An artificial intelligence's interpretation of complex high-resolution *in situ* transmission electron microscopy data

Xingzhi Wang[a,b], Chang Yan[a,b,c], Justin C. Ondry[a,d,e], Viraj Bodiwala[f], Peter Ercius[g], A. Paul Alivisatos[a,b,h,i,*]

[a]*Department of Chemistry, University of California, Berkeley, Berkeley, California, 94720, United States*
[b]*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, United States*
[c]*Department of Chemistry and Center for Ultrafast Science and Technology, School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China*
[d]*Kavli Energy NanoScience Institute, Berkeley, California, 94720, United States*
[e]*James Franck Institute, Chicago, Illinois, 60637, United States*
[f]*Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois, 60637, United States*
[g]*National Center for Electron Microscopy, The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, United States*
[h]*Department of Materials Science and Engineering, University of California, Berkeley, Berkeley, California, 94720, United States*
[i]*Department of Chemistry, The University of Chicago, Chicago, Illinois, 60637, United States*

Abstract

Complicated nano- and atomic-scale processes with sub-Angstrom spatial resolution and millisecond time resolution visualized by *in situ* transmission electron microscopy (TEM) are often highly dynamical and time-consuming to analyze and interpret. Here, we report how variational autoencoders (VAEs) can provide an artificial intelligence's interpretation of high-resolution in situ TEM data by condensing and deconvoluting complicated atomic-scale dynamics into a latent space with reduced dimensionality. We designed a VAEs model with high latent dimensions capable of deconvoluting information from complex high-resolution TEM data. We demonstrate how this model with high latent dimensions trained on atomically resolved TEM images of lead sulfide (PbS) nanocrystals is able to capture movements and perturbations of periodic lattices in both simulated and real in situ TEM data. The VAEs model shows capability of detecting and deconvoluting dynamical nanoscale physical processes, such as the rotation of crystal lattices and intraparticle ripening during the annealing of semiconductor nanocrystals.

*Corresponding author and lead contact
Email address: paul.alivisatos@uchicago.edu (A. Paul Alivisatos)

1. Introduction

Transmission electron microscopy (TEM) has been a standard method for the characterization of nanomaterials, allowing researchers to directly visualize nanoscale structures and processes.[1,2] TEM imaging is powerful, yet conclusions drawn from it must be carefully drawn due to issues relating to small samples sizes, selection bias, and in some cases difficulty quantifying complex features.[2–5] Introduction of automated TEM data acquisition has drastically improved the amount of data that can be collected for a sample in a reasonable amount of time.[2–10] Further, advancement in techniques such as *in situ* imaging[5,11–13] and aberration correction[14,15] allows for the acquisition of TEM data with high spatial and/or temporal resolution with relative ease. Enabled by such enhancements, it may be possible to achieve an autonmous electron microscopy workflow to analyze, or even fabricate, the structure of nanomaterials with atomic precision on a statistical scale[8,16].

However, a significant challenge is the development of data analysis protocols to elucidate the desired information from sufficiently large sample set in a time-efficient manner. High-resolution TEM images contain a high density of multiplexed information which typically requires meticulous effort from a human expert, either through manual analysis or development of custom written analysis software, to elucidate statically significant results, as exemplified by many previous works.[17–21] Such a workflow for data analysis can be time consuming to implement, especially when the dataset is large and the desired information is convoluted. In a recent work by Sainju et al.[17], it took three researchers 20 weeks to manually analyze and label 1200 time-correlated TEM images, while these data could be collected in as short as 12 seconds using a 103 frame per second *in situ* detector.

A valuable solution to this challenge would be an artificial intelligence algorithm that can extract information from TEM data with minimum human supervision. Many works have been focused on the application of artificial intelligence and machine learning for the analysis of large and/or complex data.[5,16,22–31] In particular, several recent works have been focused on identifying and segmenting crystal lattices from high-resolution transmission or scanning transmission electron microscopy data.[23,29,31] However, the algorithms presented in many of the previous works require either a human labeled training data set, or human-made assumptions about the information contained in the data.[5,16,24–26,29–31] Such requirements on the input of human knowledge and assumptions limit the applicability of these algorithms to datasets collected by automated TEM data acquisitions workflows, which not only are large in size, but also could contain unexpected information. Maskov *et al.* demonstrated the use of variational autoencoders (VAEs) to differentiate between different types of defects in two-dimensional WS_2 from high-angle annular dark-field scanning transmission electron microscopic (HAADF-STEM) images without the need of human

assumptions.[23] VAEs are a type of deep neural network designed to find the hidden (latent) variables governing the distribution of a dataset.[32] A VAEs model encodes input data into a correlated latent space with much lower dimensionality, and subsequently reconstructs the input data from the latent space.[23,32] After the network is trained, the latent space becomes a dimensionally reduced representation of the input data, and previous studies have found that the latent variables in the latent space can often be correlated to real physical descriptors underlying the input data.[23,33] As such, the latent space provided by VAEs can be seen as an artificial intelligence's interpretation of the input data. Since the output of the network is reconstruction of the input data, the training of VAEs does not require a human labeled training dataset, making it an instance of self-supervised machine learning. VAEs have shown great performance on HAADF-STEM datasets. However application of VAEs to phase contrast HRTEM datasets, which are more complicated due to coherent elastic scattering giving raise to the image, has not been demonstrated. Phase contrast HRTEM is more amenable to quick collected of large, atomic-resolution datasets due to the parallel image formation process, and is more widely accessible suggesting an unmet need in unsupervised machine learning algorithms.

Here, we demonstrate the use of VAEs to analyze and interpret an *in situ* HRTEM dataset of lead sulfide (PbS) semiconductor nanocrystals (NCs) while being annealed by the electron beam. Annealing of semiconductor NCs has been studied extensively due to their implications for the achievement of defect-free "artificial solids".[18,20,34–39] Because of the highly dynamic nature of these processes, this *in situ* TEM data is often challenging to interpret, and often requires extensive time and efforts from a human expert to fully analyze.[18,34] In this work, we will show that VAEs trained on high-resolution TEM images of PbS NCs can produce a latent space consisting of latent variables encoding the lattice structures of these NCs. By tracking the changes of these latent variables as a function of time over the course of the annealing of a pair of NCs, one can deconvolute nanoscale dynamics with physical interpretability from the complicated annealing process. In this way, the VAEs provide an artificial intelligence approach to analyze and interpret the highly complex nanoscale annealing processes. We have found the VAEs are capable of accurately pinpointing locations where lattice structures are shifting during the annealing processes, as well as recognizing dynamics that are more complex and convoluted. As such, we believe that the VAEs show potential to serve as an artificial intelligence tool for the automated analysis of TEM datasets.

2. Results and discussion

2.1. Training of the VAEs

VAEs are a type of deep neural network that typically consists of two sets of 2-dimensional convolutional neural networks. The first one, called an encoder, passes the input data through a series of 2-dimensional convolutional layers with reducing dimensionality, until the input data is compressed into a vector

consisted of an arbitrary number of latent variables z_n, hence constructing a latent space, a low-dimensional representation of the original data. The second network, called a decoder, takes samples from the latent space as input, and passes them through a series of convolutional layers that mirrors the layers of the encoder, to reconstruct the input data. In this work, we trained our VAEs model using 64-by-64 pixels segments (Figure 1a) of high-resolution TEM images of PbS NCs. The 4096-by-4096 pixel HRTEM images were taken at a pixel scale of 0.014 nm/pixel, and hence each image represents a field of view of about 59-by-59 nm. A total of 20,400 segments containing lattice fringes of PbS obtained from 27 HRTEM images were used to train the VAEs, and this dataset is henceforth referred to as the training dataset. The 27 HRTEM images used for training the model were collected manually in about 1 h on a FEI Tecnai T20 TEM. It is estimated that about 200 recognizable PbS nanoparticles were present in the 27 HRTEM images. Importantly, the *in situ* HRTEM data analyzed by the VAEs model presented in this manuscript, including those presented in the *Supplementary Data*, were not included in the training dataset. This is to prevent the overtraining of the model on the *in situ* data. For a detailed discussion regarding how the training dataset was generated, the reader is referred to the *Supplementary Information*.

Detailed discussions regarding the training of the VAEs, including the tuning of hyperparameters, are presented in the *Supplementary Information* (Figure S2, S3). Figure 1b shows the changes in the training losses during the training of the VAEs model. Over the course of the training, the reconstruction loss steadily decreased as the Kullback–Leibler divergence increased. The decrease in reconstruction loss indicates that the network was learning to reconstruct the input images, while the increase in the KL divergence is expected as the distribution of the latent variables is expected to be more complicated than the assumed normal distribution built into the model. Of note, the increase in KL divergence is also expected, because the distributions of the latent variables representing the information in the training dataset are expected to be different from the normal distributions used to modulate the latent space. The reconstruction loss converged after the model has been trained for 20 training steps (epochs), at which point the training was stopped. Inspections of the reconstruction performance of the trained VAEs confirm that the model is capable of a reasonable reconstruction of training images, verifying that training has been sufficient (Figure 1a).

Figure 1c shows the variances of the latent variables encoding the training dataset. The latent variables were indexed based on the magnitude of their variances. Out of the 64 latent variables in the latent space, about 34 of them have significantly higher variances than the rest. This indicates that 34 latent dimensions are sufficient to encode all the information in the training dataset. However, this does not indicate that the 30 latent variables that do not have significant variances are unnecessary. In fact, the performance of the VAEs model worsens both qualitatively and quantitatively as latent dimension decreases (Figure S1). While it remains uncertain whether an optimum number of latent variables exists, it is clear that the latent variables contribute to the

overall performance of the model even if they do not have non-trivial variances. Of note, previous studies using VAEs to analyze STEM data used much lower latent dimensions.[23,40] We believe that moving to a higher latent dimension is crucial for adapting the VAEs model to the much more complex HRTEM data.

It is important to note that, during the training of the VAEs model, each latent variable was treated equally by the model, with no pre-assigned meaning or role. To distinguish the latent variables from each other, we assigned an index to each latent variable based on its variance on the training dataset (Figure 1c), such that z_{63} is the latent variable with the highest variance and z_0 is the one with the lowest. Although the latent variables were degenerate before the model was trained, the fact that they had vastly different variances after the training shows that they likely encoded different information. In the following section, we will explore how information was encoded in the latent space, and whether physical interpretation can be associated to the latent variables.

2.2. Interpretation of the latent space

To confirm that the latent space has encoded information with physical interpretability, we constructed simulated HRTEM data. PbS lattice fringes were simulated with Gaussian-distributed dark contrast on a bright background with the same periodicity as PbS lattices under HRTEM. Simple geometric permutations, including translations and rotations, were introduced to the simulated lattices to observe how the latent variables change in response to these permutations.

As shown in Figure 2, when the simulated lattices were going through different permutations, the latent variables changed accordingly. indicating that the latent space is indeed correlated to the physical positions and orientations of the simulated PbS lattices. During the training of the VAEs model, a regularization factor (β) was applied to the KL loss to enforce orthorgonality onto the latent space. In an ideal orthorgonal latent space, each latent variable should be uniquely correlated to one physical factor or process underlying the dataset. By inspecting the variances of the latent variables in Figure 2, it can be seen that, for each of the three permutations, a few latent variables showed significantly higher variances than the rest, and, importantly, the sets of latent variables that show higher variances for different permutations have minimum overlaps. This shows that while imperfect, orthorgonality is present in the latent space. This is further confirmed by the fact that by permuting a single latent variable, one can introduce a specific type of permutation to the lattice in the reconstructed image (Figure 3). Of note, while z_{43} showed high variances in both translation along the x-axis (Figure 2b) and rotation (Figure 2c), this can be explained by the fact that during the rotation of the simulated lattice, the center of rotation did not align perfectly with the center of the lattice, and hence the lattice after rotation was offset from the original lattice by a fraction of one lattice vector, resembling the effect of a translation.

The orthogonality of the latent variables has important implications for the generalizability of the VAEs model, as it suggests that the latent variables could serve as an "orthogonal basis set". Assuming it allows comparisons to be drawn

between different datasets. In the *Supporting Information*, we demonstrate an example of such a comparison between the dataset shown here, and a dataset of PbS NCs imaged by aberration-corrected transmission electron microscopy (Figure S10,S11,S12).

Finally, to confirm that the changes in the latent variables are indeed correlated to the permutations in the simulated lattices, we examined the latent variables that showed the highest variances in each type of permutation in detail (Figure S4). Taking translation in the y-axis as an example (Figure 2a, Figure S4a), we can see that as the simulated lattice was translated by one period, the change in latent variable (z_{40}), the latent variable with the highest variance during the permutation, was also periodic. Such periodicity can also be observed in the other two permutations (Figure 2b,c, Figure S4b,c). This further proves that the latent space is encoding information with real interpretable physical meaning.

It is of note that the latent variables that did not show high variances in the geometric permutations could also carry important information. For example, z_{63} did not show significant variances in Figure 2, but varying its value caused significant changes in the contrast of the simulated lattice (Figure 3d). The effects of changing the value of z_{63} on the reconstructed lattices (Figure 3d), and the distribution of high positive values of z_{63} along the edges of the nanocrystals (Figure S14), indicate that z_{63} is correlated with the position of the edges of lattices. When analyzing real *in situ* TEM data using the VAEs model, it is important to take into account the information carried by these latent variables, as will be demonstrated in the following section.

2.3. Analysis of in situ HRTEM data

After confirming that the VAEs model is capable of encoding physical information in simulated lattices into the latent space, we applied the model to analyze real high resolution *in situ* TEM data observing the dynamics of PbS NCs while being annealed by the electron beam. The annealing of heavy metal chalcogenide NCs has been studied extensively to understand mechanisms through which crystal defects can be removed from these materials.[18,20,34–39] It is well-known that when two NCs attach to each other, a variety of crystal defects can arise if the orientations of the two nanocrystals are not epitaxial, or if step edges exist on the surface of one or both of the NCs. These defects represent kinetic products and if given enough thermal energy to activate atom rearrangement, the defects can move to the surface of the nanocrystal dimer and be eliminated.[18,20,34–39] Such annealing processes can be observed under *in situ* TEM, with the electron beam of the TEM providing excitation which mimics thermal annealing.[18,34,41] Studies on these processes have provided significant insights into the fundamental mechanisms into the formation of defect-less lattices and bear important implications for the design of artificial solids.[18,34,42] Compared to the materials used in previous *in situ* TEM studies (PbTe[18], PbSe[20,37,38], CdSe[34]), PbS NCs used in this study offers the unique advantage of high air stability, allowing them to be handled in much larger quantities with less stringent requirements on inert environments, hence leading to

a much faster rate of TEM data generation. While the rock salt lattice of PbS is relatively easy to visualize under HRTEM, due to the smaller atomic weight of sulfur, compared to selenium and tellurium, the lattices of PbS NCs show lower contrast under HRTEM compared to other lead chalcogenides, making them more challenging to be analyzed using traditional methods. However, as demonstrated in Figure 1, the VAEs model is able to accurately recognize and reconstruct these lattices, showing that AI models such as the VAEs model have the potentials to reduce the requirements on the quality of the data from which information can be extracted.

In this work, we collected *in situ* HRTEM data of PbS based on the methods reported by Ondry et al.[18] The NCs, about 5 nm in size, were imaged at close to Scherzer focus under *in situ* HRTEM to visualize the reorganization of the lattices. The datasets which showed clear annealing processes were selected manually to be encoded by the VAEs model. It is important to note that, unlike STEM, the periodic contrast in HRTEM data (eg. Figure 4a) does not correspond to the physical locations of atomic columns in NCs, and should not be interpreted as such.[43,44] However, the contrast does reflect the lattice periodicity of the NCs, and therefore movement of periodic contrast seen under *in situ* HRTEM can be interpreted as a representation of the lattice reorganization happening in the annealing NCs.[18,43,44]

To analyze the *in situ* HRTEM data of annealing NCs with the VAEs model, we performed a sliding frame analysis (Figure 4b). A sliding frame 64-by-64 pixels in size was applied to the *in situ* HRTEM data with a step size of 8 pixels. Then, each time frame in each sliding frame was encoded into the latent space by the VAEs model, and the variance of each latent variable over time were calculated for each sliding frame (Figure 1d). To rule out the influences of random background fluctuations in the HRTEM data on the latent variables, Spearman's ρ correlation score was computed for each latent variable in each sliding frame as a function of time. The computed Spearman's ρ was multiplied to the variance, and the product was squared to make the final result a positive number. This squared product is henceforth referred to as the time correlated variance. Through such a sequence of analysis, each *in situ* HRTEM video was transformed to a 2D array of time correlated variance (Figure 1d).

Here, we will focus our analysis on the *in situ* HRTEM video in Figure 4a. Figure S6 shows the same analysis performed on another in situ video, demonstrating that the analysis method is generalizable. Figure 4a shows the oriented attachment of two PbS NCs along the {100} direction. Initially, the two NCs were offset by 13.7°, preventing them from achieving a strain-free attachment geometry. When exposed to the electron beam in TEM, the NCs received energy from electron beam irradiation, leading to an annealing process in which strain was reduced as the lattices reorganized and dislocations moved in the lattice. Over the course of 38 s, the angle between the two NCs was reduced by 4.4°, which can be measured both directly from the HRTEM images, and from the 2D Fourier transform of the images. While it is easy for a human expert to recognize the beginning and end point of the annealing, delineating the exact pathways of the annealing process on an atomic level by manual analysis

is challenging. However, we will demonstrate that insights on the annealing pathways can be gained by an analysis of the time correlated variances of the latent variables.

Figure 4b shows the time correlated variance maps of three latent variables: z_{29}, z_{31} and z_{63}. A quality inspection of the three maps shows that while z_{29} shows very little variance within the boundaries of the annealing nanocrystals, z_{31} has a spatial distribution of variances that roughly correlates to the distribution of lattices inside the nanocrystals, and z_{63} shows variances mainly on the boundaries of the nanocrystals. This observation further demonstrates that the VAEs model is capable of deconvoluting distinct information into different latent variables. This can be seen further by inspecting the time correlated variance maps of all latent variables (Figure S8). While z_{29} does not seem to have encoded significant information in this example, in the following discussions, we will attempt to understand the information encoded in latent variables z_{31} and z_{63}, and correlate them to interpretable physical information.

A visual inspection of the time correlated variance map of z_{31} showed that the distribution of high time correlated variances corresponds with the distribution of lattice periodicity of the NCs. Taking into account the observed annealing process of the two NCs (Figure 4a) and the behaviors of z_{31} in simulated lattices (Figure 2c), one can hypothesize that the observed changes in z_{31} is correlated to the rotation of the lattices during the annealing of the NCs. To confirm the hypothesis, we constructed two simulated lattices that are offset by a similar degree, measured as the angle between the two sets of lattice planes, as the *in situ* HRTEM data, and gradually rotated them to reduce the offset angle in a similar time frame (Figure 5c). We then selected a 64-by-64 pixels region of the *in situ* HRTEM data where the reorganization of the lattices is clearly visible (Figure 5a), and found a similar region in the simulated data, and compared how z_{31} changes in the two regions (Figure 5b,c). While z_{31} of the *in situ* HRTEM data showed fluctuations, after taking the mean of every 10 time steps, it can be seen that the trends of z_{31} in *in situ* HRTEM data closely agree with each other (Figure 5b). As such, it is likely that the changes in latent variable z_{31} indeed correspond to rotational lattice movement captured in the *in situ* HRTEM data.

Latent variable z_{63}, on the other hand, showed high time correlated variances mainly on the edges and corners of the NCs (Figure 4c). To understand what physical processes z_{63} may be capturing, we performed a Fourier transform analysis on the *in situ* HRTEM data. The lattice frequencies of the two NCs were separated and rotated such that the {100} axis is in the upright position, allowing us to obtain the traces of the lattices of the two NCs as a function of time (Figure 6c,d). By overlaying the traces of the lattices, it can be observed that the regions where z_{63} showed high time correlated variances (Figure 6a,b) correspond to the regions in which lattices were shrinking or growing over the course of the annealing process (Figure 6c,d, Figure S15). This can be further seen by comparing the changes in the values of z_{63} to the changes of the areas occupied by the lattice in the same region, computed by performing an image thresholding after Fourier filtering (Figure 6e). Since z_{63} is correlated to the po-

sitions of the edges of lattices (Figure 3d, S14), it seems likely that the changes in z_{63} correspond to the movement of atoms on the surfaces of the NCs over the course of annealing, a process known as intra-particle ripening[36,42]. Since no temporal correlation was including in the training dataset, and no *a priori* physical knowledge was built into the model, the VAEs model had no opportunity to "learn" about intra-particle annealing during its training. As such, in a sense, the VAEs model "discovered" intra-particle annealing without any human intervention or preconceived human knowledge.

The ability of the model to "interpret" without the need of interpretation makes it a promising candidate as a component of an autonomous TEM data acquisition and analysis pipeline. The minimum requirement on human input and intervention enables the model to efficiently extract physically meaningful information from large volumes of data, while also allowing it to "notice" information that a human analyzer may miss. Such capabilities allow the model to potentially play the essential role of an on-the-fly experimental decision maker in an autonomous electron microscopy setup[9,10]. For example, Figure S9 demonstrates how variances of latent variables can inform a human operator, or an autonomous image acquisition algorithm, on where nanoscale dynamics might be happening, and hence worthy of further data collection, based on a relatively short (18 s) *in situ* HRTEM video data. Hence, in the future, an automated TEM could screen over a relatively large field of view to take short *in situ* HRTEM videos at various locations. The VAEs model can then analyze the *in situ* data on the fly, and inform the automated TEM the most likely locations at which nanoscale dynamics could be happening, and are hence most worthy of further data collection. As such, a fully autonomous TEM data acquisition workflow can be achieved. Additionally, the relatively small volume of data required to train the VAEs model potentially allows it to be re-trained for different electron microscopes or material systems with relative ease, making the model suitable for a generalizable workflow for automated electron microscopy. As examples, in the *Supporting Information*, we demonstrated the performance of the model on two different datasets (Figure S11, S12, S13), one of which involves a material different from PbS (Figure S13). We envision that in the near future, the VAEs model will become an integral part of an autonomous laboratory focused on automated data acquisition and analysis by electron microscopy.

3. Conclusions

In this work, we demonstrated that the VAEs model can be trained to faithfully reconstruct atomically-resolved TEM images, and construct a latent space that encodes physically interpretable information contained in the TEM data. We showed that the latent space can capture permutations and reorganizations of lattices in both simulated and real HRTEM data. Based on the insights provided by the latent variables, we proposed a mechanism for the annealing of PbS NCs, which can be seen as an artificial intelligence's interpretation of the HRTEM data. The model shows great potential to serve as part of an autonomous workflow for automated electron microscopy.

4. Methods

4.1. Resource availability

4.1.1. Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, A. Paul Alivisatos (paul.alivisatos@uchicago.edu).

4.1.2. Materials availability

This study did not generate new unique reagents.

4.1.3. Data and codes availability

Unprocessed HRTEM data and *in situ* TEM data, and processed data for training the VAEs model are available on Dryad (`https://datadryad.org/stash/share/kLvdk-AartSZf1GOw1slklZ0mLmPCAkQeYorCbeBMdI`). Trained VAEs models, and sample codes for training and applying them, are available on Github (`https://github.com/realxingzhiwang/VAEs-for-HRTEM`).

4.2. Materials

Lead chloride ($PbCl_2$), lead nitrate ($Pb(NO_3)_2$) elemental sulfur, oleic acid (OA), oleylamine (OLA), hexanes, toluene, acetone were purchased from Sigma-Aldrich and used without further purification. Aqueous ammonium sulfide ($(NH_4)_2S$, 40 *wt*%) was purchased from Sigma-Aldrich and titrated with $Pb(NO_3)_2$ to determine the actual sulfide concentration based on methods reported by Zhang et al.[45] Molecular sieves (3 Å) were purchased from Sigma-Aldrich and activated at 300 °C for at least 12 hours before being used.

4.3. Synthesis of PbS NCs

Cubic PbS NCs about 6 nm in diameter were synthesized based on a two step approach. First, pseudo-spherical PbS NCs were prepared based on the methods reported by Weidman et al.[46] Briefly, 2.5 g $PbCl_2$ was mixed with 7.5 mL OLA. The mixture was degassed at room temperature for 10 min, then at 110 °C for 5 min. After degassing, the mixture was heated to 120 °C to form a milky white suspension. The sulfur source was prepared by adding 7.5 mL OLA to 40 mg of S. The mixture was degassed for 20 minutes at 120 °C, and then cooled to close to 30 °C to form a dark orange solution. To form the NCs, 2.25 mL of the S solution was swiftly injected into the Pb solution. When a dark color was observed in the mixture, the reaction was allowed to continue for 10 min at 120 °C. Then, the reaction mixture was removed from heat source, and 20 mL of hexanes was injected into the mixture to quench the reaction. PbS NCs were collected by adding 10 mL IPA and 5 mL MeOH to the reaction mixture, and centrifuge at 4000 rpm for 3 min. Solid products collected were resuspended in 10 mL hexanes. Then, 20 mL OA was added to the resuspended products. The mixture was sonicated for 5 min, and centrifuged at 4000 rpm for 3 min to collect solid products. Collected products were washed with OA two more times, or until the supernatant is clear, and then one more time with

10 mL IPA and 5 mL MeOH. The resultant suspension of PbS NCs in hexanes was then centrifuged at 4000 rpm for 3 min once more to remove remaining solid $PbCl_2$. The suspension was then stored in an air-free environment for at least 24 h, after which the suspension was once again centrifuged at 4000 rpm for 3 min to remove any remaining $PbCl_2$. UV-Vis spectroscopy and and TEM characterization showed that the collected NCs were 6 nm spheroidal NCs with an absorption peak at 0.79 eV with a narrow size distribution (Figure S5a,b), agreeing with previous reports by Weidman et al.[46]

To modify the spheroidal NCs to form cubic NCs, we adopted a method reported by Zhang et al.[45] Under an air-free environment, 3 mg spheroidal PbS NCs were suspended in 4 mL hexanes with 0.2 mL OA. A $(NH_4)_2S$/OLA solution was prepared by adding 0.3 mmol $(NH_4)_2S$ to 10 mL OLA. The solution was then dried by ∼2 g of 3 Å molecular sieves under an inert environment for 30 min. The solution was kept at 30-40 °C to prevent solidification. Then in an inert environment, 4 mL of the $(NH_4)_2S$/OLA solution was injected into the PbS suspension. The mixture was allowed to react for 30 min. Then 20 mL acetone was added and the mixture was centrifuged at 4000 rpm for 3 min to collect cubic PbS NCs (Figure S5c). The NCs were suspended in 5 mL hexanes and washed with 5 mL acetone two more times, after which they were resuspended in 5 mL toluene and used to prepare TEM samples immediately.

4.4. Preparation of TEM samples

The suspension of cubic PbS NCs was diluted with toluene until the suspension was light yellow. Then, 2 μL of the diluted suspension was dropcasted onto a plasma cleaned Ted Pella 400 mesh ultrathin carbon support Au TEM grid. The grid was let to dry under air-free condition for at least 2 hours. Then, the grid was dipped into a 0.45 M solution of $(NH_4)_2S$ in methanol for 30 s, and then in pure methanol for another 30 s. The grid was then dried in vacuum for at least 2 hours before imaging.

4.5. In situ TEM imaging

In situ TEM data of the annealing of PbS NCs were acquired using a FEI Tecnai T20 transmission electron microscope equipped with a Gatan RIO16IS camera and a LaB_6 filament. All data were acquired under 200 kV accelerating voltage. Atomically resolved *in situ* videos were recorded close to Scherzer focus. The dose rates of electrons the samples were exposed to during imaging were estimated after the *in situ* data were collected. A conversion value of 124 was used to convert CCD counts to electrons.

4.6. Training of the VAEs model

The VAEs model was constructed in Python 3 using the Keras package. The encoder network was consisted of four 2D convolutional layers with a kernel size of 4, followed by a reshape layer and another convolutional layer that projects the output of the previous layer onto the 64-dimensional latent space. The

decoder network has the same set of layers with the reverse sequence, and with the convolutional layers replaced by inverse convolutional layers.

The model was trained by minimizing the reconstruction loss measured by binary cross entropy (BCE) and the original data. To modulate the distribution of the latent variables, Kullback–Leibler (KL) divergence between the distribution of the latent variables and a normal distribution is added to BCE during the training process as a constraint.[23,32] A factor of 2 was multiplied to the KL divergence as a regularization factor (β) to enforce orthogonality in the latent space. The model was trained on the training dataset for 20 epochs, and the resulting latent space was analyzed for physical insights.

5. Acknowledgement

The synthesis of nanomaterials, TEM data acquisition, and algorithm development were supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02-05-CH11231 within the Physical Chemistry of Inorganic Nanostructures Program (KC3103). Transmission aberration-corrected electron microscopy work done at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. J.C.O gratefully acknowledges the support of the Kavli Philomathia Graduate Student Fellowship. The authors would also like to thank Dr. Jakob C. Dahl, Samuel Gleason, Prof. Vida Jamali, Dr. Michelle F. Crook, Dr. Katherine Sytwu, Prof. Mary Scott for useful discussions.

6. Author contributions

Conceptualization, X.W. and A.P.A.; Methodology, J.C.O. and P.E.; Software, X.W. and V.B.; Formal Analysis, X.W.; Investigation, X.W. and C.Y.; Writing – Original Draft, X.W.; Writing – Review and Editing, C.Y., J.C.O., P.E., and A.P.A.; Supervision, C.Y., J.C.O., P.E., and A.P.A.; Funding Acquisition, A.P.A.

7. Declaration of interests

The authors declare no competing interests.

8. Inclusion and diversity statement

We support inclusive, diverse, and equitable conduct of research.

References

[1] Ruska, E. (1987)., The Development of the Electron Microscope and of Electron Microscopy (Nobel Lecture), Angewandte Chemie International Edition in English 26 595–605. doi:10.1002/anie.198705953.

[2] Spurgeon, S. R., Ophus, C., Jones, L., Petford-Long, A., Kalinin, S. V., Olszta, M. J., Dunin-Borkowski, R. E., Salmon, N., Hattar, K., Yang, W. C. D., Sharma, R., Du, Y., Chiaramonti, A., Zheng, H., Buck, E. C., Kovarik, L., Penn, R. L., Li, D., Zhang, X., Murayama, M., Taheri, M. L. (2021)., Towards data-driven next-generation transmission electron microscopy, Nature Materials 20 274–279. doi:10.1038/s41563-020-00833-z.

[3] Mulligan, S. K., Speir, J. A., Razinkov, I., Cheng, A., Crum, J., Jain, T., Duggan, E., Liu, E., Nolan, J. P., Carragher, B., Potter, C. S. (2015)., Multiplexed TEM Specimen Preparation and Analysis of Plasmonic Nanoparticles, Microscopy and Microanalysis 21 1017–1025. doi:10.1017/S1431927615014324.

[4] Schorb, M., Haberbosch, I., Hagen, W. J., Schwab, Y., Mastronarde, D. N. (2019)., Software tools for automated transmission electron microscopy, Nature Methods 16 471–477. doi:10.1038/s41592-019-0396-9.

[5] Cheng, Z., Wang, C., Wu, X., Chu, J. (2022)., Review in situ transmission electron microscope with machine learning, Journal of Semiconductors 43 (2022). doi:10.1088/1674-4926/43/8/081001.

[6] Tan, Y. Z., Cheng, A., Potter, C. S., Carragher, B. (2016)., Automated data collection in single particle electron microscopy, Microscopy (Oxford, England) 65 43–56. doi:10.1093/jmicro/dfv369.

[7] House, S. D., Chen, Y., Jin, R., Yang, J. C. (2017)., High-throughput, semi-automated quantitative STEM mass measurement of supported metal nanoparticles using a conventional TEM/STEM, Ultramicroscopy 182 145–155. doi:10.1016/j.ultramic.2017.07.004.

[8] Kalinin, S. V., Ziatdinov, M., Spurgeon, S. R., Ophus, C., Stach, E. A., Susi, T., Agar, J., Randall, J. (2022)., Deep learning for electron and scanning probe microscopy : From materials design to atomic fabrication, MRS Bulletin 47 931–939. doi:10.1557/s43577-022-00413-3.

[9] Kalinin, S. V., Mukherjee, D., Roccapriore, K., Blaiszik, B., Ghosh, A., Ziatdinov, M., Al-Najjar, A., Doty, C., Akers, S., Rao, N. S., Agar, J., Spurgeon, S. R. (2023)., Deep Learning for Automated Experimentation in Scanning Transmission Electron Microscopy, arXiv preprint arXiv:2304.02048.

[10] Olszta, M., Hopkins, D., Fiedler, K. R., Oostrom, M., Akers, S., Spurgeon, S. R. (2022)., An Automated Scanning Transmission Electron Microscope Guided by Sparse Data Analytics, Microscopy and Microanalysis 28 1611–1621. doi:10.1017/S1431927622012065.

[11] Hauwiller, M. R., Ondry, J. C., Alivisatos, A. P. (2018)., Using graphene liquid cell transmission electron microscopy to study in situ nanocrystal etching, Journal of Visualized Experiments 2018 1–9. doi:10.3791/57665.

[12] Taheri, M. L., Stach, E. A., Arslan, I., Crozier, P. A., Kabius, B. C., LaGrange, T., Minor, A. M., Takeda, S., Tanase, M., Wagner, J. B., Sharma, R. (2016)., Current status and future directions for in situ transmission electron microscopy, Ultramicroscopy 170 86–95. doi:10.1016/j.ultramic.2016.08.007.

[13] Yuk, J. M., Park, J., Ercius, P., Kim, K., Hellebusch, D. J., Crommie, M. F., Lee, J. Y., Zettl, A., Alivisatos, A. P. (2012)., High-Resolution EM of Colloidal Nanocrystal Growth Using Graphene Liquid Cells, Science 336 61–65.

[14] Hawkes, P. W. (2009)., Aberration correction past and present, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 367 3637–3664. doi:10.1098/rsta.2009.0004.

[15] Scott, M. C., Chen, C. C., Mecklenburg, M., Zhu, C., Xu, R., Ercius, P., Dahmen, U., Regan, B. C., Miao, J. (2012)., Electron tomography at 2.4-ångström resolution, Nature 483 444–447. doi:10.1038/nature10934.

[16] Unruh, D., Kolluru, V. S. C., Baskaran, A., Chen, Y., Chan, M. K. (2022)., Theory+AI/ML for microscopy and spectroscopy: Challenges and opportunities, MRS Bulletin 47 1024–1035. doi:10.1557/s43577-022-00446-8.

[17] Sainju, R., Chen, W. Y., Schaefer, S., Yang, Q., Ding, C., Li, M., Zhu, Y. (2022)., DefectTrack: a deep learning-based multi-object tracking algorithm for quantitative defect analysis of in-situ TEM videos in real-time, Scientific Reports 12 1–14. doi:10.1038/s41598-022-19697-1.

[18] Ondry, J. C., Hauwiller, M. R., Alivisatos, A. P. (2018)., Dynamics and Removal Pathway of Edge Dislocations in Imperfectly Attached PbTe Nanocrystal Pairs: Toward Design Rules for Oriented Attachment, ACS Nano 12 3178–3189. doi:10.1021/acsnano.8b00638.

[19] Hauwiller, M. R., Frechette, L. B., Jones, M. R., Ondry, J. C., Rotskoff, G. M., Geissler, P., Alivisatos, A. P. (2018)., Unraveling Kinetically-Driven Mechanisms of Gold Nanocrystal Shape Transformations Using Graphene Liquid Cell Electron Microscopy, Nano Letters 18 5731–5737. doi:10.1021/acs.nanolett.8b02337.

[20] Wang, Y., Peng, X., Abelson, A., Zhang, B. K., Qian, C., Ercius, P., Wang, L. W., Law, M., Zheng, H. (2019)., In situ TEM observation of neck formation during oriented attachment of PbSe nanocrystals, Nano Research 12 2549–2553. doi:10.1007/s12274-019-2483-8.

[21] Yan, C., Byrne, D., Ondry, J. C., Kahnt, A., Moreno-Hernandez, I. A., Kamat, G. A., Liu, Z. J., Laube, C., Crook, M. F., Zhang, Y., Ercius, P., Alivisatos, A. P. (2022)., Facet-selective etching trajectories of individual semiconductor nanocrystals, Science Advances 8 1–11. doi:10.1126/sciadv.abq1700.

[22] Wang, X., Li, J., Ha, H. D., Dahl, J. C., Ondry, J. C., Moreno-Hernandez, I., Head-Gordon, T., Alivisatos, A. P. (2021)., AutoDetect-mNP: An Unsupervised Machine Learning Algorithm for Automated Analysis of Transmission Electron Microscope Images of Metal Nanoparticles, JACS Au 1 316–327. doi:10.1021/jacsau.0c00030.

[23] Maksov, A., Dyck, O., Wang, K., Xiao, K., Geohegan, D. B., Sumpter, B. G., Vasudevan, R. K., Jesse, S., Kalinin, S. V., Ziatdinov, M. (2019)., Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS2, npj Computational Materials 5 1–8. doi:10.1038/s41524-019-0152-9.

[24] Jamali, V., Hargus, C., Ben-Moshe, A., Aghazadeh, A., Ha, H. D., Mandadapu, K. K., Alivisatos, A. P. (2021)., Anomalous nanoparticle surface diffusion in LCTEM is revealed by deep learning-assisted analysis, Proceedings of the National Academy of Sciences of the United States of America 118 (2021). doi:10.1073/pnas.2017616118.

[25] Botifoll, M., Pinto-huguet, I., Arbiol, J. (2022)., Machine learning in electron microscopy for advanced nanocharacterization : current developments , available tools and future outlook, Nanoscale Horizons Advance Ar (2022). doi:10.1039/d2nh00377e.

[26] Baskaran, A., Kautz, E. J., Chowdhary, A., Ma, W., Yener, B., Lewis, D. J. (2021)., The Adoption of Image-Driven Machine Learning for Microstructure Characterization and Materials Design: A Perspective, JOM 73 3639–3657.

[27] Prifti, E., Klie, R., Buban, J. (2022)., Deep Learning Computer Vision for Anomaly Detection in Scanning Transmission Electron Microscopy, Microscopy and Microanalysis 28 3018–3020. doi:10.1017/s1431927622011291.

[28] Ziatdinov, M., Ghosh, A., Wong, C. Y., Kalinin, S. V. (2022)., AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy, Nature Machine Intelligence 4 1101–1112. doi:10.1038/s42256-022-00555-8.

[29] Cheng, Z., Xie, X., Yang, Y., Wang, C., Luo, C., Bi, H., Wang, Y., Chu, J., Wu, X. (2023)., Neural network approach for ferroelectric hafnium oxide phase identification at the atomistic scale, Materials Today Electronics 3 100027. doi:10.1016/j.mtelec.2023.100027.

[30] Ou, Z., Wang, Z., Luo, B., Luijten, E., Chen, Q. (2020)., Kinetic pathways of crystallization at the nanoscale, Nature Materials 19 450–455. doi:10.1038/s41563-019-0514-1.

[31] Sytwu, K., Groschner, C., Scott, M. C. (2022)., Understanding the Influence of Receptive Field and Network Complexity in Neural Network-Guided TEM Image Analysis, Microscopy and Microanalysis 28 1896–1904. doi:10.1017/S1431927622012466.

[32] Kingma, D. P., Welling, M. (2014)., Auto-Encoding Variational Bayes, arXiv 1–14.

[33] Lu, P. Y., Kim, S., Soljačić, M. (2020)., Extracting Interpretable Physical Parameters from Spatiotemporal Systems Using Unsupervised Learning, Physical Review X 10 1–19. doi:10.1103/PHYSREVX.10.031056.

[34] Ondry, J. C., Philbin, J. P., Lostica, M., Rabani, E., Alivisatos, A. P. (2019)., Resilient Pathways to Atomic Attachment of Quantum Dot Dimers and Artificial Solids from Faceted CdSe Quantum Dot Building Blocks, ACS Nano 13 12322–12344. doi:10.1021/acsnano.9b03052.

[35] Cui, J., Pan, Y. E., Koley, S., Shamalia, D., Waiskopf, N., Remennik, S., Popov, I., Oded, M., Banin, U. (2019)., Colloidal quantum dot molecules manifesting quantum coupling at room temperature, Nature Communications 10 5401–5411. doi:10.1038/s41467-019-13349-1.

[36] Cui, J., Koley, S., Panfil, Y. E., Levi, A., Ossia, Y., Waiskopf, N., Remennik, S., Oded, M., Banin, U. (2021)., Neck Barrier Engineering in Quantum Dot Dimer Molecules via Intraparticle Ripening, Journal of the American Chemical Society 143 19816–19823. doi:10.1021/jacs.1c08863.

[37] Walravens, W., Solano, E., Geenen, F., Dendooven, J., Gorobtsov, O., Tadjine, A., Mahmoud, N., Ding, P. P., Ruff, J. P., Singer, A., Roelkens, G., Delerue, C., Detavernier, C., Hens, Z. (2019)., Setting Carriers Free: Healing Faulty Interfaces Promotes Delocalization and Transport in Nanocrystal Solids, ACS Nano 13 12774–12786. doi:10.1021/acsnano.9b04757.

[38] A. van Huis, M., T. Kunneman, L., Overgaag, K., Xu, Q., Pandraud, G., W. Zandbergen, H., Vanmaekelbergh, D. (2008)., Low-Temperature Nanocrystal Unification through Rotations and Relaxations Probed by in Situ Transmission Electron Microscopy, Nano Letters 8 3959–3963. doi:10.1021/nl8024467.

[39] Smeaton, M. A., El Baggari, I., Balazs, D. M., Hanrath, T., Kourkoutis, L. F. (2021)., Mapping Defect Relaxation in Quantum Dot Solids upon in Situ Heating, ACS Nano 15 719–726. doi:10.1021/acsnano.0c06990.

[40] Creange, N., Dyck, O., Vasudevan, R. K., Ziatdinov, M., Kalinin, S. V. (2022)., Towards automating structural discovery in scanning transmission electron microscopy, Mach. Learn.: Sci. Technol. 3 015024. doi:10.1088/2632-2153/ac3844.

[41] Van Dyck, D., Lobato, I., Chen, F. R., Kisielowski, C. (2015)., Do you believe that atoms stay in place when you observe them in HREM?, Micron 68 158–163. doi:10.1016/j.micron.2014.09.003.

[42] Ondry, J. C., Philbin, J. P., Lostica, M., Rabani, E., Alivisatos, A. P. (2021)., Colloidal Synthesis Path to 2D Crystalline Quantum Dot Superlattices, ACS Nano 15 2251–2262. doi:10.1021/acsnano.0c07202.

[43] O'Keefe, M. A., Buseck, P. R., Iijima, S. (1978)., Computed crystal structure images for high resolution electron microscopy, Nature 274 322–324. doi:10.1038/274322a0.

[44] Geuens, P., Van Dyck, D. (2002)., The S-state model: A work horse for HRTEM, Ultramicroscopy 93 179–198. doi:10.1016/S0304-3991(02)00276-0.

[45] Zhang, H., Yang, J., Hanrath, T., Wise, F. W. (2014)., Sub-10 nm monodisperse PbS cubes by post-synthesis shape engineering, Physical Chemistry Chemical Physics 16 14640–14643. doi:10.1039/c4cp01531b.

[46] Weidman, M. C., Beck, M. E., Hoffman, R. S., Prins, F., Tisdale, W. A. (2014)., Monodisperse, air-stable PbS nanocrystals via precursor stoichiometry control, ACS Nano 8 6363–6371. doi:10.1021/nn5018654.

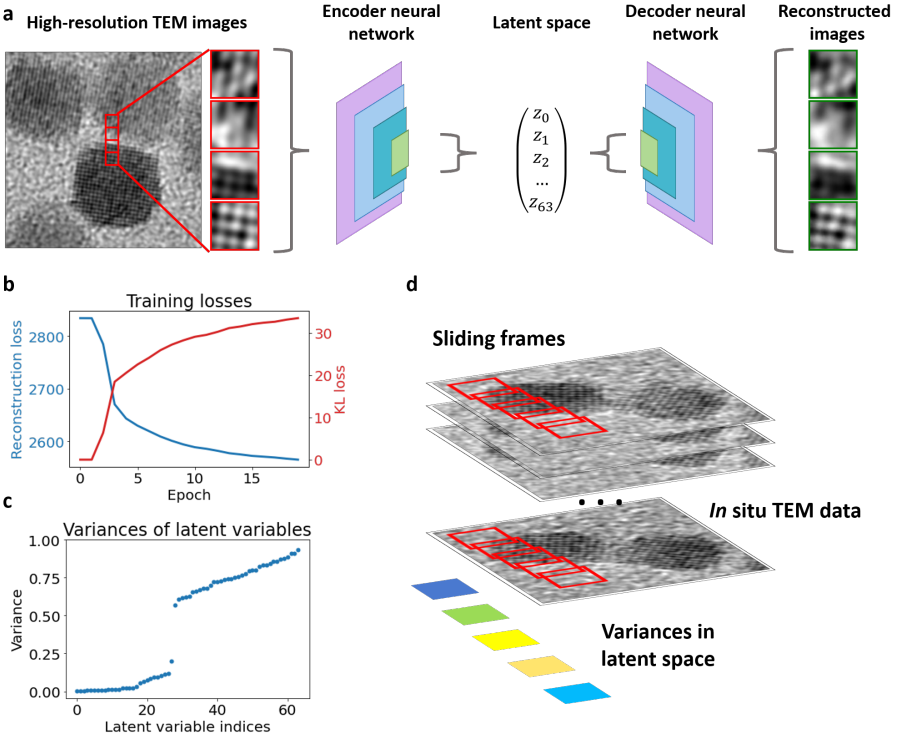

Figure 1: a) Schematic of a 2D VAEs trained on high-resolution TEM images of PbS nanocrystals. b) Training curves of the VAEs showing both the binary cross entropy reconstruction loss and the Kullback–Leibler divergence as functions of epochs trained. c) Variances of the latent variables on the training set. Latent variables indexed based on magnitude of their variances. d) Schematic of sliding frame latent space variance analysis used to annalyze *in situ* TEM data.

9. Figures

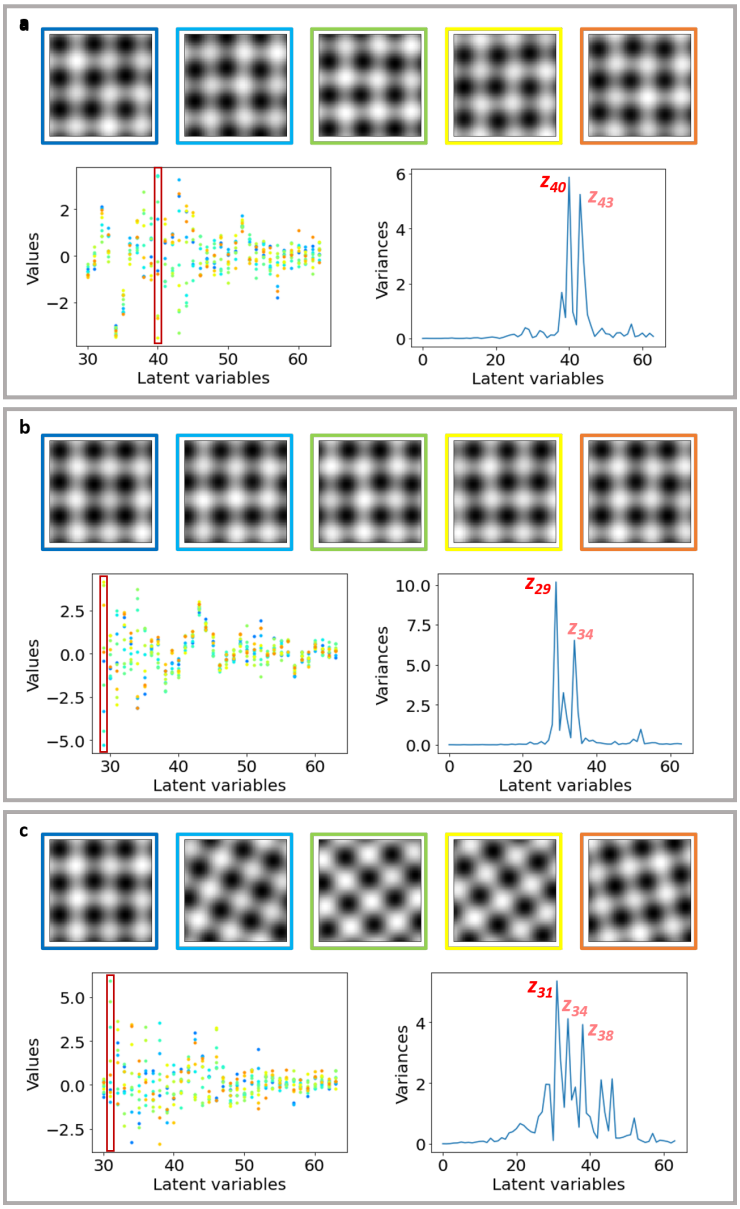

Figure 2: Distributions and variances of latent variables in simulating *in situ* TEM data showing a) translation by one lattice period in the y-axis, b) translation by one lattice period in the x-axis, c) rotation by 90deg clockwise. Snapshots of the simulated in situ TEM data at different time points are shown at the top of each subfigure. Red texts and boxes in each subfigure denotes the latent variables with the highest variance in each permutation. Other latent variables with notably high variances are denoted by pink texts.

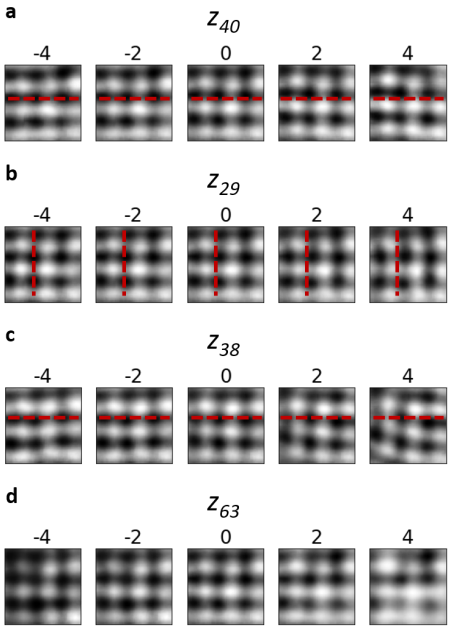

Figure 3: Effects of permuting one latent variable while keeping all the others constant on the reconstruction of a simulated lattice. The red dotted lines were drawn at the same pixel locations in each panel, serving as a visual aid for comparing the positions of the lattices. The latent variables permuted were the ones that showed highest variances during the a) translation in the y-axis, b) translation in the x-axis, c) rotation clockwise permutations of the simulated lattices. d) A latent variable that did not show large variances in simple geometric permutations of the simulated lattices, but capable of "erasing" the lattice on the edge. The same analysis on all latent variables is shown in Figure S7.

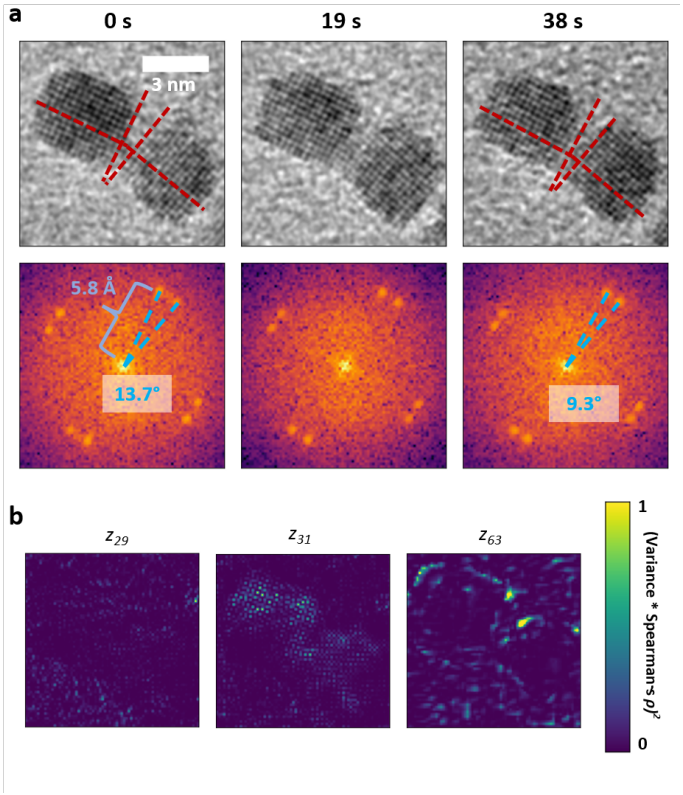

Figure 4: a) Snap shots of *in situ* TEM data showing the annealing of two PbS nanocrystals, with the corresponding power spectra obtained from 2D Fourier transforms. The *in situ* was taken under an electron dose rate of $\sim 1250\ e^-/Å^2 \cdot s$. b) Maps of time correlated variances of latent variables z_{29}, z_{31} and z_{63} in each sliding frame, demonstrating how the two latent variables are capturing different dynamics in the *in situ* TEM data.

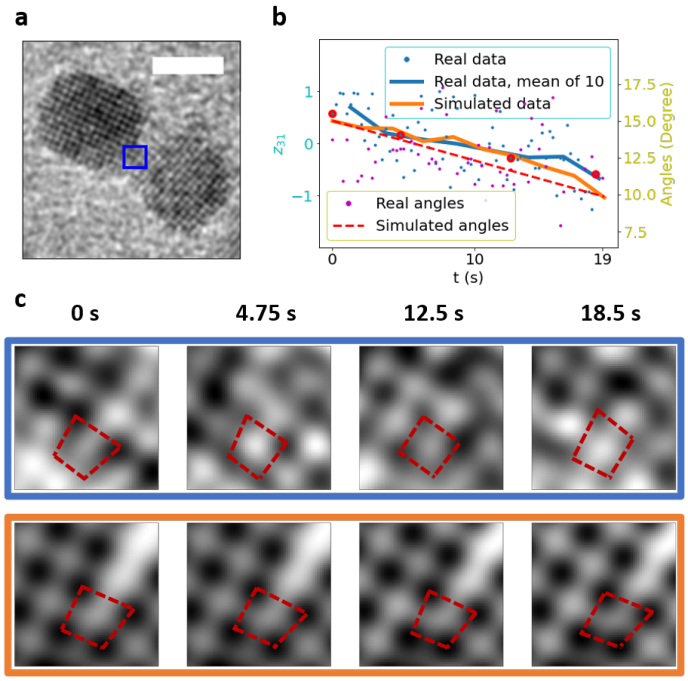

Figure 5: a) A 64-by-64 pixels subregion (blue box) of the *in situ* TEM data of PbS nanocrystals annealing in which the reorganization of the two lattices can be seen. Scale bar represents 3 nm. b) (Left axis) The values of latent variable z_{31} in the subregion (real data) as a function of time, compared to those of simulated data showing the same annealing behaviors of two lattices. (Right axis) Angles between the two nanocrystals measured from the Fourier transform of the TEM data, and the angles between the two simulated lattices in simulated data, as functions of time. c) Comparisons between the real and simulated data at selected time points (denoted by red dots in b). Red dotted lines represents sets of lattice vectors forming a close circuit, shown as a visual aid.

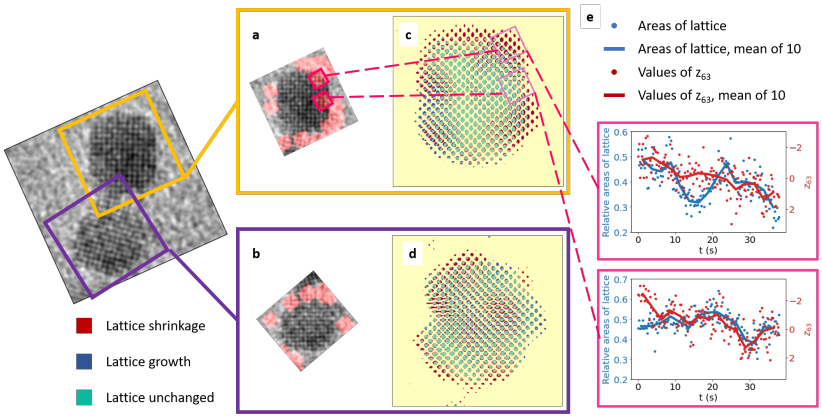

Figure 6: a, b) The two annealing nanoparticles. The particles are rotated such that the {100} facets are facing upward. Red regions denotes regions where latent variable z_{63} shows high variances. c-d) Traces of lattices of the two annealing nanoparticles in the first and last frame of the *in situ* data. Regions where lattice where present in the first frame, but not the last frame (red) were considered regions of lattice shrinkage. Regions where lattice where present in the last frame, but not the first frame (blue) were considered regions of lattice growth. If lattices were present in both frames, the regions were considered regions of lattice unchanged (cyan). Note that the lattice frequencies in the two frames were not exactly the same, likely due to a fluctuation in the stigmatism of electron lenses during data collection. e) Comparisons between the values of z_{63} and the relative areas occupied by the lattice as functions of time at two selected subregions of the *in situ* data.