Enhanced single-shot inversion of the diffraction phase problem with the diffusion model

Seungjun Lee[1], Xi Yu[1], Yuewei Lin[1], Shinjae Yoo[1], and Ian Robinson[2]

March 24, 2026

[1]Computing and Data Sciences, Brookhaven National Laboratory, Upton, NY 11973, USA
[2]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA

Corresponding authors: Seungjun Lee (slee19@bnl.gov) and Ian Robinson (irobinson@bnl.gov)

1 Abstract

We present an advance in computational solution of the crystallographic "phase problem". The diffraction data used in Bragg Coherent Diffraction Imaging experiments are "oversampled" because there are many more measurement points than unknown pixel values. Their inversion in the presence of noise often leads to multiple solutions. Here we reexamine diffraction data measured previously on thin films of superconducting $La_{1.93}Sr_{0.07}CuO_4$ material on its (103) Bragg peak with bursts of single pulses of X-ray Free Electron Laser radiation. We demonstrate that the diffusion model is able to filter these solutions sufficiently to give reproducible images. The images show nanometer-sized regions of material phase-shifted with respect to each other. These are expected from the use of the Atomic Layer Epitaxy methods used to prepare the samples.

2 Introduction

X-ray Bragg diffraction with a focused coherent beam is a powerful experimental method for imaging domains in thin films. Domains, which form during the growth of thin films, are phase-shifted with respect to each other when they contribute to Bragg diffraction spots. The phase shifts give rise to interference "speckles" in the diffraction, which encode the structure. Inverting the diffraction to an image is difficult because of the "phase problem" inherent to

all diffraction, that only the intensity of the speckles can be measured and the relative phase of the X-rays hitting is lost in the detection process. Nevertheless, knowledge of the domain structure is important to understand the growth mechanism of the films, so it can be optimized for applications, which motivates the development of computational inversion algorithms.

We first need to consider that the phase-domain problem is more difficult to solve than the more conventional application of Bragg Coherent Diffractive Imaging (BCDI) to moderately strained crystals [1], which fall in the "weak phase" limit. Distortions of a crystal smaller than a half a lattice spacing lead to shifts, ϕ, of their contribution to the diffraction of $-\pi/2 < \phi < \pi/2$. Up to this level of distortion, these signals can be considered to be *in phase* with the rest of the crystal and modulate the signal due to the shape of the crystal, in the form of fringes surrounding the central Bragg peak. These diffraction patterns can be inverted by iterative algorithms, which switch back and forth between complex real-space images and reciprocal-space diffraction patterns by alternating Fourier transforms. On each iteration, the diffraction signal is updated by overwriting the measured magnitude (square root of intensity) and retaining the phase, while the image is updated by a "real-space constraint" [2]. This is most commonly the application of a "support" mask to zero (or overdrive) the image outside a known region. The support, which is central to the success of these algorithms because it enforces the Shannon-Nyquist "oversampling" of the diffraction pattern signal, can be dynamically updated during the iterative cycling by the "shrink-wrap" method [3]. This method has become the standard procedure [1] for BCDI and has recently been augmented by AI methods based on convolutional neural networks (CNNs) [4, 5, 6].

As soon as the phase shift goes beyond $-\pi/2 < \phi < \pi/2$, cancellation of the diffraction starts to take place by interference. This is seen directly in the diffraction data as the start of peak splitting, where the central peak becomes broken into multiple maxima. Phase domain structures have unlimited phase shifts, so they can be represented by complex numbers with $-\pi < \phi < \pi$. It was shown by qualitative arguments, under certain assumptions, that the number of central peaks in the diffraction pattern is equal to the number of domains in the structure [7], which is a useful guide. It is found experimentally that both the iterative methods and CNNs described above start to fail when this situation arises. The failure is characterized by apparent non-uniqueness of the solution: different random seeding of the algorithm leads to different solutions. In simpler cases, for structures containing just a few domains, similar-looking solutions are generated.

Theoretical considerations of the uniqueness of this phase problem [8] have concluded that, since it is mathematically overdetermined by the Shannon-Nyquist oversampling, a unique solution should exist in two (2D) or three dimensions (3D). One-dimensional phase problems are non-unique in the sense that the complex diffraction pattern can be expressed as a polynomial function, which can be factorized by solving for the roots; each factor in the expansion product of the polynomial can be conjugated, so there are 2^n possible n'th-order polynomials that have the same magnitude. Since factorizing 2D and 3D func-

tions is rarely possible, these are protected from non-uniqueness. According to Bates [8], these situations would be "pathological" exceptions. This argument, however, does not apply to experimental data that always contains noise. There is generally no solution to the phase problem for noisy data. Nevertheless, a good algorithm, with an appropriately chosen loss function, for example, the Euclidean distance squared (chi-square), should get close to a solution, but there is no guarantee this is unique. Noise is therefore understood to lead to a distribution of solutions, consistent with the collected wisdom of a large number of experimental trials by a large number of practitioners of BCDI.

This provides an opportunity which has been incorporated into BCDI phasing procedures, incorporated into more recent software packages [9, 10]. Good data with less noise should lead to a narrower distribution of solutions, but there will always be outliers, for example, when the conjugate solution (indistinguishable from the original by the symmetry of the Fourier transform) becomes mixed with the true solution. A scatter approach of performing many reconstructions and filtering the results by some objective function, then averaging the *good* results has shown success. The "cdi-utils" package employs the "mean-to-maximum" ratio to select the good data and performs a common-mode selection of the best [10]. The "Guided HIO" method [9] selects based on the sharpness of the image, then breeds a new generation of starting models for further cycles. In both cases, the "twin" conjugate images have to be detected and corrected before merging solutions. In this current work, we accept there will always be a range of solutions from experimental data and adopt the idea of filtering a distribution of solutions to obtain the "best" result.

3 Background

We previously developed a convolutional neural networks (CNN) to invert diffraction patterns from a thin film of $La_{2-x}Sr_xCuO_4$ cuprate measured with single pulses of X-rays from an X-ray Free-electron Laser (XFEL) [11]. The experiment was described in the previous publication, but summarized here. The superconducting sample, with a thickness of 40 repeats of the La and Cu layers along the c-axis, was grown layer-by-layer in an Atomic Layer Epitaxy (ALE) growth chamber. Coherent X-ray diffraction patterns were measured at the (103) Bragg peak using the MID endstation [12] of the European XFEL in Hamburg using 140-pulse bursts of 50-fs X-ray pulses at a 2-MHz repetition rate. Beam attenuators were used so that the sample was not damaged by the beam, allowing 2000 frames of 140 pulses to be collected. The observed response included possible thermal fluctuation of the double-crystal monochromator, which was in place. Big fluctuations of intensity both between trains and within the pulse trains were observed, attributed to the Self Amplified Spontaneous Emission (SASE) process of the XFEL. Full details of the experiment are provided in ref [11].

The overall goal of this work is to understand the origin of the fluctuations seen in the data, particularly to establish whether variations of the beam position on the sample were important. This requires inversion of the diffraction

patterns to real-space images, which can be interpreted. The data, which are inverted using the diffusion model here, consisting of a few photons per pulse, were merged into two diffraction patterns as described below in section 4.3. The diffraction patterns are strongly speckled with an elongated diagonal ellipse shape for both the speckles and their envelope describing the Bragg peak itself; this is explained by the sample geometry and the beam footprint [11]. The speckle pattern is interpreted as coming from phase domains, which are due to the epitaxial growth of the sample. Crystalline regions of the $La_{2-x}Sr_xCuO_4$ film nucleate at step edges and other defect sites in the substrate, which lead to phase shifts between the resulting crystal domains. The number of speckles, corresponding to the approximate number of domains in the beam [7], is about 50. This number was therefore used to create simulated diffraction patterns, which were then used for training of the diffusion model, as described below in section 4.5.

The previous analysis [11] used a UNet consisting of three convolutional layers of encoder and decoder and skip connections between them. A single-channel diffraction pattern was applied to the first layer of the encoder, and two channels appeared at the output of the decoder. Rather than using separate amplitude and phase image decoder channels, these were kept together in a combined complex image. Because of GPU hardware limitations, the complex arithmetic was performed explicitly on the real and imaginary sub-layer channels in the decoder. The CNN was trained on diffraction patterns from simulated phase domain arrays with randomly generated phase shifts multiplied by an elliptical beam shape, for which the ground truth was available. This complex CNN was found to perform better than the traditional real-valued CNN, presumably by keeping the related amplitude and phase image information together. After training, realistic-looking images of domains were obtained with the experimental data.

Diffusion-based generative models have recently emerged as the state of the art in sample quality and training stability [13, 14, 15]. Building on these advances, diffusion models are now widely used as a powerful regularizer for inverse problems, where they gradually generate solutions from a learned data distribution while enforcing measurement consistency. In the image domain, [16] introduces a guided diffusion framework that enables posterior sampling for a broad class of forward operators, while [17, 18] extend this framework to efficiently integrate constraint enforcement within the diffusion process. Similarly, [19] adapts diffusion models to physics-constrained reconstruction tasks such as MRI and CT, demonstrating improved robustness under severe undersampling. Another important application area is the reconstruction of full spatial or spatiotemporal fields from sparse measurements, where diffusion models are combined with partial differential equation (PDE) constraints to guide the recovery of physically consistent solutions. [20, 21]. Closely related inverse problems arise in ptychography for electron microscopy [22] and X-ray imaging [23], where the goal is to recover complex-valued images from measured diffraction intensities in the presence of missing phase information.

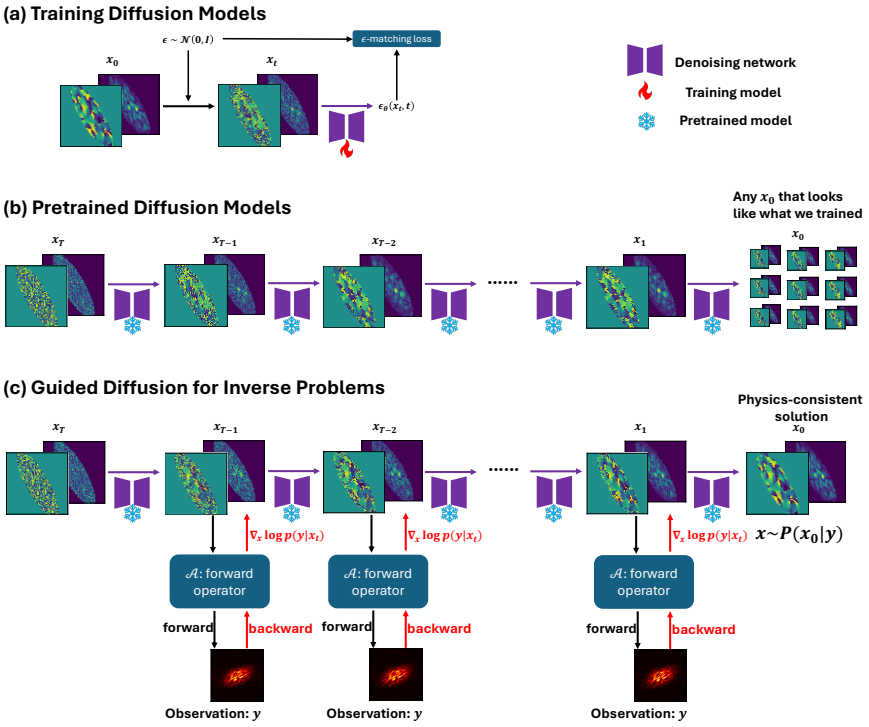

Figure 1: Overview of proposed physics-guided reconstruction framework. **a** During training, clean data samples x_0 are progressively perturbed with Gaussian noise to obtain x_t, and a denoising network is optimized to predict the injected noise through an ϵ-matching loss. **b** Then, the pretrained diffusion model progressively removes noise to synthesize samples x_0 that reflect the learned data distribution. **c** During inference, at each reverse-diffusion step, the model integrates measurement constraints via the forward operator $\mathcal{A}$. Gradients of the likelihood, $\nabla_{x_t} \log p(y|x_t)$, guide the sampling so that the final reconstruction x_0 is both physically consistent with the observation y while adhering to the learned prior.

4 Method

An overview of the proposed physics-guided diffusion reconstruction framework is shown in Figure 1. This section provides a detailed description of the method.

4.1 Diffusion models

Diffusion models [13, 14, 15, 24] are a class of generative models that progressively learn the data distribution by simulating both forward and reverse stochastic processes. The forward process gradually perturbs a data sample $\mathbf{x}_0 \sim p_0$ into Gaussian noise through a stochastic differential equation (SDE):

$$\begin{cases} d\mathbf{x}_t = -\dfrac{\beta_t}{2}\,\mathbf{x}_t\,dt + \sqrt{\beta_t}\,dw_t, & \text{(forward SDE)}, \\ d\mathbf{x}_t = \Big[-\dfrac{\beta_t}{2}\mathbf{x}_t - \beta_t \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\Big]dt + \sqrt{\beta_t}\,d\bar{w}_t, & \text{(reverse SDE)}, \end{cases} \tag{1}$$

where $t \in [0, T]$, β_t denotes the noise schedule, $\mathbf{w}_t$ and $\bar{\mathbf{w}}_t$ are independent Wiener processes, and $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ is the score function [14].

The score is approximated by a neural network $\mathbf{s}_\theta(\mathbf{x}_t, t)$ trained with the denoising score-matching objective [25]:

$$\min_\theta \; \mathbb{E}_{\mathbf{x}_0, t, \mathbf{x}_t \sim p(\mathbf{x}_t|\mathbf{x}_0)} \Big[\big\| s_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) \big\|_2^2 \Big]. \tag{2}$$

Once the score model is learned for all t, integrating the reverse SDE yields a clean sample $\mathbf{x}_0$ starting from pure noise.

4.2 Guided Diffusion models for inverse problems

For an inverse problem, the objective is to sample from the posterior distribution $p(\mathbf{x}|\mathbf{y})$ given a measurement $\mathbf{y}$. This is achieved by modifying the reverse SDE in Eq. (1) to incorporate the posterior score $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y})$ [16]:

$$d\mathbf{x}_t = \Big[-\tfrac{\beta_t}{2}\mathbf{x}_t - \beta_t \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y})\Big]dt + \sqrt{\beta_t}\,d\bar{\mathbf{w}}_t. \tag{3}$$

Using Bayes' rule, the posterior score decomposes as

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t). \tag{4}$$

The prior score $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$ is provided by the pretrained diffusion model. From [16], the likelihood score is approximated via the gradient of a measurement-alignment, such as $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\hat{\mathbf{x}}_0) = -\lambda \nabla_{\mathbf{x}_t} \big[\|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0)\|_2^2\big]$, where λ is a guidance weight and $\mathcal{A}$ denotes the forward operator. $\hat{\mathbf{x}}_0$ is denoised estimate of x_t from Tweedie's formula [24, 26]. This guided reverse diffusion ensures that reconstructed samples remain consistent with the measurement $\mathbf{y}$ while adhering to the prior distribution learned by the diffusion model.

4.3 Dataset and preprocessing

We consider two types of coherent diffraction datasets, including synthetic data and experimental data.

Synthetic data. The first is a synthetic dataset generated according to the procedure in [11]. Each sample contains both the diffraction intensities and the corresponding object image, represented by its amplitude x_{amp} and phase x_{phs}. This dataset is used for training the diffusion model and for evaluating its reconstruction accuracy. The trained model further serves as a pretrained prior for inference on real experimental data. The diffraction patterns are first converted to amplitudes by taking the square root of the measured intensities, $\mathbf{y} = \sqrt{I_{\text{syn}}}$. These amplitudes are then normalized by the global maximum to yield $\tilde{\mathbf{y}} = \frac{\mathbf{y}}{\max_n \mathbf{y}_n}$. On the object side, we also normalize the object images by their global maximum and compute their real and imaginary components using the following transformation

$$\mathbf{x} = \left[x_{\text{re}},\, x_{\text{im}}\right] = \left[x_{\text{amp}} \cos(x_{\text{phs}}),\, x_{\text{amp}} \sin(x_{\text{phs}})\right] \in \mathbb{R}^{2\times 64\times 64}, \tag{5}$$

where $\mathbf{x}$ is the data representation for training the diffusion models. Conversely, to compute the forward model described in Sec . 4.6, the inverse transformation from real and imaginary components back to amplitude and phase is

$$x_{\text{amp}} = \sqrt{x_{\text{re}}^2 + x_{\text{im}}^2}, \qquad x_{\text{phs}} = \tan^{-1}\left(\frac{x_{\text{im}}}{x_{\text{re}}}\right) \tag{6}$$

Together, these transformations provide a bijective mapping between $(x_{\text{re}}, x_{\text{im}})$ and $(x_{\text{amp}}, x_{\text{phs}})$.

Experimental data. The experimental data used for the second test were measured during the same X-ray Free Electron Laser (XFEL) beamtime as our previous work [11]. To summarise, we used the Materials Imaging and Dynamics (MID) instrument of the European XFEL facility [12]. The 8.8 keV photon energy beam was monochromated and focused with compound refractive lenses onto a 75 nm thin-film sample of $La_{1.93}Sr_{0.07}CuO_4$ (LSCO), epitaxially grown on an $LaSrAlO_4$ (LSAO) substrate [27]. The sample was aligned on its (103) Bragg peak, whose diffraction was recorded on the MID adaptive gain integrating pixel detector (AGIPD) situated 8 m from the sample at a 2θ angle of 29°. The diffraction patterns from 140 sequential pulses of X-rays were recorded at a repetition rate of 2.2 MHz. Strong speckles were found to arise from the domain structure of the epitaxial thin film.

In this work, we have chosen to follow pulse train 50 of run 108, which was particularly strong. The data were corrected using the detector calibrations, then background subtracted and cleaned by omitting pulses that correlated with the average with a Pearson coefficient less than 0.25. This yielded 79 out of 140 frames with stronger intensity. These were merged into two groups upon the suggestion that there was some visible change taking place in the middle of the train. Dataset 1 was the sum of the first 24 frames of the train, while dataset 2 was the sum of the last 55 frames of the cleaned sequence. By comparing images

obtained from both datasets, separated in time by about $32\mu s$, we sought to understand what caused this change during the experiment.

Unlike the synthetic dataset, these measurements contain only diffraction intensities and no paired object images. Thus, experimental data are used exclusively for inference with the pretrained diffusion model. Each experimental diffraction pattern has dimensions $I_{\text{exp}} \in \mathbb{R}^{128\times102}$, where non-positive pixels are set to zero. The diffraction patterns were then symmetrically padded along the horizontal direction to a width of 128 and subsequently downsampled by a factor of two in both dimensions to obtain 64×64 images.

3D synthetic data. To evaluate scalability to volumetric phase retrieval, we additionally generate a 3D synthetic dataset of size 32^3. Each 3D object is constructed by superimposing 50 randomly shifted 3D Gaussian blobs with random complex phases inside an ellipsoidal Gaussian support. The support mask is obtained by thresholding a 3D Gaussian field at 0.15 with parameters $(\sigma_x, \sigma_y, \sigma_z) = (6.0, 10.0, 8.0)$ and a fixed in-plane rotation. The diffraction measurement is simulated as the magnitude of the Fourier transform of the masked complex object. In total, 10,000 samples are generated. Following the 2D protocol, we use the raw object amplitude and phase as ground truth and apply global normalization over the training set.

4.4 Diffusion Models and Architecture

We employ a Gaussian diffusion model operating on complex-valued objects represented as two channels (real and imaginary), $\mathbf{x}_0 \in \mathbb{R}^{2\times64\times64}$, and conditioned on measured diffraction data $\mathbf{y} \in \mathbb{R}^{64\times64}$. We adopt a Gaussian diffusion process with $T = 500$ steps. For each training example and a randomly sampled $t \in \{1, \dots, T\}$, we construct a noised input $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}$, where $\bar{\alpha}_t$ is the cumulative product of the noise schedule and $\boldsymbol{\epsilon}$ is standard Gaussian noise. The denoising network then takes the noised input $\mathbf{x}_t$ along with the timestep t and predicts the noise $\boldsymbol{\epsilon}$.

The denoising model processes inputs at a spatial resolution of 64×64 pixels and uses a hierarchical architecture, UNet [28, 15], in which the number of feature channels increases as the spatial resolution decreases. Starting from a base width of $C = 128$ channels, the model expands the feature capacity by factors of $1, 2, 3$, and 4 at successive stages, resulting in feature dimensions of $128, 256, 384$, and 512 channels. At each level, the network applies two residual blocks, and spatial self-attention is applied at the 16×16 resolution to capture long-range spatial relations. Timestep conditioning is provided by a sinusoidal embedding followed by a two-layer MLP of width 512, injected into all residual blocks. The model outputs a 2-channel noise estimate matching the shape of $\mathbf{x}_t$. The final prediction head consists of a normalization layer, a SiLU nonlinearity [29], and a 3×3 convolution.

3D extension. For the 3D experiments, we keep the diffusion posterior sampling algorithm unchanged and replace the 2D UNet backbone with a 3D UNet. All convolutional, residual, and up/down-sampling operations are extended from 2D to 3D using $3 \times 3 \times 3$ kernels, while preserving the hierarchi-

cal encoder–decoder structure and skip connections. The 3D network operates on volumetric inputs of size 32^3 and maintains the same timestep conditioning mechanism as in the 2D model. The model takes a two-channel condition formed by duplicating the 3D diffraction amplitude and predicts the real and imaginary components of the 3D object volume. The output therefore has shape $\mathbf{x}_t \in \mathbb{R}^{2\times32\times32\times32}$, from which amplitude and phase are computed.

4.5 Training

Although diffusion models are traditionally formulated using denoising score matching (Section 4.1), modern implementations commonly adopt the equivalent noise-prediction ($\boldsymbol{\epsilon}$-prediction) parameterization [13], which we follow in our work. This formulation is mathematically equivalent to the score-matching objective, while being simpler to implement and empirically more stable. At each timestep t, the network predicts the Gaussian noise $\boldsymbol{\epsilon}_\theta(x_t, t)$ injected into the noised sample x_t. The training loss is therefore:

$$\mathcal{L}_{\text{train}}(\theta) = \mathbb{E}_{n,t,\boldsymbol{\epsilon}}\Big[\big|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(x_t, t)\big|_2^2\Big]. \tag{7}$$

Training is conducted on an NVIDIA A100 GPU using mixed-precision (FP16) arithmetic to improve computational efficiency. An exponential moving average (EMA) of the model parameters is maintained throughout training, following standard practice in diffusion-model training [15, 13]. We employ the Adam optimizer [30] with learning rate 10^{-4} and batch size 100.

4.6 Inference

At test time, inference is performed using the pretrained diffusion model under a measurement-guided reverse diffusion process. Given an observed measurement $\mathbf{y}$, samples are drawn from the learned reverse diffusion process while enforcing the measurement-alignment constraint introduced in Eq. 4.2. To approximate the likelihood score, we employ the combination of ℓ_1 and ℓ_2 loss to compute measurement-alignment, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\hat{\mathbf{x}}_0) = -\nabla_{\mathbf{x}_t}\big[\lambda_1\|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0)\|_1 + \lambda_2\|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0)\|_2\big]$, where the forward operator $\mathcal{A}(\cdot)$ is defined as

$$\mathcal{A}(\mathbf{x}) = \big\|\mathrm{FFT}\big(x_{\text{amp}}\, e^{j2\pi x_{\text{phs}}}\big)\big\|. \tag{8}$$

Here, $\mathcal{A}(\mathbf{x})$ computes the magnitude of the Fourier transform of the complex-valued object constructed from its amplitude x_{amp} and phase x_{phs}. ℓ_1 loss promotes robustness and considers the inherent sparsity of the diffraction patterns, while the ℓ_2 loss provides stability and encourages smooth convergence during reconstruction. We set $\lambda_1 = \lambda_2 = 0.5$ to balance their contribution equally. Incorporating this measurement consistency into the reverse diffusion process ensures that reconstructed samples align with the physical constraints imposed by $\mathbf{y}$, while remaining consistent with the learned data prior. For real experimental data, reconstructed samples are further refined through gradient-based

measurement-alignment steps applied after the diffusion process. An approximate support constraint is imposed using a Gaussian-shaped binary mask G_{sup}, and refinement is performed for a fixed number of iterations with a step size of 10^{-4} to enhance alignment with the observed diffraction pattern.

4.7 Evaluation Metrics

Reconstruction Errors. To evaluate the consistency between predictions and ground truth, we compute a discrepancy metric between the reconstructed output and the reference. For each of the N samples, let x^i_{pred} denote the reconstructed quantity (e.g., diffraction intensity or objects), and let x_{gt} denote the corresponding ground truth. Following the implementation of χ^2, both quantities are first normalized by their maximum absolute values, and the reconstruction error is defined as

$$\chi^2 = \frac{\sum \left(\tilde{x}^i_{\text{pred}} - \tilde{x}_{\text{gt}}\right)^2}{\sum \left(\tilde{x}^i_{\text{pred}}\right)^2 + \varepsilon}, \tag{9}$$

where $\tilde{x}^i_{\text{pred}} = \frac{x^i_{\text{pred}}}{\max(x^i_{\text{pred}})}$, $\tilde{x}_{\text{gt}} = \frac{x_{\text{gt}}}{\max(x_{\text{gt}})}$, and ε is a small constant for numerical stability. The final error is reported as the average χ^2 over all samples.

Correlation Analysis. The statistical consistency of the reconstructions is evaluated through pairwise correlation matrices for amplitude, phase, and predicted diffraction patterns. For a set of N samples, we construct an $N \times N$ correlation matrix $M_{i,j}$, extracting only the upper triangle values where $1 \leq i < j \leq N$ to avoid redundant self-correlation and duplicate pairs. While the standard Pearson correlation coefficient is used for amplitude and observed diffraction patterns, phase data requires a circular similarity metric to account for its periodic nature. Before comparison, phases are aligned to remove global constant shifts estimated within the object mask. The similarity between two phase images, ϕ_1 and ϕ_2, is computed by representing the pixels as complex numbers on the unit circle:

$$d_{\text{circ}}(\phi_1, \phi_2) = \frac{1}{S} \left| \sum_{j=1}^{S} e^{i(\phi_{1,j} - \phi_{2,j})} \right|, \tag{10}$$

where S denotes the total number of pixels.

Modal Decomposition. To analyze structural reliability, modal decomposition is performed on the $N = 100$ samples. For the amplitude, which consists of strictly non-negative values ($x_{\text{amp}} \geq 0$), Non-negative Matrix Factorization (NMF) [31, 32] is employed to factorize the data matrix $x_{\text{amp}} \in \mathbb{R}^{N \times S}$ into a product of weight matrix W and basis matrix H:

$$x_{\text{amp}} \approx WH, \quad \text{subject to } W, H \geq 0, \tag{11}$$

where $W \in \mathbb{R}^{N\times k}$ represents the weights for each sample and $H \in \mathbb{R}^{k\times F}$ contains the k basis modes. For the phase, we utilize a circular singular value decomposition (SVD) which is constrained within the periodic interval $[-\pi, \pi]$ [33]. The phases are embedded on the unit circle $Z = e^{i\phi} \in \mathbb{C}^{N\times F}$, centered by the pixel-wise circular mean μ_{unit}, and decomposed via SVD on the concatenated real-imaginary embedding,

$$x'_{\text{phs}} = [\text{Re}(Z_c), \text{Im}(Z_c)] = U\Sigma V^T, \tag{12}$$

where $Z_c = Z/\mu_{\text{unit}}$. The k-th physical phase mode ϕ_k is retrieved by projecting the complex samples onto the k-th left-singular vector u_k (the k-th column of U) and reverting the centering process, $\phi_k = \text{Arg}\left(\mu_{\text{unit}} \cdot \sum_{i=1}^{N} u_{k,i} Z_{c,i}\right)$.

5 Results

5.1 Results on synthetic data

Comparison with existing baselines. Figure 2 compares the proposed diffusion posterior sampling method with two representative baselines, including a conventional iterative algorithm, the hybrid input–output (HIO) method [9], and a deterministic data-driven baseline based on UNet. For a fair comparison, both HIO and the proposed method are obtained from 500 iterations of measurement-alignment optimization. Despite this, the HIO reconstruction exhibits noticeable artifacts and larger amplitude and phase errors. Its performance is sensitive to initialization, and due to the inherent ill-posedness of the phase retrieval problem, the reconstructed phase shows significant discrepancies relative to the ground truth. The UNet baseline produces a single deterministic prediction that usually averages over multiple valid solutions consistent with the observation. Although it recovers the overall object structure, this averaging behavior smooths out fine-scale amplitude features and fails to accurately capture phase discontinuities. Such behavior is typical of deterministic neural network predictions in ill-posed inverse problems, where multiple reconstructions can be consistent with the same measurement. Although the proposed method uses the same number of measurement-alignment iterations, it differs from HIO by integrating a learned generative prior that supplies well-informed intermediate estimates during the reverse diffusion process. This results in sharper amplitude reconstructions with improved preservation of fine-scale structures and substantially reduced phase errors. The corresponding error maps show consistently lower errors for both amplitude and phase. Figure 2 (b) presents the distribution of reconstruction metrics in both diffraction and object across 100 predictions. While HIO achieves near-perfect diffraction correlation and minimal diffraction error, this does not lead to accurate object recovery, as evidenced by its low object correlation and broad error distribution. The proposed method maintains low diffraction error due to the diffusion prior, but substantially outperforms HIO in the object, achieving consistently higher correlation and lower error with narrower distributions, indicating both accuracy and stability. Meanwhile,

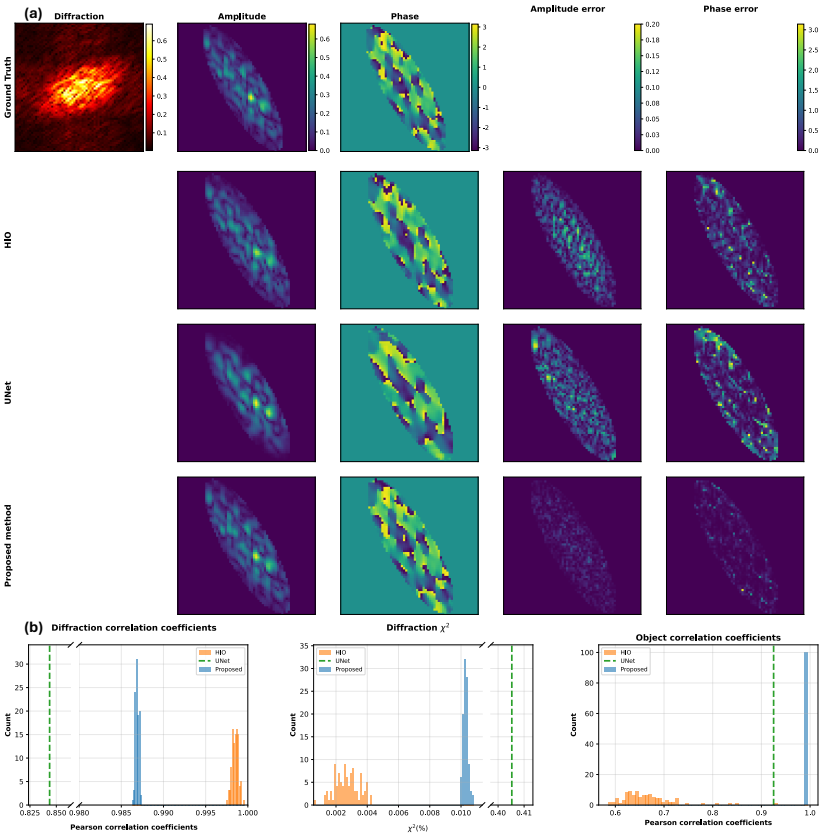

Figure 2: **a** Comparison of reconstruction results for different methods. From left to right: measured diffraction pattern, reconstructed amplitude, reconstructed phase, amplitude error, and phase error. Rows correspond to ground truth, HIO, U-Net, and the proposed method, respectively. The proposed method shows improved fidelity with reduced amplitude and phase errors compared to baseline approaches. **b** Quantitative evaluation of reconstruction performance, including diffraction Pearson correlation coefficients, diffraction χ^2 error, and object-amplitude Pearson correlation coefficients. Despite strong diffraction consistency, HIO exhibits poor object-amplitude reconstruction, whereas the proposed method achieves consistently high correlation and low error, demonstrating robustness; in contrast, UNet performs poorly across all metrics and, being deterministic, yields a single-valued result for each metric.

UNet performs poorly across all metrics and produces single-valued outputs due to its deterministic nature, highlighting its lack of robustness.

Robustness against noise. Figure 3 (a) illustrates the robustness of the proposed diffusion-based method under varying levels of Gaussian noise, $\sigma \in \{0.0, 0.01, 0.05, 0.1, 0.2\}$, applied to the observed diffraction patterns. Despite the progressive degradation of the measurements as the noise level increases, the proposed method consistently recovers the overall amplitude structure and maintains stable phase reconstructions across noise levels up to $\sigma = 0.1$. These results demonstrate that the diffusion-based formulation, which enforces measurement consistency during the reverse diffusion process, remains robust to moderate measurement noise and continues to produce physically plausible reconstructions. At the highest noise level, $\sigma = 0.2$, where the diffraction measurements are severely corrupted, both amplitude and phase reconstructions degrade noticeably, leading to increased reconstruction errors. The quantitative reliability of these reconstructions is further elucidated by the probability density distributions in Figure 3 (b). For noise levels ranging from $\sigma = 0.0$ to $\sigma = 0.05$, the performance distributions for both amplitude correlation and phase similarity are nearly identical, with the majority of the mass highly concentrated above 0.99. However, as the noise increases toward $\sigma = 0.2$, the distributions exhibit a distinct shift toward lower mean values and a notable increase in variance, reflecting a broader spread in reconstruction quality and a graceful degradation of robustness under severe measurement corruption.

5.2 Results on experimental data

Performance on experimental data. As shown in Figure 4, the diffusion-based method achieves consistent reconstructions across both datasets. Although the diffraction patterns from the two temporal windows appear slightly different due to the distinct frame-summing ranges, the recovered amplitude and phase structures remain remarkably similar across multiple sampling instances. Notably, the reconstructed diffraction patterns in both cases align closely with the experimental measurements, with diffraction error maps remaining consistently low throughout the illuminated regions. These results indicate that the diffusion prior acts as a robust preconditioner, successfully guiding the optimization toward a physically plausible solution space and effectively handling the inherent noise and artifacts present in real-world XFEL data.

Statistical and Modal Analysis on experimental data. The performance and stability of the diffusion-based inverse solver were evaluated using a set of 100 samples generated from a single ground-truth diffraction observation. In Figure 5 (a), the solver demonstrates high fidelity in satisfying physical constraints, evidenced by the distribution of correlation coefficients where the observed diffraction patterns are tightly centered near 0.97. In contrast, the reconstructed amplitude and phase distributions exhibit broader variances and lower mean correlations, approximately 0.87 and 0.72, reflecting the inherent ambiguity in phase retrieval where multiple solutions can map to the same diffraction pattern. This high-precision adherence to the forward model is fur-

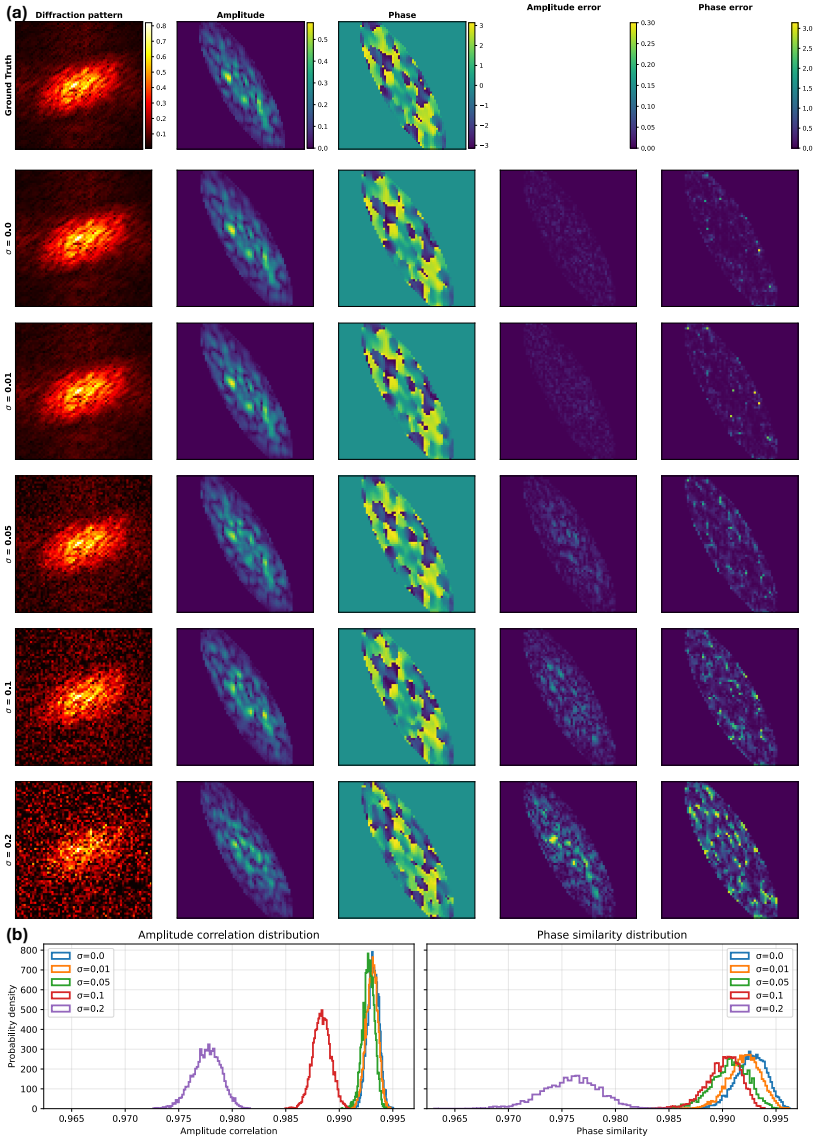

Figure 3: **a** Each subsequent row corresponds to reconstructions obtained from diffraction measurements corrupted by additive Gaussian noise with increasing standard deviation, $\sigma = \{0.0, 0.01, 0.05, 0.1, 0.2\}$. The proposed method maintains stable reconstructions under moderate noise and degrades gracefully as the noise level increases. **b** Probability density distributions for amplitude correlation and phase similarity across the evaluated noise levels. The results remain nearly identical and highly concentrated above 0.99 for $\sigma \leq 0.05$. As σ increases toward 0.2, the distributions shift toward lower mean values and an increase in variance, reflecting a degradation in robustness.

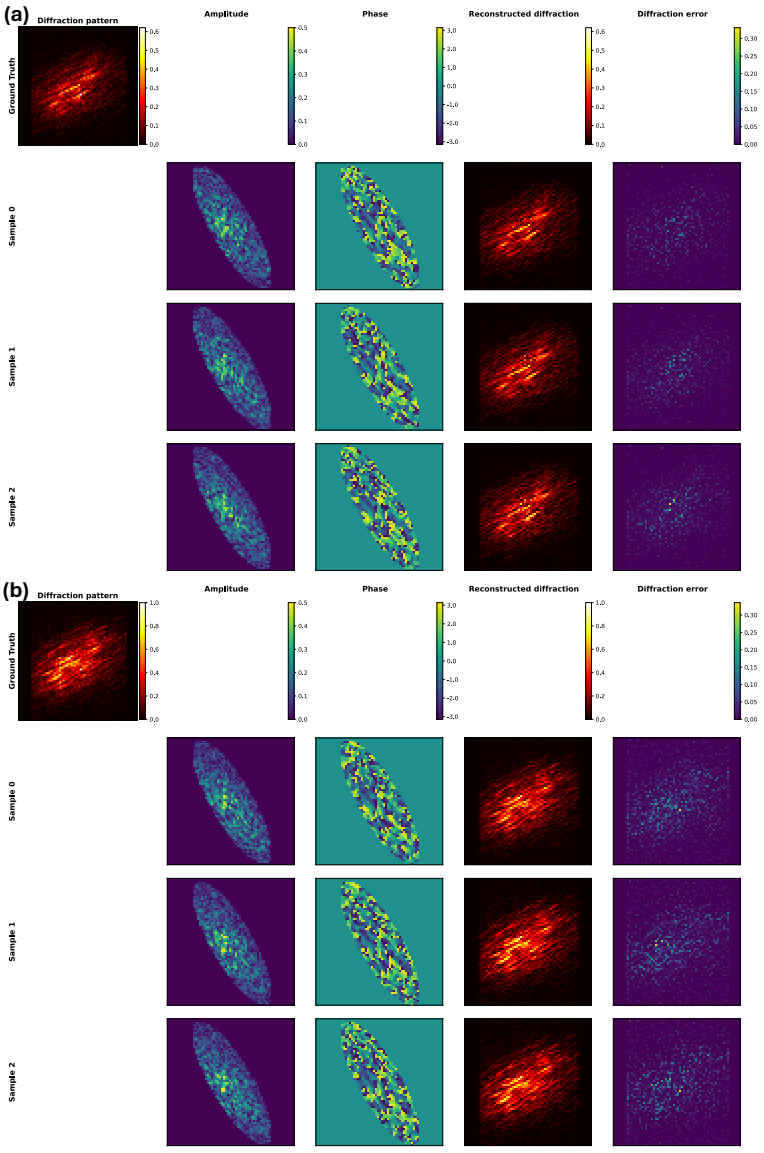

Figure 4: **a** Predictions for three samples from experimental data 1 (sum of diffraction pattern frames 1-24) and **b** data 2 (sum of diffraction pattern frames 25-79), derived from the same experimental pulse but representing different temporal windows. Despite the variations in the input diffraction patterns caused by different frame compression, the recovered amplitudes and phases remain highly similar, with only subtle variations between samples. The reconstructed diffraction patterns and corresponding error maps demonstrate that the errors are consistently low, indicating that the diffusion prior helps provide a robust preconditioner for accurate experimental retrieval.

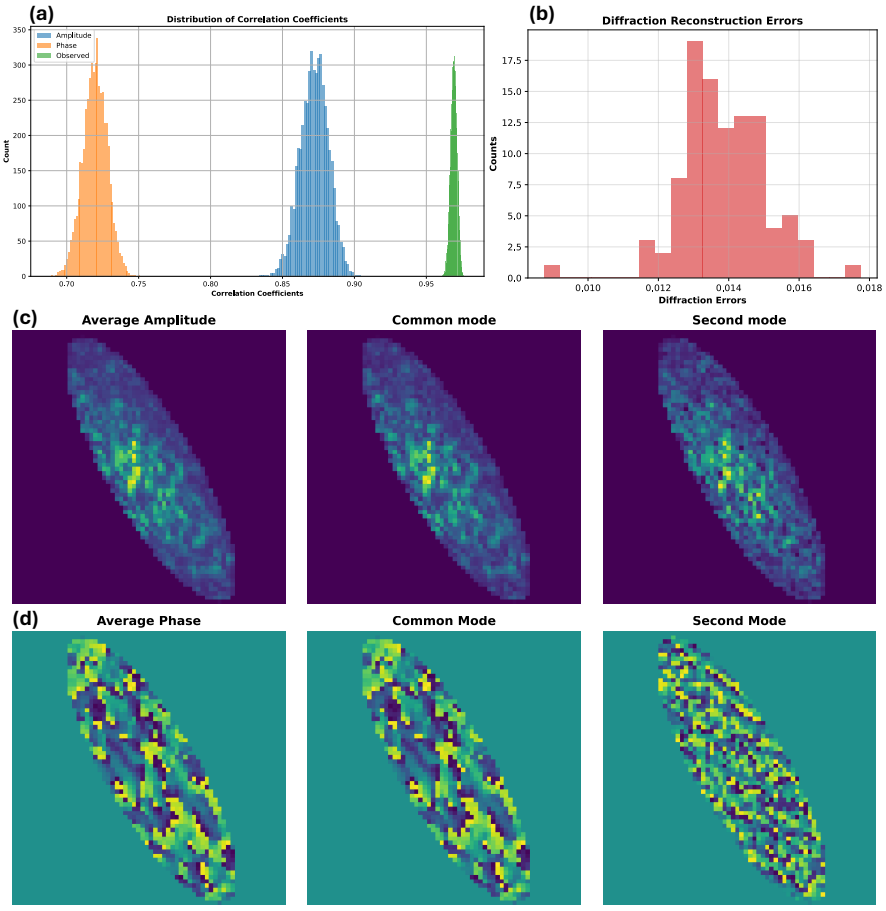

Figure 5: Statistical and modal analyses on experimental data. **a** Histograms of correlation coefficients for amplitude, phase, and reconstructed diffraction patterns across 100 samples. Bin widths are adjusted individually for each distribution to yield comparable histogram heights. **b** Distribution of diffraction reconstruction errors, demonstrating tight convergence to the ground-truth observation. **c** Amplitude modal decomposition via NMF, showing the average amplitude, the dominant common mode, and the second mode. **d** Phase modal decomposition via circular SVD, showing the average phase, the first, and second eigenmodes.

ther confirmed by the diffraction reconstruction errors shown in Figure 5 (b), where the residuals are concentrated between 0.012 and 0.015, indicating robust convergence and numerical reliability.

To provide a more in-depth structural reliability analysis of the reconstructed amplitude and phase, modal decomposition was performed to separate stable features from the stochastic variations inherent in the generative process. For the amplitude, which consists of strictly non-negative values, NMF was employed to identify the dominant structural components. In Figure 5(c), an NMF with $k = 2$ basis captures 90.4% of the total amplitude energy. The leading common mode accounts for 81.9% of the reconstructed energy and closely resembles the averaged amplitude. The second mode contributes the remaining 18.1% of the energy and corresponds to amplitude fluctuations, which are attributed to stochastic variations. For the phase, which is ranged $[-\pi, \pi]$, a circular SVD was employed for modal analysis. In Figure 5(d), the leading phase mode explains 54.9% of the total variance and closely aligns with the circularly averaged phase, indicating a dominant phase structure. The subsequent mode captures a substantially smaller fraction of the variance and exhibits higher spatial frequencies, reflecting the stochastic phase variations.

5.3 Extension to three dimensions

BCDI data are typically measured in 3D to provide 3D spatial images, so we tested whether the proposed method generalizes to 3D by simulating an ellipsoidal object with 50 phase regions, using a method similar to that described in Section 4.3. Figure 6 demonstrates the effectiveness of the proposed diffusion posterior sampling method on the full 3D phase retrieval problem. The predicted object, both amplitude and phase, exhibits strong agreement with the ground truth, with small reconstruction errors across all three spatial dimensions. These results indicate that the proposed method successfully extends from 2D to 3D settings without significant degradation. Importantly, the reconstructed solutions remain consistent with the observed diffraction measurements while preserving physically meaningful 3D structures. This demonstrates that the diffusion posterior framework scales effectively to higher-dimensional inverse problems and can recover both amplitude and phase information in volumetric settings.

5.4 Speed of implementation

We report the average inference time required to reconstruct a single sample from a given diffraction pattern. The HIO and UNet baselines require 0.234 s and 0.264 s per reconstruction, respectively, while our proposed method requires 6.61 s per reconstruction.

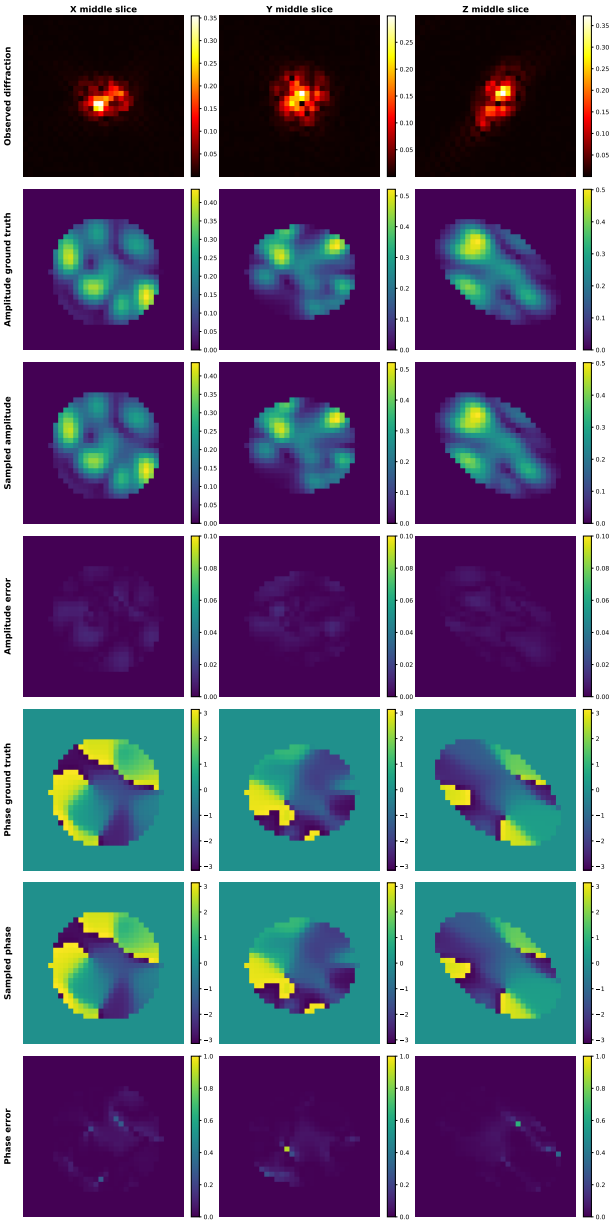

Figure 6: Representative results for the 3D phase retrieval problem. The first row shows the observed diffraction pattern at the X, Y, and Z middle slices. The second, third, and fourth rows present the ground-truth object amplitude, the sampled amplitude generated by the proposed method, and the absolute errors. The fifth, sixth, and seventh rows present the ground-truth phase, the sampled phase generated by the proposed method, and the absolute errors. All quantities are visualized at the central slices along each spatial dimension.

6 Conclusions and Outlook

This work represents an advance in computational solution of the crystallographic "phase problem". As we stated in the introduction, the reconstruction of images of domains from the magnitude of their diffraction pattern is a difficult computational problem. While the data used here and in Bragg Coherent Diffraction Imaging in general are well-oversampled in the sense that there are many more measurement points than unknown pixel values, their inversion in the presence of noise often leads to multiple solutions. Here, we have demonstrated that the diffusion model is able to filter these solutions sufficiently to give reproducible images.

The amplitude images obtained reproducibly in Figure 5 show clusters of nearby domains with a characteristic size of about 300 nm. The clusters are concentrated consistently within a region smaller than the assumed elliptical support area used in the calculation. This is consistent with an image of the beam footprint dropping smoothly off to the sides. Each domain corresponds to a block of flat phase in the phase image obtained as the second output channel of the algorithm; there are sharp steps in phase between the blocks, which represent domain boundaries. These boundaries lie along the minima of the amplitude image. It is expected at finite resolution that destructive interference would lead to amplitude fading close to locations where the phase changes abruptly. We note that the presence of domains and their average size could be estimated from the statistical properties of the speckles in the diffraction pattern. This picture of touching phase domains confirms the domain counting ideas previously published [7].

Domains of this kind are expected in epitaxially grown thin films. Domain physics lies at the heart of many of the theoretical ideas about the macroscopic properties of materials. The scattering of electrons from domain walls is an extrinsic contribution to the resistivity of metals at low temperatures. It might be possible to understand the anomalous transverse resistivity of cuprate thin films discovered a few years ago [27]. Similarly, the dielectric constant of an insulator can be underestimated by ignoring the extrinsic contribution of domain walls. It was discovered that the 90° domain wall structures favored in nanocrystals of $BaTiO_3$ can migrate without needing to change any internal bonds, so they enhance the dielectric constant several-fold, even up to GHz frequencies. This is the foundation of a new Multilayer Ceramic Capacitor industry with a large fraction of the commercial market [34].

In the classical picture of the expected structure of epitaxial thin films[35], islands of the nucleating film material should be locally lattice-matched to the substrate, aligned near their centers. Misfit should build up from the center towards the edges of the domains. When they touch, the domains of the film will no longer be in lattice registration with each other, showing as a phase shift at the domain wall. The lattice parameter difference with the substrate also causes the Bragg peaks of the thin film to be shifted in-plane with respect to the substrate, as is commonly seen[36]. Grain boundaries appear as strain within the film, accommodating any misfit between the film and the substrate.

Because thin films usually have 2D arrays of domains, simple 2D coherent X-ray diffraction patterns of thin-film Bragg peaks contain all the information about the domain structure, as revealed by the methods presented here.

The demonstrated application of the diffusion model is computationally expensive in comparison with neural-network and iterative phasing methods. However, its superior performance in generating reproducible images from different random starting arrays makes it valuable for practical applications. The method is expected to scale to larger array sizes where needed and is expected to work at least as well for three-dimensional data. Its faithful reproducibility behavior may allow it to be used in the future to develop faster algorithms with efficient shortcuts to save on computational resources.

7 Data and Code availability

The experimental data and implementation code used to generate the results presented in this study are available on GitHub [37].

8 Acknowledgements

Work at Brookhaven National Laboratory was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-SC0012704. Work performed at UCL was supported by EPSRC and the ERC. We thank Longlong Wu for developing the codes for data extraction from the XFEL facility. We also thank Jiecheng Diao, Jialun Liu, Jörg Hallmann, Ulrike Boesenberg, Johannes Möller, Anders Madsen, Tadesse Assefa, Emil S. Bozin, Yue Cao, Hoydoo You, Dina Sheyfer, Stephan Rosenkranz, Samuel Marks, Paul Evans, David Keen, and Mark P. M. Dean for help with the measurement of the XFEL data used in this work, previously described in [11]. We thank Ivan Božović for providing the sample.

9 Author Contributions

S.L. and I.R. wrote the main manuscript text. S.L. prepared all figures and conducted the calculations. X.Y. assisted with the generation and processing of simulated data. Y.L. and S.Y. contributed to discussions on improving the model performance. All authors reviewed the manuscript.

References

[1] I. Robinson and R. Harder, "Coherent x-ray diffraction imaging of strain at the nanoscale," *Nature materials*, vol. 8, no. 4, pp. 291–298, 2009.

[2] J. R. Fienup, "Reconstruction of an object from the modulus of its fourier transform," *Optics letters*, vol. 3, no. 1, pp. 27–29, 1978.

[3] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. Spence, "X-ray image reconstruction from a diffraction pattern alone," *Physical Review B*, vol. 68, no. 14, p. 140101, 2003.

[4] L. Wu, S. Yoo, A. F. Suzana, T. A. Assefa, J. Diao, R. J. Harder, W. Cha, and I. K. Robinson, "Three-dimensional coherent x-ray diffraction imaging via deep convolutional neural networks," *npj Computational Materials*, vol. 7, p. 175, 2021.

[5] M. J. Cherukara, Y. S. Nashed, and R. J. Harder, "Real-time coherent diffraction inversion using deep generative networks," *Scientific reports*, vol. 8, no. 1, p. 16520, 2018.

[6] M. Townsend, B. Frosik, H. Taylor, L. Schnebly, R. L. Sandberg, A. Pateras, J. Clark, W. Cha, and R. J. Harder, "High performance phase retrieval code for 3d nanometer scale strain mapping," in *OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP), OSA Technical Digest*, Optica Publishing Group, 2021.

[7] I. Robinson, T. A. Assefa, Y. Cao, G. Gu, R. Harder, E. Maxey, and M. P. Dean, "Domain texture of the orthorhombic phase of $La_{2-x}Ba_xCuO_4$," *Journal of Superconductivity and Novel Magnetism*, vol. 33, no. 1, pp. 99–106, 2020.

[8] R. Bates, "Global solution to the scalar inverse scattering problem," *Journal of Physics A: Mathematical and General*, vol. 8, no. 8, p. L80, 1975.

[9] C.-C. Chen, J. Miao, C. Wang, and T. Lee, "Application of optimization technique to noncrystalline x-ray diffraction microscopy: Guided hybrid input-output method," *Physical Review B—Condensed Matter and Materials Physics*, vol. 76, no. 6, p. 064113, 2007.

[10] C. Atlan, C. Chatelier, and K. Olson, "A python package to help coherent diffraction imaging (cdi) practitioners in their analysis," 2023.

[11] X. Yu, L. Wu, Y. Lin, J. Diao, J. Liu, J. Hallmann, U. Boesenberg, W. Lu, J. Möller, M. Scholz, *et al.*, "Ultrafast bragg coherent diffraction imaging of epitaxial thin films using deep complex-valued neural networks," *npj Computational Materials*, vol. 10, no. 1, p. 24, 2024.

[12] A. Madsen, J. Hallmann, G. Ansaldi, T. Roth, W. Lu, C. Kim, U. Boesenberg, A. Zozulya, J. Moeller, R. Shayduk, M. Scholz, A. Bartmann, A. Schmidt, I. Lobato, K. Sukharnikov, M. Reiser, K. Kazarian, and I. Petrov, "Materials imaging and dynamics (mid) instrument at the european x-ray free-electron laser facility," *Journal of Synchrotron Radiation*, vol. 28, pp. 637–649, 2021.

[13] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[14] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021.

[15] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[16] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *The Eleventh International Conference on Learning Representations*, 2023.

[17] Z. Wu, Y. Sun, Y. Chen, B. Zhang, Y. Yue, and K. Bouman, "Principled probabilistic imaging using diffusion models as plug-and-play priors," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[18] B. Zhang, W. Chu, J. Berner, C. Meng, A. Anandkumar, and Y. Song, "Improving diffusion inverse problem solving with decoupled noise annealing," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20895–20905, 2025.

[19] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *International Conference on Learning Representations*, 2022.

[20] J. Huang, G. Yang, Z. Wang, and J. J. Park, "DiffusionPDE: Generative PDE-solving under partial observation," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[21] J. Yao, A. Mammadov, J. Berner, G. Kerrigan, J. C. Ye, K. Azizzadenesheli, and A. Anandkumar, "Guided diffusion sampling on function spaces with applications to PDEs," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[22] C. Belardi, C.-H. Lee, Y. Wang, J. Lovelace, K. Weinberger, D. Muller, and C. Gomes, "Improving multislice electron ptychography with a generative prior," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3603–3613, 2025.

[23] R. M. Cam, J. Deng, R. Kettimuthu, M. J. Cherukara, and T. Bicer, "Ptychographic image reconstruction from limited data via score-based diffusion models with physics-guidance," in *2025 IEEE 35th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, 2025.

[24] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021.

[25] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[26] B. Efron, "Tweedie's formula and selection bias," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1602–1614, 2011.

[27] J. Wu, A. T. Bollinger, X. He, and I. Božović, "Spontaneous breaking of rotational symmetry in copper oxide superconductors," *Nature*, vol. 547, pp. 432–435, 2017.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[29] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural networks*, vol. 107, pp. 3–11, 2018.

[30] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[31] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[32] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, vol. 13, 2000.

[33] M. Schmidt, S. Rajagopal, Z. Ren, and K. Moffat, "Application of singular value decomposition to the analysis of time-resolved macromolecular x-ray data," *Biophysical journal*, vol. 84, no. 3, pp. 2112–2129, 2003.

[34] A. F. Suzana, S. Liu, J. Diao, L. Wu, T. A. Assefa, M. Abeykoon, R. Harder, W. Cha, E. S. Bozin, and I. K. Robinson, "Structural explanation of the dielectric enhancement of barium titanate nanoparticles grown under hydrothermal conditions," *Advanced Functional Materials*, vol. 33, no. 19, p. 2208012, 2023.

[35] C. V. Thompson, "Structure evolution during processing of polycrystalline films," *Annual review of materials science*, vol. 30, no. 1, pp. 159–190, 2000.

[36] P. F. Fewster, "X-ray diffraction from low-dimensional structures," *Semiconductor science and technology*, vol. 8, no. 11, p. 1915, 1993.

[37] S. Lee, "Diffusion4xray," 2026. Implementation and data for Enhanced single-shot inversion of the diffraction phase problem with the Diffusion model.