. conceived the project. Z.S. performed all the calculations in this manuscript, with technical help from C.Z.W.T., K.A. and N.D.L. The manuscript was written by N.D.L and Z.S. with input from K.A.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-79589-0.

Correspondence and requests for materials should be addressed to N.D.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.