IUCrJ
ISSN 2052-2525
PHYSICS | FELS

Unsupervised learning approaches to characterizing heterogeneous samples using X-ray single-particle imaging

Yulong Zhuang,[a,b] Salah Awel,[c] Anton Barty,[c] Richard Bean,[d] Johan Bielecki,[d] Martin Bergemann,[d] Benedikt J. Daurer,[e] Tomas Ekeberg,[f] Armando D. Estillore,[c] Hans Fangohr,[a,b,d,r] Klaus Giewekemeyer,[d] Mark S. Hunter,[g] Mikhail Karnevskiy,[d] Richard A. Kirian,[h] Henry Kirkwood,[d] Yoonhee Kim,[d] Jayanath Koliyadu,[d] Holger Lange,[i,j] Romain Letrun,[d] Jannik Lübke,[c,i,k] Abhishek Mall,[a,b] Thomas Michelat,[d] Andrew J. Morgan,[l] Nils Roth,[c,k] Amit K. Samanta,[c] Tokushi Sato,[d] Zhou Shen,[e] Marcin Sikorski,[c] Florian Schulz,[i,s] John C. H. Spence,[h] Patrik Vagovic,[c,d] Tamme Wollweber,[a,b,i] Lena Worbs,[c,k] P. Lourdu Xavier,[a,c,i] Oleksandr Yefanov,[c] Filipe R. N. C. Maia,[f,m] Daniel A. Horke,[c,i,n] Jochen Küpper,[c,i,k,o] N. Duane Loh,[e,p] Adrian P. Mancuso,[d,q] Henry N. Chapman[c,i,k] and Kartik Ayyer[a,b,i]*

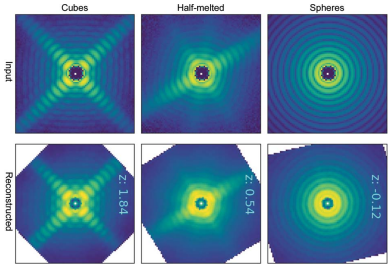
Received 15 September 2021
Accepted 30 November 2021

Edited by T. Ishikawa, Harima Institute, Japan

Keywords: coherent X-ray diffractive imaging (CXDI); single particles; XFELs.

[a]Max Planck Institute for the Structure and Dynamics of Matter, 22761 Hamburg, Germany, [b]Center for Free-Electron Laser Science, 22761 Hamburg, Germany, [c]Center for Free-Electron Laser Science CFEL, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany, [d]European XFEL, 22869 Schenefeld, Germany, [e]Center for Bio-Imaging Sciences, National University of Singapore, 117557, Singapore, [f]Department of Cell and Molecular Biology, Uppsala University, 75124 Uppsala, Sweden, [g]Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA, [h]Department of Physics, Arizona State University, Tempe, AZ 85287, USA, [i]Hamburg Center for Ultrafast Imaging, Universität Hamburg, 22761 Hamburg, Germany, [j]Institute of Physical Chemistry, Universität Hamburg, 20146 Hamburg, Germany, [k]Department of Physics, Universität Hamburg, 22761 Hamburg, Germany, [l]Department of Physics, University of Melbourne, Victoria 3010, Australia, [m]NERSC, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [n]Institute for Molecules and Materials, Radboud University, 6525 AJ Nijmegen, Netherlands, [o]Department of Chemistry, Universität Hamburg, 20146 Hamburg, Germany, [p]Department of Physics, National University of Singapore, 117551, Singapore, [q]Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia, [r]University of Southampton, Southampton SO17 1BJ, United Kingdom, and [s]Institute of Nanostructure and Solid State Physics, University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. *Correspondence e-mail: kartik.ayyer@mpsd.mpg.de

One of the outstanding analytical problems in X-ray single-particle imaging (SPI) is the classification of structural heterogeneity, which is especially difficult given the low signal-to-noise ratios of individual patterns and the fact that even identical objects can yield patterns that vary greatly when orientation is taken into consideration. Proposed here are two methods which explicitly account for this orientation-induced variation and can robustly determine the structural landscape of a sample ensemble. The first, termed common-line principal component analysis (PCA), provides a rough classification which is essentially parameter free and can be run automatically on any SPI dataset. The second method, utilizing variation auto-encoders (VAEs), can generate 3D structures of the objects at any point in the structural landscape. Both these methods are implemented in combination with the noise-tolerant expand–maximize–compress (*EMC*) algorithm and its utility is demonstrated by applying it to an experimental dataset from gold nanoparticles with only a few thousand photons per pattern. Both discrete structural classes and continuous deformations are recovered. These developments diverge from previous approaches of extracting reproducible subsets of patterns from a dataset and open up the possibility of moving beyond the study of homogeneous sample sets to addressing open questions on topics such as nanocrystal growth and dynamics, as well as phase transitions which have not been externally triggered.



1. Introduction

X-ray single-particle imaging (SPI) is a method to reconstruct 3D structures of isolated nanoscale objects by collecting a large number of diffraction patterns using bright X-ray pulses.

The diffraction patterns sample the object's 3D Fourier transform along randomly oriented spherical slices, which enables a Fourier synthesis of the 3D structure as long as the orientations can be determined (Neutze *et al.*, 2000). However, the X-ray pulses are bright enough to destroy the sample after each shot, and so each pattern is collected from a different particle. If the particles are reproducible up to the resolution of interest, then determination of the 3D structures is fairly straightforward, involving the determination of the orientation and incident X-ray fluence for each shot. Various algorithms have been proposed and implemented for this purpose, including the expand–maximize–compress (*EMC*) algorithm (Loh & Elser, 2009), which has been used for a number of experimental demonstrations (Ekeberg *et al.*, 2015; Rose *et al.*, 2018; Shi *et al.*, 2018; Ayyer *et al.*, 2019; Cho *et al.*, 2021; Ayyer *et al.*, 2021). Other methods utilizing intensity correlations have also been demonstrated experimentally (Kurta *et al.*, 2017; von Ardenne *et al.*, 2018).

However, one of the challenges in analysing a serial dataset like that produced in an SPI experiment is the proper classification of patterns in terms of their structures. This is a necessary step in order to obtain a high-resolution structure, since the underlying assumption of the above-mentioned algorithms is that the particles are identical, which is never true in practice. On the other hand, the framework used for this classification can be used to study datasets where the heterogeneity is not just a drawback, uncovering the landscape of structural variations in the sample. This requires the detection of discrete classes of object shapes representing contaminants, aggregates *etc.* and even the detection of continuous deformations, depending on the scientific problem being studied.

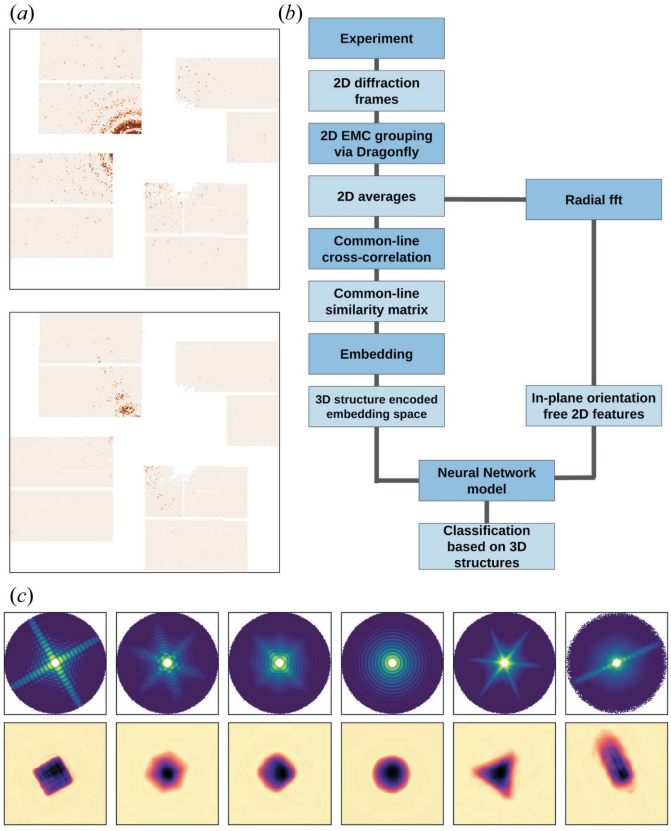


Figure 1
(*a*) Examples of diffraction patterns in the data set. The colour scale maximizes at four photons. (*b*) The workflow of the CLPCA method. (*c*) The upper row shows some typical frame averages in the dataset on a logarithmic scale, while the lower row shows their corresponding real-space 2D density projections via phase retrieval. The real-space field of view is 132.4 nm.

In past years, different machine learning algorithms have been developed for use in this classification of structural variability. These algorithms are usually applied to the patterns themselves using methods including spectral clustering (Yoon *et al.*, 2011), support vector machines (Bobkov *et al.*, 2015) and convolutional neural networks (Zimmermann *et al.*, 2019; Ignatenko *et al.*, 2021).

In this work, we take an unsupervised learning approach to analyse an experimental dataset of more than 2.4 million diffraction patterns of nominally 42 nm cubic gold nanoparticles collected at the European XFEL (Ayyer *et al.*, 2021). The goal is to study the 'structural landscape' – classifying the entire structural ensemble, including separating discrete contaminants, as well as revealing any continuous structural variations that may be present. We discuss two approaches, the first of which produces an embedding of the diffraction patterns based on their 3D structures. This approach can be applied without modification to most SPI datasets. The second is a generative method using variational auto-encoders (VAEs) which enables us to visualize the 3D structure of the particle at any point along its landscape. Both methods are applied to the dataset to study the continuous XFEL-induced deformation of the cubic nanoparticles to spheres.

2. Results

2.1. Experiment and dataset information

The dataset discussed in the rest of this work was collected as part of the experiment described by Ayyer *et al.* (2021), which we review in brief: The SPI experiment was performed with the megahertz-rate European XFEL (Decking *et al.*, 2020) on the SPB/SFX (single particles, biomolecules and clusters/serial femtosecond crystallography) instrument (Mancuso *et al.*, 2019) with 6 keV photons in pulses with an average energy of 2.5 mJ (2.6×10^{12} photons) measured upstream of the focusing optics. The adaptive gain integrating pixel detector (AGIPD) (Allahgholi *et al.*, 2019) was placed 705 mm downstream of the interaction region to collect each diffraction pattern up to a scattering angle of 8.3°. In this analysis, we will use data from the central region of the detector up to 1.8°.

The samples were nominally gold cubes with edge lengths 42 nm and were injected into the X-ray beam via an electrospray aerosolization aerodynamic lens-stack sample delivery system. For this sample, 34 197 950 frames were recorded, of which 2 451 068 were judged to contain sample diffraction (hit ratio ∼7.17%). Part of the dataset was collected with the European XFEL running at an intra-train repetition rate of

1.1 MHz, during which a high fraction of the diffraction patterns originated from spherical particles. This was found to be due to the pre-exposure of particles in the wings of the previous XFEL pulse in the train, which seemed to lead to melting/rounding. This effect disappeared when reducing the repetition rate to 0.5 MHz. As a result, the sample contains diffraction patterns of cubes, spheres and, potentially, cubes at different stages of melting/rounding.

2.2. Classification of the entire ensemble

There are two major challenges in understanding the sample ensemble from the diffraction data. First, many of the patterns are too weak (100–1000 photons per pattern) to apply single-shot analysis, as illustrated in Fig. 1(*a*). Second, the dataset is not structurally homogeneous, not just in that the sample set included non-cubic particles, but also that a fraction of the cubic particles suffered varied levels of melting due to pre-exposure by the previous pulse in the train.

We first developed a common-line principal component analysis (CLPCA) method to understand SPI datasets from an arbitrary ensemble of particles. Fig. 1(*b*) shows our workflow, which we discuss in detail below.

2.2.1. Generation of 2D averages via bootstrapping. Due to the weak scattering signal in single diffraction frames, the first step is to average similar 2D detector frames together (accounting for in-plane rotations), which was done with the 2D classification procedure implemented in *Dragonfly* (Ayyer *et al.*, 2016). The code implements a modification of the *EMC* algorithm that classifies all 2D frames into a given number of averages (models, termed classes in *Dragonfly*). This averaging improves the signal-to-noise ratio, corrects for incident fluence variations and fills in detector panel gaps, at the potential cost of washing out some structural variations. In order to mitigate this last effect, one can use a very large number of models, but this can lead to instabilities in the iterative reconstruction and reduced signal-to-noise ratios in individual class averages.

In order to get more pattern averages, a bootstrapping method was used by running the reconstruction five times with 200 models, each time with a random subset of 80% of the frames. In this way, each of the 1000 models are composed of a different group of similar 2D frames. In Fig. 1(*c*), we show an example of some typical 2D diffraction averages and corresponding projected electron-density maps after phase retrieval using a combination of the difference map and error reduction algorithms similar to that employed by Ayyer *et al.* (2019). This highlights the variety both in the samples and also in the diffraction patterns from the same samples in different orientations, such as the rotated versions of identical cubes in the first two columns of Fig. 1(*c*).

2.2.2. Common line 3D classification. The 2D *EMC* method can only group similar 2D frames together but, as mentioned above, diffraction patterns of the same object can look very different depending on the orientation. In order to understand the variations of structures in the sample, we need to classify the averages further through their 3D features, rather than considering them just as 2D images.

At small scattering angles, each average is a Fourier transform of a projection of the target object at a given orientation. According to the Fourier slice theorem, this means that each average represents a slice through the 3D Fourier transform of the object. Any two patterns of different orientations from the same object should share a common intersection line (at larger angles, these lines become arcs), as shown in the illustration of Fig. 2(*a*). The similarity of the diffraction intensities along the best 'common lines' between two patterns should be correlated to the similarity of their 3D structures. For each pair of 2D averages we calculated the cross correlation (CC) between their radial intensity profile lines at different angles. We define the common-line similarity between two 2D averages as the value of the highest CC coefficient. This yields a common-line similarity matrix (CC matrix) with a size of $N\times N$ for N 2D intensity averages. We find that the relation of the CC value of one average to all other averages acts as a signature of its 3D shape, namely that all particles with the same shape, regardless of orientation, have high values between each other and lower values to distinctly shaped objects.

Common lines have been used previously for orientation determination (Shneerson *et al.*, 2008; Singer & Shkolnisky, 2011; Yefanov & Vartanyants, 2013), where the optimal angles tell one how to fit the two slices in the 3D Fourier space. Here, we are interested in the similarity index of the common lines

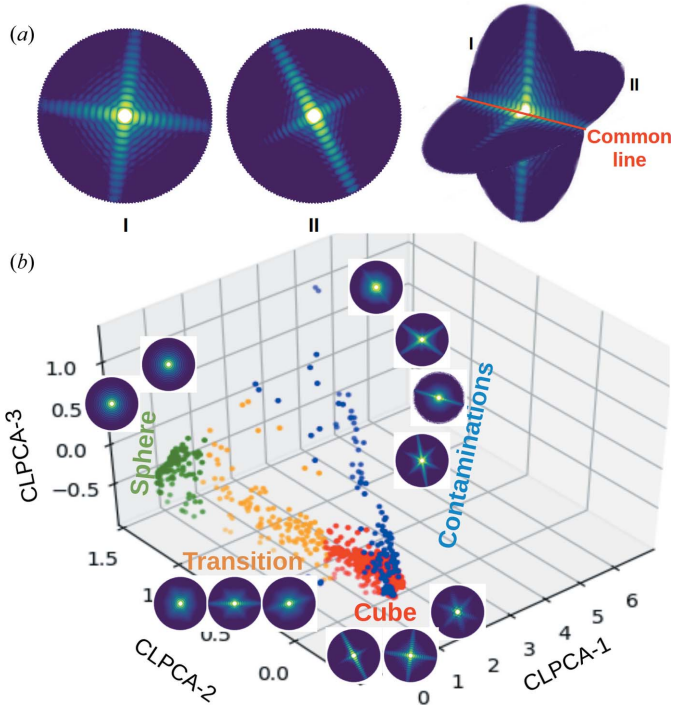


Figure 2
(*a*) An illustration of the common line between two patterns of similar-shaped objects but with different orientations. (*b*) The distribution of 1000 averages in the 3D CLPCA space. Different colours are manually divided groups from the first two components: contaminants (blue), cubes (red), transition (yellow) and spheres (green). Typical patterns for averages in each group are also shown on a logarithmic scale.

as a tool for structural similarity analysis, rather than the relative angles at which they are maximized.

The 3D shape distribution was visualized using principal component analysis (PCA), the inputs to which are the row vectors of the CC matrix. Fig. 2(*b*) shows a 3D embedding of the 1000 averages. In the rest of this article, we refer to this 3D embedding as 'CLPCA space'. Until this point, the entire procedure is fully unsupervised and can be run automatically for any SPI dataset.

By comparing with 2D patterns and their locations in CLPCA space, we then qualitatively divided the CLPCA space into four groups as shown in Fig. 2(*b*), where red denotes cubes, green spheres, blue contaminants and yellow rounded cubes. In the embedded space, CLPCA-1 separates these assorted patterns from those originating from spherical/cubic particles, CLPCA-2 separates spherical and cubic particles, and for spherical particles, CLPCA-3 is associated with their diameter (see Appendix *D*). In addition, we see a sequence where particles transition from cubes to spheres. This is a clear trace of the pre-melting processes observed in the experiment. Without using any *a priori* knowledge about the sample set, we are thus able to obtain a rough classification of the dataset according to their 3D structures.

We also note from Fig. 2(*b*) that the dense cluster of cube patterns are correctly identified as being from the same-shaped particles, even though the patterns themselves vary extensively at different orientations. For the sake of simplicity we only perform the embedding with the PCA method, but other dimensionality reduction methods could also be used interchangeably. In Appendix *A* we show a comparison between PCA and other embedding methods, showing no strong preference in terms of separating structural classes for this dataset.

2.2.3. Absolute embedding of images. The CLPCA method provides a way of classifying unknown sample frames according to their 3D structures. But the 'similarity' used for generating the landscape is a relative concept, *i.e.* the distributions in the embedded space depend on the sample one chooses. This limits the classification and comparison of certain subsamples. For example, one might want to look into some subgroup in detail while still being able to relate them to the whole-sample embedding, or generate new sets of averages with bootstrapping. In addition, the time complexity of calculating the similarity matrix scales as the square of the number of models, which imposes a computational hurdle to using too many models at once.

In order to get a quicker and more universal measure of frame features after obtaining the embedding of a typical set of averages, a neural network-based regression was used to map any previously unseen averages into the defined embedding space. Firstly, we extracted the relevant features of the patterns: for each of our 1000 averages we Fourier-transformed its azimuthal intensity variation at every radius and kept the absolute values to make the pattern invariant to in-plane rotations. We then kept frequency signals within the spatial resolution at each radius as the training features, as shown in Figs. 3(*a*)–3(*c*). We used the coordinates of the averages in CLPCA space as training labels, as shown in Fig. 3(*e*).

The neural network used for fitting the relation between the pattern of 2D averages and their coordinates in CLPCA space has four fully connected hidden layers with 512, 128, 64 and 32 nodes per layer, respectively. From the 1000 patterns used to calculate the similarity matrix, 800 were used to train the model, which was then validated with the rest of the 200 patterns with mean-square errors of 0.088, 0.059 and 0.042 for components 1, 2 and 3, respectively. With this model, we were quickly able to find the absolute embedding of any single 2D intensity model and zoom into arbitrary subsets of patterns while still retaining a reference to the full data set. An example is shown in Fig. 3(*d*) where we use CLPCA on a manually selected subset of averages in the 'cube' region of CLPCA space. The frames which contributed to these selected averages were classified using *Dragonfly*. The embedding of

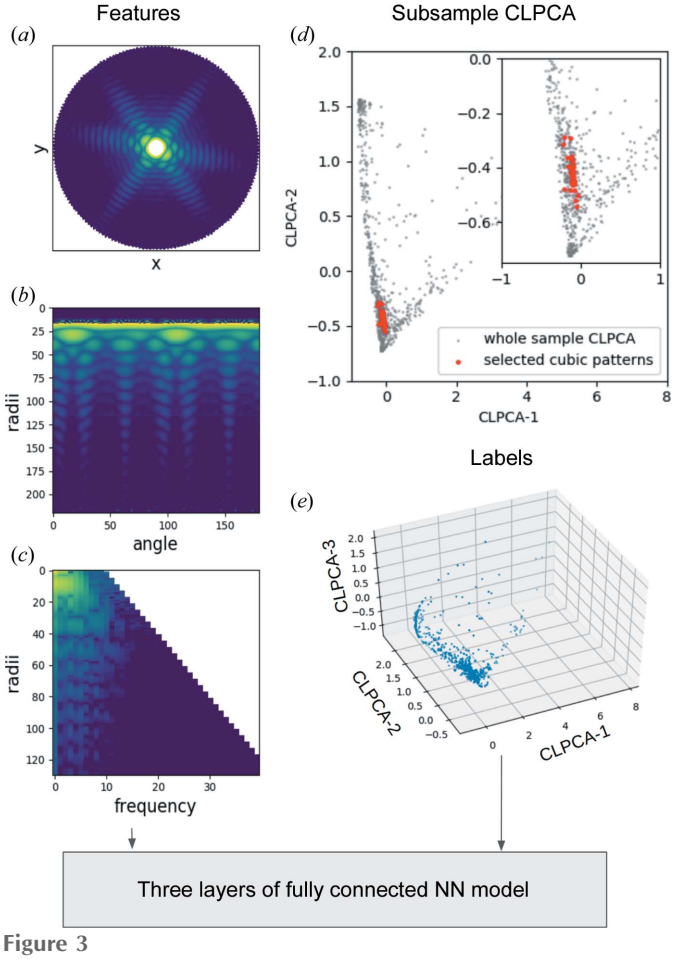


Figure 3
A brief illustration of the absolute embedding neural network (NN) model. (*a*) The average pattern from *Dragonfly*. (*b*) A polar representation of the pattern. (*c*) A stack of 1D Fourier transform magnitudes along the angular axis for each radial bin. The odd frequency components (due to inversion symmetry) and the higher frequencies for signals at smaller radii have been removed. These represent the feature vectors for the neural network. (*d*) An example of using absolute CLPCA on a selected cubic subset of frames (red dots). The grey dots represent the embedding of the pattern averages from the whole dataset. (*e*) Training labels from CLPCA.

these new averages is shown in red with reference to the whole-dataset CLPCA embedding.

In summary, the CLPCA method provides a robust and mostly parameter-free framework to classify and visualize the structural landscape of an arbitrary set of coherent diffraction patterns.

2.3. 3D reconstruction of heterogeneous models

In the previous section we found a sequence where the shape of the particles transitioned from cubic to spherical. To understand this transition further, we would like to be able to reconstruct the 3D models along the sequence. Given a set of diffraction patterns from a discrete set of reproducible objects, it is relatively straightforward to generate multiple 3D models using the *EMC* algorithm (Ayyer *et al.*, 2021; Cho *et al.*, 2021). Here, this approach is difficult for two reasons: (i) it needs to assume a number of discrete heterogeneity models, which does not qualitatively capture the continuous cube–sphere transition, and (ii) we do not have enough patterns located in the transition region to reconstruct 3D structures via *EMC*.

Fortunately, previous studies in cryo-electron microscopy single-particle analysis provide a good way of modelling the continuous heterogeneity. Zhong *et al.* (2021) developed *cryoDRGN*, a variational auto-encoder (VAE) for the efficient reconstruction of heterogeneous complexes and continuous trajectories of protein motions. Inspired by their paper, we developed a similar deep learning model by combining a VAE and convolutional neural networks (CNNs) to model the continuous shape transition along this 'melting' sequence.

2.3.1. Variational auto-encoders. Variational auto-encoders (VAEs) are a neural network architecture introduced by Kingma & Welling (2019), extensively used as generative networks and to understand the internal relationships of a dataset. They are a form of auto-encoder which are neural networks which attempt to recover the input data using a network with a bottleneck layer consisting of only a few neurons, representing the dimensionally reduced dataset. Fig. 4(*a*) shows the structure of our VAE neural network, consisting of a 2D CNN pattern encoder to encode 2D patterns, along with their orientation estimates, into distributions of latent parameters, and a 3D transposed convolution network as volume decoder to generate 3D intensity volumes from latent numbers. This setting allows the neural networks to learn the 3D-heterogeneity structure-encoded latent numbers from the diffraction patterns.

The inputs of the VAE network are the 2D intensity averages used as input for the CLPCA method in the previous section and their associated relative orientations. The former are obtained from the *Dragonfly* output discussed in Section 2.2.1, while the latter are calculated as described below. We start with the 3D intensity volume of an ideal 42 nm cubic particle, slicing it with 16 407 orientations distributed uniformly in quaternion space within the O_h subgroup to account for the symmetry of the object. This subgroup was chosen for this sample since we observed no symmetry breaking in either the single-model reconstruction in Ayyer *et al.* (2021) or in any of the 2D averages. The procedure following the choice of samples is identical even if there is no symmetry expected. The cross-correlation coefficient (CC) of each 2D model is calculated with all the orientations. The orientation with the highest CC is recorded for each model. In addition, another parameter which helps in training is termed 'gain', referring to the ratio of the best to the average CC over all orientations. This parameter can also indicate the 'cubicness' of the particle since patterns from cubes fit very well to the synthetic model.

The VAE works in the following manner: For each input average X, its orientation Ω and gain $\mathcal{G}$, the encoder network generates a Gaussian distribution of its latent variable values $N(\mu, \sigma)$. The fact that this is not just a single latent vector z but a distribution makes the auto-encoder variational and enforces the smoothness of the latent space. It also enables the network to use information from neighbouring regions in the space to update regions with limited data. A latent vector z is sampled using this distribution and used by the decoder to generate a 3D intensity volume V_{3D}, which is then symmetrized by the octahedral point group. The known orientation is then used to slice the generated 3D volume to get a reconstructed 2D pattern X'. The goal of training is to minimize the difference between X and X' (further details in Appendix *B*). Three-dimensional information is obtained since the same latent space region is sampled by averages in a variety of orientations.

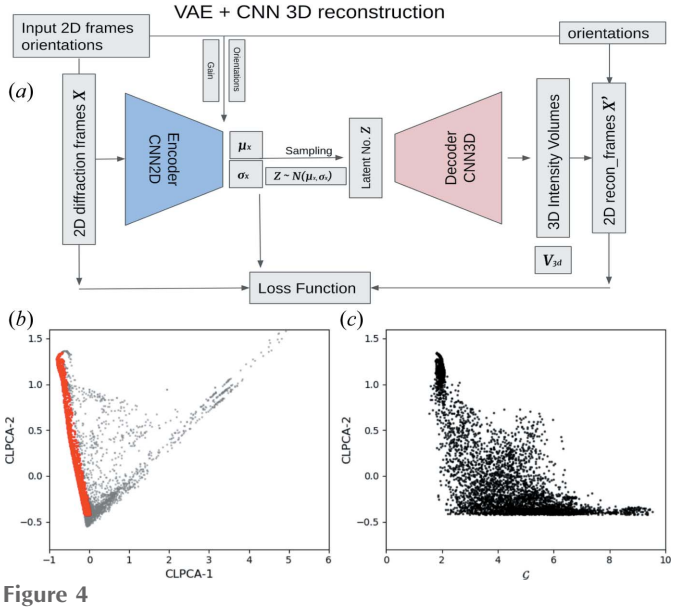


Figure 4
(*a*) A schematic diagram of the variational auto-encoder (VAE). It consists of two convolutional neural networks as encoder and decoder, and the model is fitted by comparing the similarity between the input and decoder-generated 2D slices with additional regularization by assuming the latent parameter follows an $N(0, 1)$ distribution. (*b*) The distribution of the 10 000 bootstrapped 2D average patterns in CLPCA space (grey). Red dots are selected average patterns along the melting sequence used for the VAE analysis. (*c*) CLPCA-2 plotted against gain for the selected patterns from panel (*b*) along the melting sequence.

Note that although the $\mathcal{G}$ is not strictly necessary for the VAE network to reconstruct 3D volumes, it helps to regularize the latent space of z.

When the model is trained, we can encode the 2D diffraction patterns into the latent space via the encoder network. The VAE network not only reconstructs the 3D structure of any given input 2D diffraction pattern but can also do so for any chosen location in the latent space. In this way, it can be used to study the continuous transition between cube and sphere in the melting sequence, including in regions where there are not sufficient patterns to isolate and obtain a conventional reconstruction.

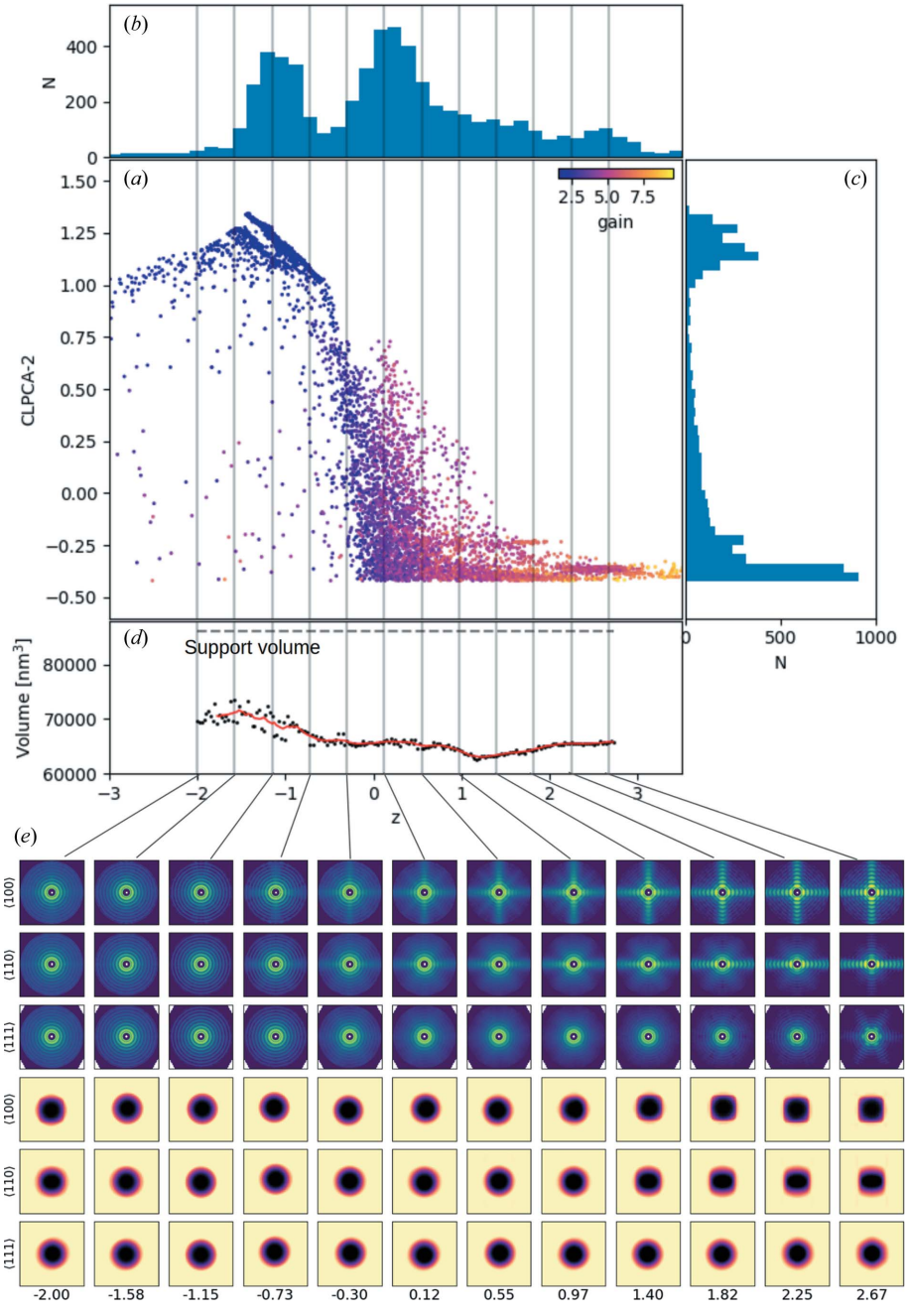


Figure 5
(*a*) VAE-encoded 1D latent number z plotted against CLPCA-2 for the input sample patterns. The colour coding is the gain parameter of each pattern. (*b*) A histogram of latent number z. (*c*) A histogram of CLPCA-2. (*d*) The volume 'evolution' along z, with volumes calculated with a density threshold equal to 10^{-4} of the total mass. The dashed horizontal line is the size of the support volume in phase retrieval. Vertical grey lines show the locations of the 12 selected z numbers. (*e*) The top three rows are the VAE-generated intensity volumes from the 12 selected z numbers on a logarithmic scale, the three rows showing slices in (from row 1 to row 3) the ⟨100⟩, ⟨110⟩ and ⟨111⟩ directions, respectively. The bottom three rows show their corresponding density projections in real space.

2.3.2. Tracing the melting sequence. As mentioned before, using *Dragonfly* to generate 2D averages comes at the cost of potentially averaging out structural variations in individual frames due to the limited number of classes. To generate the CLPCA landscape, 1000 averages were deemed to be enough to cover most of the major shapes in the sample, but in order to trace the continuous shape variation along the melting sequence more detailed minor variation information needs to be retained. To keep more of this variation information along the melting sequence we repeat the same bootstrapping plus *Dragonfly* method as described in Section 2.2.1 to generate 10 000 average patterns.

Using the CLPCA method in combination with the absolute embedding approach discussed in Section 2.2.3, we selected 5965 of these averages located along the melting sequence. Fig. 4(*b*) shows the distribution of these averages and the selected melting sequence averages in CLPCA space. As discussed above, the CLPCA-2 is roughly aligned with the melting sequence, which can be used as a coarse estimator of the melting process. Fig. 4(*c*) shows the comparison between CLPCA-2 and the gain $\mathcal{G}$ for the selected melting sequence models. Although correlated with the melting sequence at a certain level, by visual inspection $\mathcal{G}$ seems to be good at tracing the cubes, while CLPCA-2 is better at isolating the spheres.

In order to reduce computational cost, the input 2D average intensities were downsampled and truncated slightly at high q. Since the intensity distribution of a compact object is heavily weighted to low q, the input data were normalized by the azimuthally averaged intensity over the whole dataset. Details of the VAE network structure and various pre-processing steps are given in Appendix *B*.

For simplicity, we start by modelling with a 1D latent space, so the VAE

model is forced to trace the strongest variation along the melting sequence/cubic–spherical transition. Fig. 5(*a*) shows the VAE-encoded 1D latent parameter z plotted against CLPCA-2 for the input models, colour coded by the $\mathcal{G}$ parameter: brighter yellow indicates that the particles are more anisotropic. Figs. 5(*b*) and 5(*c*) are the histograms of z and CLPCA-2, respectively. Though there exists a strong correlation between the latent parameter z and CLPCA-2, they show different distributions in tracing the cubic–spherical transition. First, the VAE further separates the cubic side cluster classified by CLPCA (CLPCA-2 < -0.25). Secondly, at lower than $z \simeq -1.5$ another sequence is visible which is not clearly seen in CLPCA space.

One advantage of the VAE network is that it allows us to reconstruct/generate the 3D intensity volume for any given z. Here, we selected 12 z numbers of equal steps between −2.00 and 2.67. In Fig. 5(*e*), the upper three rows show three special (⟨100⟩, ⟨110⟩ and ⟨111⟩) slices of the 12 VAE-generated intensity volumes on a logarithmic scale. We see a clear and smooth cubic–spherical transition from $z \simeq 2.7 \rightarrow -1.5$, which shows that the VAE network is able to trace the early melting stages much better than the CLPCA. For the second 'sequence' at $z < 1.5$, the lower contrast in the intensities and the less-spherical structures indicate they are likely to be contaminated particles at late melting stages, the low contrast being due to the fact that they contain various shapes and a lack of accurate orientation estimates. Compared with the CLPCA method that roughly classifies cubic and spherical particles, the VAE is able to provide a detailed smooth modelling of the entire transition process.

The bottom three rows of Fig. 5(*e*) show projections of phase-retrieved density maps in real space along the same three directions. We also find the volume increases by around 10% from cubes to spheres along the melting sequence by calculating the volume evolution of 3D reconstructed volumes at 200 different finely sampled z values, shown as black dots in Fig. 5(*d*). This is consistent with the density difference between crystalline gold (19.32 g cm^{-3}) and molten/randomly close-packed atoms (17.31 g cm^{-3}). This is also in agreement with the direct size fitting on 2D average patterns of spherical and cubic particles, shown in Fig. 9.

This volume analysis allows us to observe an additional feature, namely that the melting sequence seems to show two different stages: Stage 1 is from $z \simeq 3 \rightarrow -0.3$, where most of the shape change occurs, but with no significant volume changes. Stage 2 is from $z \simeq -0.3 \rightarrow -1.5$, where all particles are mostly round but the volume increases. This is to be expected, since edges and vertices require lower energies to be disrupted than for complete melting, where the bulk crystal structure is lost (Cahn, 1986; Chen *et al.*, 2021). As discussed below, the melting sequence here is only based on the 3D structure of the particles. Thus, the 'two stages' might only suggest that the shape changing happened 'before' the size expansion along z (exposure intensity).

We note that the melting sequence here is purely based on the 3D structure of the particles. Since we observe a continuous transition, we assume a monotonic relationship between the latent variable z and the incident fluence in the previous pulse. However, without additional information about the expected distribution of particles with a given incident fluence, we cannot obtain the precise mapping between the two. As all pre-exposures happened 880 ns before the imaging pulse, one should not view the melting sequence as a time-dependent shape variation, but rather an incident fluence- or temperature-dependent behaviour.

We have shown that the 1D latent space VAE network is capable of modelling the major transition, but it is not the only variation in the dataset. To trace more variations we now model it with a higher-dimensional latent space. Fig. 6(*a*) shows the 2D VAE-encoded latent parameters of the sample. Neither the first nor the second latent parameter traces the cubic–spherical transition independently. This shows the encoding of more secondary variations, *e.g.* contaminants, size *etc*.

We selected 200 evenly distributed points in the latent space and generated their 3D intensity volumes. Fig. 6(*b*) shows the ⟨100⟩ direction projection maps of the phase-retrieved VAE-generated intensity volumes of the 200 selected latent numbers z. One can clearly see that the transition between spherical and cubic features from right to left is roughly encoded in the first z component. However, the trend is twisted in 2D space and we obtained multiple transition sequences along the second component. In regions with no or

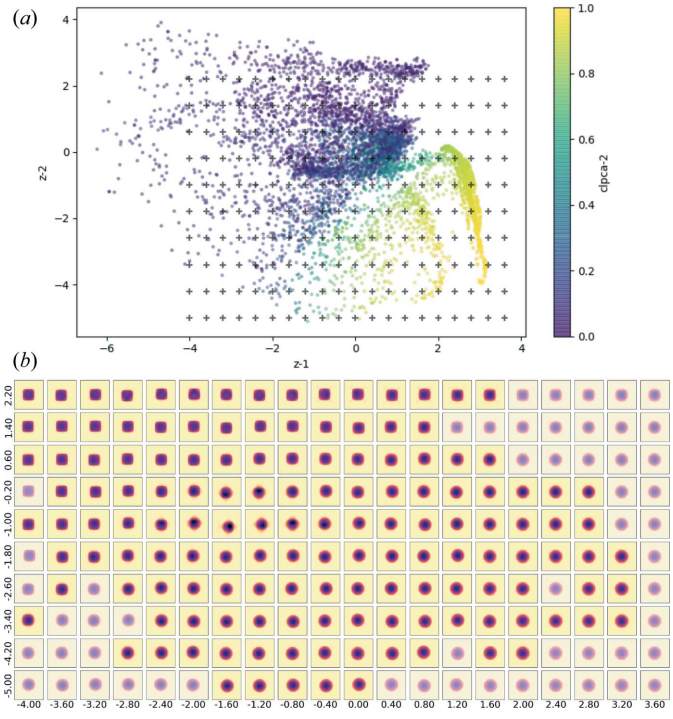


Figure 6
(*a*) The distribution of the dataset in 2D latent space, colour coded by CLPCA-2. Black crosses in the plot are the 200 selected latent numbers z. (*b*) The ⟨100⟩ direction projections of the 200 real-space densities retrieved from the logarithmic intensity volumes generated from the 200 selected latent numbers z [black crosses in panel (*a*)]. Semi-transparent slices are volumes generated from latent regions with no or very few data points.

very few data points (semi-transparent slices), the results are less reliable as the VAE network could not generate reliable intensity volumes.

3. Conclusions

In this work, we have demonstrated two methods to study heterogeneous ensembles of samples using X-ray single-particle imaging by analysing the relationships between the 3D structures of the samples rather than the diffraction patterns directly. The first, termed CLPCA, uses the fact that diffraction patterns from ideal particles share a common line/arc from the Ewald construction. The similarity of the best common lines allows us to visualize a low-dimensional embedding of the dataset which roughly represents 3D similarity regardless of orientation. A fully connected neural network was trained against the embedding in order to embed arbitrary 2D averages with respect to the rest of the data set.

This was applied to an experimental dataset from gold nanoparticles, containing 2.4 million patterns with only a few thousand photons per frame. In order to improve the signal-to-noise ratio and fill in detector panel gaps, 2D classification in *Dragonfly* was used to combine similar patterns up to in-plane rotations, and a bootstrapping approach was used to increase the number of 2D averages with subsets of the data.

The second method involves a generative VAE network that models continuous structural deformations, and can be used to generate 3D structures at arbitrary points along the landscape. In the studied dataset, this network was able to recover a continuous melting sequence induced by pre-exposure of the particles to the previous pulse in the European XFEL pulse train. At low fluences, we observed a rounding of the particle shape, while at higher fluences we additionally observed a volume expansion, probably caused by atomic scale disordering at higher temperatures. Our two methods open up the possibility of studying samples of heterogeneous particles and, potentially, tracing the dynamic motion of particles in SPI experiments.

APPENDIX *A*
Different embeddings of CC matrix

In Fig. 7 we show a comparison between PCA and other standard dimensional reduction methods implemented in the *scikit-learn* Python package (Pedregosa *et al.*, 2011). For the sample set considered, all four embeddings are able to separate the four basic shape groups (cube, sphere, transition and contaminant), although this may not be true for other, more complex, continuous structural variations.

APPENDIX *B*
Details of the VAE model

The details of the VAE network and preprocessing steps are described below. The code is available at https://github.com/AyyerLab/SPIEncoder.

B1. Preprocessing the input patterns

The 2D averages from *Dragonfly* have an initial size of 441×441 pixels, which is computationally redundant considering the data are highly oversampled. For the results shown in this paper, we used the following procedures to reduce the size. We first downsampled the size to 243×243, and then we cut off the high-q part to reduce the size further to 161×161.

In addition to the size reduction, diffraction patterns of compactly supported objects are overwhelmingly dominated by the low-q signal. To weight the higher-q shape information appropriately, we divided the 2D intensities by the radial average intensity over the whole dataset before feeding them into the VAE, and multiplied it back when generating the 3D intensities.

B2. Model parameters

The depth and nodes of our encoder/decoder network depend on the quality of the sample and the spatial resolution of the reconstructed 3D volumes we need. In our code we provide three different reconstructed volume size options: low volume (81^3), intermediate volume (161^3) and high volume (243^3). The loss function contains the mean-square error plus binary cross entropy and a Kullback–Leibler divergence as the regularization term. For the results shown in this paper, the low-volume model is used, in which the encoder network has four 2D convolution layers with (16×81^2), (32×27^2), (64×9^2) and (128×9^2) nodes and (3×3) kernel size per layer. The decoder consists of six 3D transposed convolution layers with (128×9^3), (64×9^3), (32×27^3), (16×81^3),

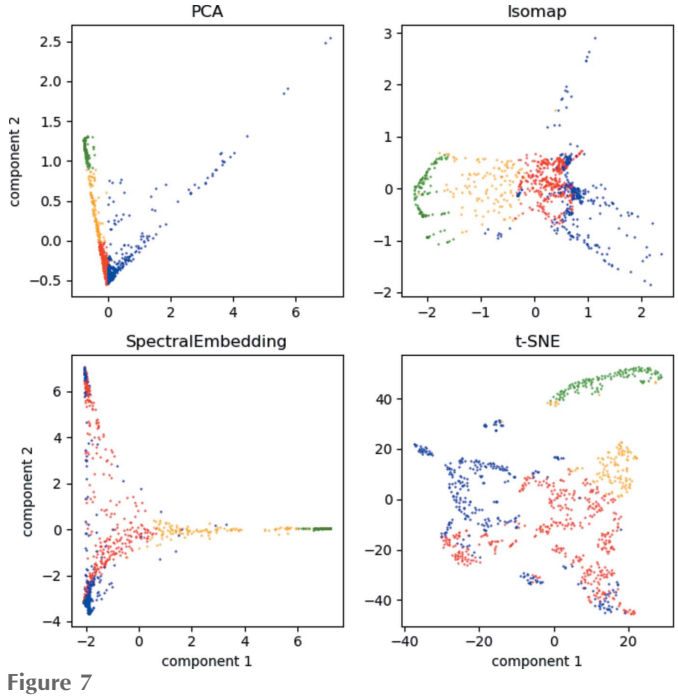


Figure 7
Comparison between four different common dimensionality reduction methods, (top left) PCA, (top right) isomap, (bottom left) spectral embedding and (bottom right) t-SNE. Colour coding is same as in Fig. 2 by manually divided groups from the first two components of PCA space.

(8×161^3) and (1×161^3) nodes and ($3 \times 3 \times 3$) kernel size per layer.

In Fig. 8 we show a comparison of input 2D patterns X and the corresponding 2D slices of their reconstructed 3D intensities X' of our VAE model at four representative latent z values from the model shown in Section 2.3.2. Considered together with Fig. 5, the first pattern is probably from half-melted contaminant particles, where the reconstruction is not reliable because the number of patterns of that shape is low and the orientation estimate by fitting against a cube is inaccurate. The other rows are well reconstructed spheres, melting cubes and cubes.

B3. Computation devices

The network was implemented in *PyTorch* and run on a single node with four NVidia V100 GPUs with 32 GB memory each. For our 5672 2D averages, the 161^3 volume model takes less than 2 h to calculate.

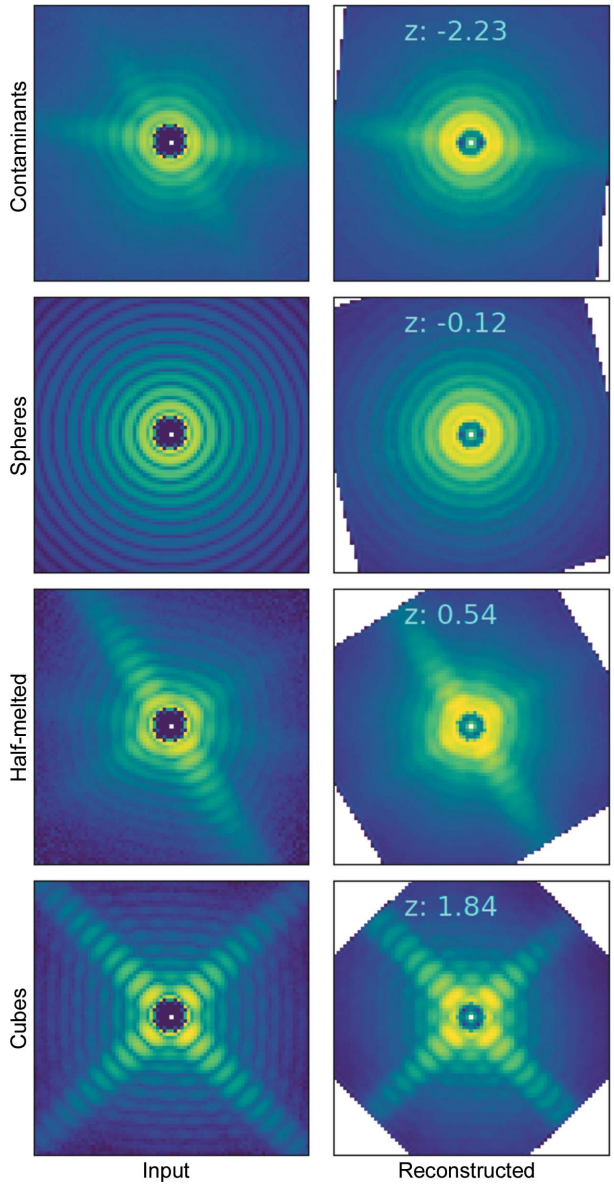


Figure 8
The left-hand column shows input intensities X to our VAE model, and the right-hand column shows VAE-reconstructed intensity patterns X' on a logarithmic scale.

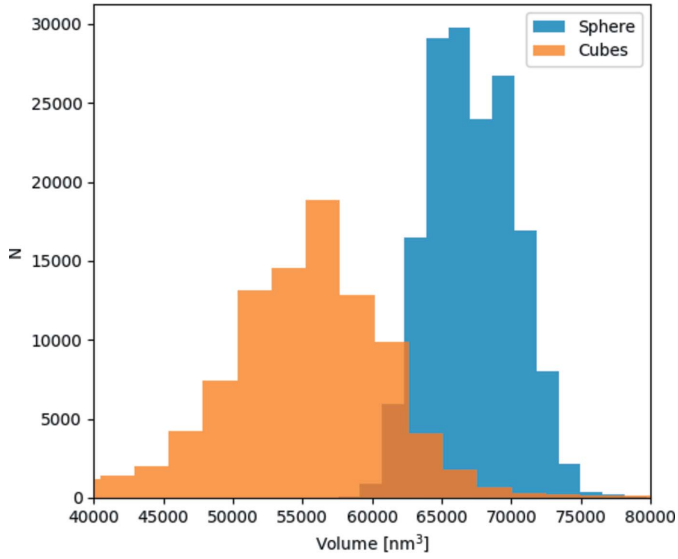


Figure 9
Volume distribution of cubic particles (orange) and spherical particles (blue).

APPENDIX C
Direct size comparison of cubes and spheres

Fig. 9 shows the volume comparison of classified pure cubes and pure spheres. The cube sizes were calculated directly from frames with clear streak fringes, which were assumed to have at least one set of faces parallel to the X-ray beam. The larger spread of cube sizes can be accounted for by the inclusion of frames where the faces were not exactly parallel to the beam. Nevertheless, similar to the VAE model, a clear size increase is observed from cubes to spheres.

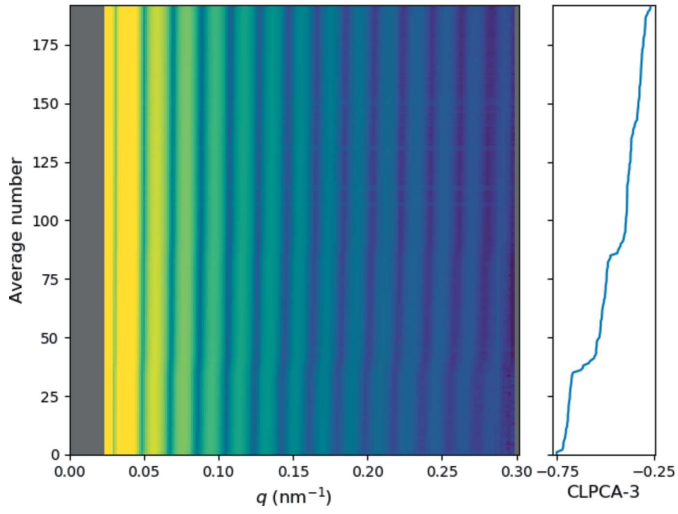


Figure 10
Correlation of radial intensity distribution with the CLPCA-3 component for the spherical 2D averages. The left-hand panel shows radial intensities versus q for 192 2D class averages. The right-hand panel shows the corresponding CLPCA-3 values. One can observe the stretching of the intensity distributions for higher values of CLPCA-3.

APPENDIX *D*
Correlation of CLPCA-3 with sphere diameter

By observation, we note that the CLPCA-3 component in Fig. 2 is associated with the diameter of the spherical particles highlighted in green. To validate this observation, we examined the radial intensity distributions for these spherical classes versus their CLPCA-3 values. This is shown in Fig. 10 for 192 2D averages in that region of CLPCA space.

Acknowledgements

We acknowledge the European XFEL in Schenefeld, Germany, for provision of X-ray free-electron laser beamtime on Scientific Instrument SPB/SFX (Single Particles, Clusters and Biomolecules, and Serial Femtosecond Crystallography) and would like to thank the staff for their assistance.

Funding information

This work has been supported by the Clusters of Excellence 'Center for Ultrafast Imaging' (CUI, EXC 1074, ID 194651731) and 'Advanced Imaging of Matter' (AIM, EXC 2056, ID 390715994) of the Deutsche Forschungsgemeinschaft (DFG). This work has also been supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013) through the Consolidator Grant COMOTION (grant No. ERC-614507-Küpper) and by the Helmholtz Gemeinschaft through the 'Impuls- und Vernetzungsfond'. John C. H. Spence and Richard A. Kirian acknowledge support from the US National Science Foundation (BioXFEL award No. STC-1231306). Filipe R. N. C. Maia acknowledges support from the Swedish Research Council, Röntgen–Ångström Cluster and Carl Tryggers Foundation for Scientific Research. P. Lourdu Xavier acknowledges a fellowship from the Joachim Herz Stiftung, Germany. P. Lourdu Xavier and Henry N. Chapman acknowledge support from the Human Frontiers Science Program, France (grant No. RGP0010/2017).

References

Allahgholi, A., Becker, J., Delfs, A., Dinapoli, R., Goettlicher, P., Greiffenberg, D., Henrich, B., Hirsemann, H., Kuhn, M., Klanner, R., Klyuev, A., Krueger, H., Lange, S., Laurus, T., Marras, A., Mezza, D., Mozzanica, A., Niemann, M., Poehlsen, J., Schwandt, J., Sheviakov, I., Shi, X., Smoljanin, S., Steffen, L., Sztuk-Dambietz, J., Trunk, U., Xia, Q., Zeribi, M., Zhang, J., Zimmer, M., Schmitt, B. & Graafsma, H. (2019). *J. Synchrotron Rad.* **26**, 74–82.

Ardenne, B. von, Mechelke, M. & Grubmüller, H. (2018). *Nat. Commun.* **9**, 2375.

Ayyer, K., Lan, T.-Y., Elser, V. & Loh, N. D. (2016). *J. Appl. Cryst.* **49**, 1320–1335.

Ayyer, K., Morgan, A. J., Aquila, A., DeMirci, H., Hogue, B. G., Kirian, R. A., Xavier, P. L., Yoon, C. H., Chapman, H. N. & Barty, A. (2019). *Opt. Express*, **27**, 37816.

Ayyer, K., Xavier, P. L., Bielecki, J., Shen, Z., Daurer, B. J., Samanta, A. K., Awel, S., Bean, R., Barty, A., Bergemann, M., Ekeberg, T., Estillore, A. D., Fangohr, H., Giewekemeyer, K., Hunter, M. S., Karnevskiy, M., Kirian, R. A., Kirkwood, H., Kim, Y., Koliyadu, J., Lange, H., Letrun, R., Lübke, J., Michelat, T., Morgan, A. J., Roth, N., Sato, T., Sikorski, M., Schulz, F., Spence, J. C. H., Vagovic, P., Wollweber, T., Worbs, L., Yefanov, O., Zhuang, Y., Maia, F. R. N. C., Horke, D. A., Küpper, J., Loh, N. D., Mancuso, A. P. & Chapman, H. N. (2021). *Optica*, **8**, 15–23.

Bobkov, S. A., Teslyuk, A. B., Kurta, R. P., Gorobtsov, O. Yu., Yefanov, O. M., Ilyin, V. A., Senin, R. A. & Vartanyants, I. A. (2015). *J. Synchrotron Rad.* **22**, 1345–1352.

Cahn, R. W. (1986). *Nature*, **323**, 668–669.

Chen, J., Fan, X., Liu, J., Gu, C., Shi, Y., Zheng, W. & Singh, D. J. (2021). *J. Phys. Chem. Lett.* **12**, 8170–8177.

Cho, D. H., Shen, Z., Ihm, Y., Wi, D. H., Jung, C., Nam, D., Kim, S., Park, S.-Y., Kim, K. S., Sung, D., Lee, H., Shin, J.-Y., Hwang, J., Lee, S. Y., Lee, S. Y., Han, S. W., Noh, D. Y., Loh, N. D. & Song, C. (2021). *ACS Nano*, **15**, 4066–4076.

Decking, W., Abeghyan, S., Abramian, P., *et al.* (2020). *Nat. Photon.* **14**, 391–397.

Ekeberg, T., Svenda, M., Abergel, C., Maia, F. R., Seltzer, V., Claverie, J.-M., Hantke, M., Jönsson, O., Nettelblad, C., van der Schot, G., Liang, M., DePonte, D. P., Barty, A., Seibert, M. M., Iwan, B., Andersson, I., Loh, N. D., Martin, A. V., Chapman, H., Bostedt, C., Bozek, J. D., Ferguson, K. R., Krzywinski, J., Epp, S. W., Rolles, D., Rudenko, A., Hartmann, R., Kimmel, N. & Hajdu, J. (2015). *Phys. Rev. Lett.* **114**, 098102.

Ignatenko, A., Assalauova, D., Bobkov, S. A., Gelisio, L., Teslyuk, A. B., Ilyin, V. A. & Vartanyants, I. A. (2021). *Mach. Learn. Sci. Technol.* **2**, 025014.

Kingma, D. P. & Welling, M. (2019). *Found. Trends Mach. Learn.* **12**, 307–392.

Kurta, R. P., Donatelli, J. J., Yoon, C. H., Berntsen, P., Bielecki, J., Daurer, B. J., DeMirci, H., Fromme, P., Hantke, M. F., Maia, F. R., Munke, A., Nettelblad, C., Pande, K., Reddy, H. K., Sellberg, J. A., Sierra, R. G., Svenda, M., van der Schot, G., Vartanyants, I. A., Williams, G. J., Xavier, P. L., Aquila, A., Zwart, P. H. & Mancuso, A. P. (2017). *Phys. Rev. Lett.* **119**, 158102.

Loh, N. D. & Elser, V. (2009). *Phys. Rev. E*, **80**, 026705.

Mancuso, A. P., Aquila, A., Batchelor, L., Bean, R. J., Bielecki, J., Borchers, G., Doerner, K., Giewekemeyer, K., Graceffa, R., Kelsey, O. D., Kim, Y., Kirkwood, H. J., Legrand, A., Letrun, R., Manning, B., Lopez Morillo, L., Messerschmidt, M., Mills, G., Raabe, S., Reimers, N., Round, A., Sato, T., Schulz, J., Signe Takem, C., Sikorski, M., Stern, S., Thute, P., Vagovič, P., Weinhausen, B. & Tschentscher, T. (2019). *J. Synchrotron Rad.* **26**, 660–676.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature*, **406**, 752–757.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. (2011). *J. Mach. Learn. Res.* **12**, 2825–2830.

Rose, B. C., Huang, D., Zhang, Z.-H., Stevenson, P., Tyryshkin, A. M., Sangtawesin, S., Srinivasan, S., Loudin, L., Markham, M. L., Edmonds, A. M., Twitchen, D. J., Lyon, S. A. & de Leon, N. P. (2018). *Science*, **361**, 60–63.

Shi, X., Levine, E., Weber, H. & Hargreaves, B. A. (2019). *Magn. Reson. Med.* **81**, 2247–2263.

Shneerson, V. L., Ourmazd, A. & Saldin, D. K. (2008). *Acta Cryst.* A**64**, 303–315.

Singer, A. & Shkolnisky, Y. (2011). *SIAM J. Imaging Sci.* **4**, 543–572.

Yefanov, O. & Vartanyants, I. (2013). *J. Phys. B At. Mol. Opt. Phys.* **46**, 164013.

Yoon, C. H., Schwander, P., Abergel, C., Andersson, I., Andreasson, J., Aquila, A., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Bozek, J., Chapman, H. N., Claverie, J., Coppola, N., DePonte, D. P., Ekeberg, T., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, E., Hartmann, R., Hauser, G., Hirsemann, H., Holl, P., Kassemeyer, S., Kimmel, N., Kiskinova, M., Liang, M., Loh, N. D., Lomb, L., Maia, F. R. N. C., Martin, A. V., Nass, K., Pedersoli, E., Reich, C., Rolles, D., Rudek, B., Rudenko, A., Schlichting, I., Schulz, J., Seibert, M., Seltzer, V., Shoeman, R. L., Sierra, R. G., Soltau, H., Starodub, D., Steinbrener, J., Stier, G.,

Strüder, L., Svenda, M., Ullrich, J., Weidenspointner, G., White, T. A., Wunderer, C. & Ourmazd, A. (2011). *Opt. Express*, **19**, 16542–16549.
Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. (2021). *Nat. Methods*, **18**, 176–185.
Zimmermann, J., Langbehn, B., Cucini, R., Di Fraia, M., Finetti, P., LaForge, A. C., Nishiyama, T., Ovcharenko, Y., Piseri, P., Plekan, O., Prince, K. C., Stienkemeier, F., Ueda, K., Callegari, C., Möller, T. & Rupp, D. (2019). *Phys. Rev. E*, **99**, 063309.