Flow-Matching-Based Posterior Sampling for Single-Shot Phase Retrieval in Cryo-EM

Rongtao Zhang [1 2] **Yixiao Yang** [3] **Joel Yeo** [1 2] **N. Duane Loh** [1 2]

[1] *The Centre for BioImaging Sciences, National University of Singapore, Singapore 117551, Singapore* [2] *Department of Physics, National University of Singapore, Singapore 117551, Singapore* [3] *School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China*. Correspondence to: N Duane Loh duaneloh@nus.edu.sg.

1. Introduction

Cryo-electron microscopy (cryoEM) reconstructs three-dimensional biomolecular structure from many two-dimensional projections of particles frozen in vitreous ice [1]. Despite its high-resolution potential, cryoEM is limited by low-dose imaging and by the fact that detectors measure only intensity, while the phase of the complex-valued electron exit-wave—carrying most specimen-dependent information—is not directly observed. Defocus-based imaging partially converts phase variations into intensity contrast, and hardware phase-contrast methods such as phase plates can further improve contrast, but at the cost of additional instrumentation and operational complexity [2].

Modern single-particle cryoEM reconstruction pipelines typically rely on linearized weak-phase, projection-based image formation models summarized by the contrast transfer function [3, 4]. Computational phase retrieval targets this limitation by explicitly recovering the complex exit wave under a physics-based forward model, and cryoEM-specific work has shown that moving beyond these linearized assumptions can mitigate systematic reconstruction artifacts without requiring changes to standard data acquisition pipelines [5]. CryoEM-specific iterative approaches such as Fresnel-zone phase retrieval illustrate both the promise and the practical limitations of classical iterative solvers in low-dose, single-image settings [6]. Nevertheless, phase retrieval remains a nonlinear, noisy and ill-posed inverse problem.

Recent likelihood-based generative models, particularly score-based diffusion models and continuous-time normalizing flows trained via flow matching, have shown that it is possible to learn a powerful prior $p_\theta(x)$ from unpaired samples of the target domain [7, 8, 9]. In these frameworks, a generative model is trained to approximate the data distribution without access to paired measurement data, and at inference time this learned prior can be combined with a differentiable forward model and a likelihood term to sample from the posterior distribution

$$p(x|y) \propto p(y|x)p_\theta(x), \tag{1}$$

using gradient guidance or sampling dynamics that steer the generative process toward measurement consistency. Because this prior is learned independently of the forward operator, this approach avoids the need for task-specific paired (x, y) datasets, and methods such as Diffusion Posterior Sampling and FlowDPS demonstrate how these ideas can be efficiently implemented for general noisy inverse problems [10, 11].

In this work, we propose a single-shot learning-based phase retrieval approach for cryoEM that leverages FlowDPS [11] style posterior sampling with a flow matching prior trained on simulated 2D phase maps and a differentiable cryoEM image formation model. Our method is designed to support phase retrieval from a single noisy measurement without specialized phase-contrast hardware. In our experiments, flow-driven posterior sampling appears to remain promising in a challenging nonlinear, noise-dominated phase retrieval setting, extending beyond the mainly linear inverse-problem regimes emphasized in prior evaluations.

2. Method

2.1 Phase retrieval problem formulation

Let $\mathbf{r}$ denote the real-space image coordinates. The specimen gives rise to a complex-valued exit wave $\psi_{\text{exit}}(\mathbf{r})$. The microscope transfers this wave through a known linear imaging operator, which we express in real space as a convolution with a point spread function $h(\mathbf{r})$,

$$\psi_{\text{img}}(\mathbf{r}) = \psi_{\text{exit}}(\mathbf{r}) \otimes h(\mathbf{r}). \tag{2}$$

The detector records an electron-count image whose statistics are approximated as Poisson:

$$y(\mathbf{r}) \sim \text{Poisson}(\lambda(\mathbf{r})), \qquad \lambda(\mathbf{r}) = \alpha\,|\psi_{\text{img}}(\mathbf{r})|^2, \tag{3}$$

where α converts intensity to expected counts (set by dose and detector gain).

Phase retrieval is the problem of recovering the complex exit wave ψ_{exit} from a single intensity measurement y. This is posed as Bayesian inference,

$$p(\psi_{\text{exit}}|y) \propto p(y|\psi_{\text{exit}})p(\psi_{\text{exit}}), \tag{4}$$

where the likelihood $p(y|\psi_{\text{exit}})$ is defined by the forward model above and $p(\psi_{\text{exit}})$ is a learned prior over physically plausible exit waves.

2.2 Flow prior via flow matching

We learn a generative prior over phase maps $x(\mathbf{r})$, and construct exit-wave candidates via

$$\psi_{\text{exit}}(\mathbf{r}) = \exp(ix(\mathbf{r})). \tag{5}$$

The phase prior $p_\theta(x)$ is modeled as a continuous normalizing flow defined by a time-dependent velocity field $v_\theta(x, t)$, where θ is the set of learnable model parameters.

The vector field is trained using Gaussian flow matching, which specifies an explicit conditional probability path $p(x_t|x_0, x_1)$ connecting samples from a simple base distribution $p_0(x)$ (often chosen to be the standard normal $\mathcal{N}(0, I)$) to samples from the data distribution $p_{\text{data}}(x)$ [9]. Concretely, we use the linear Gaussian conditional path

$$x_t = (1-t)x_0 + tx_1, \quad t \in [0,1], \tag{6}$$

where $x_0 \sim \mathcal{N}(0, I)$ and $x_1 \sim p_{\text{data}}(x)$. Along this path, the conditional velocity is given by the time derivative

$$\dot{x}_t = \frac{dx_t}{dt} = x_1 - x_0, \tag{7}$$

which serves as the target conditional vector field. The network is trained using the conditional flow matching objective [9]:

$$\mathbb{E}_{x_0,x_1,t}\left[\|v_\theta(x_t,t) - (x_1 - x_0)\|^2\right]. \tag{8}$$

After training, samples from the learned prior are obtained by integrating the ordinary differential equation

$$\frac{dx_t}{dt} = v_\theta(x_t, t), \quad t : 0 \to 1, \tag{9}$$

yielding samples $x \sim p_\theta(x)$. This learned prior captures structural regularities of physically plausible phase maps and is used for posterior inference.

2.3 Flow driven posterior sampling

To solve the inverse problem, we draw samples from the posterior

$$p(x|y) \propto p(y|x)p_\theta(x), \tag{10}$$

by modifying the flow dynamics at inference time in the spirit of FlowDPS. FlowDPS motivates posterior sampling in flow models by adding a likelihood-gradient term to the unconditional velocity field, yielding a posterior velocity of the form

$$v_t(x_t|y) = v_t(x_t) - \zeta_t \nabla_{x_t} \log p_t(y|x_t). \tag{11}$$

Since the score function at each timestep $\nabla_{x_t} \log p_t(y|x_t)$ is generally intractable, FlowDPS adopts the DPS approximation $\nabla_{x_t} \log p_t(y|x_t) \approx \nabla_{x_t} \log p\left(y|\mathbb{E}[x_0|x_t]\right)$, i.e., applying data-consistency gradients through a clean (denoised) estimate [11].

Concretely, starting from $x_0 \sim \mathcal{N}(0, I)$, we integrate the learned flow. At time t with state x_t, the model predicts $v_\theta(x_t, t)$, and we form a clean endpoint estimate using the linear-flow identity

$$\hat{x}_1 = x_t + (1-t)v_\theta(x_t, t), \tag{12}$$

which approximates a sample from the data distribution. We then enforce data consistency by performing a small number of gradient descent steps on $\hat{x}_1$,

$$\hat{x}_1 \leftarrow \hat{x}_1 - \zeta_t \nabla_{\hat{x}_1} \left\| y - \left|\exp(i\hat{x}_1(\mathbf{r})) \otimes h(\mathbf{r})\right|^2 \right\|_2^2, \tag{13}$$

and then recombine the updated clean estimate $\hat{x}_1$ with the corresponding noise component before proceeding with the flow integration. This matches the FlowDPS view that posterior sampling can be implemented by adding a likelihood-gradient correction (approximated via a clean estimate) to the unconditional flow dynamics.

2.4 Training data and model

We train the flow prior on a dataset of simulated exit-wave phase maps derived from a known macromolecular structure. Specifically, we use atomic coordinates of the *Thermoplasma acidophilum* 20S proteasome [12]. The dataset contains 20,000 phase maps of size 256×256, which are downsampled to 64×64 via binning.

The velocity field $v_\theta(x, t)$ is parameterized by a convolutional U-Net–style architecture with residual blocks and multi-head self-attention at selected spatial resolutions, following common design choices in diffusion and flow-based generative models [7, 8]. The network is trained using the flow-matching objective described in Section 2.2.

3. Initial Results

We evaluated experimental cryoEM particle images of the Thermoplasma acidophilum 20S proteasome (T20S) from the EMPIAR-10025 dataset [12], down-sampled from 256×256 to 64×64, and compared against a classical iterative phase retrieval baseline [6]. For each measurement, we generate 10 posterior samples and report their average to mitigate occasional hallucinated features. On representative particles (Fig. 1), our method often yields reconstructions that appear higher-contrast and more structured than those from the iterative baseline. Overall, these observations suggest that combining a learned flow prior with FlowDPS-style, physics-guided posterior sampling can be helpful for single-shot phase retrieval in a nonlinear, noise-dominated experimental setting.

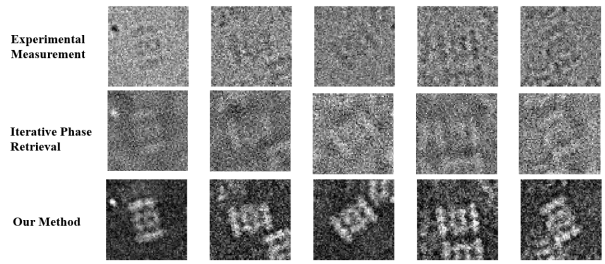

Fig. 1: Single-shot phase retrieval on experimental cryoEM particle images. Top: raw intensity measurements. Middle: classical iterative phase retrieval. Bottom: proposed flow-driven posterior sampling result.

Acknowledgments

The authors would like to acknowledge the computational resources from the NUS Centre for Bio-Imaging Sciences, as well as the NUS AI for Science MSc program.

References

[1] Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz. A primer to single-particle cryo-electron microscopy. *Cell*, 161(3):438–449, 2015.

[2] Hong-Wei Wang and Xiao Fan. Challenges and opportunities in cryo-em with phase plate. *Current Opinion in Structural Biology*, 58:175–182, 2019.

[3] J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, and S. H. W. Scheres. New tools for automated high-resolution cryo-em structure determination in relion-3. *eLife*, 7:e42166, 2018.

[4] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker. cryosparc: algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods*, 14(3):290–296, 2017.

[5] Joel Yeo, Benedikt J. Daurer, Dari Kimanius, Deepan Balakrishnan, Tristan Bepler, Yong Zi Tan, and N. Duane Loh. Ghostbuster: a phase retrieval diffraction tomography algorithm for cryo-em, 2023.

[6] K. Pant, T.-Y. Lan, M. Defrise, and V. Elser. Single-particle cryo-em phase retrieval with fresnel-zone constraints. *arXiv preprint arXiv:2105.12777*, 2021.

[7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

[8] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

[9] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling, 2022.

[10] Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems, 2022.

[11] Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. Flowdps: Flow-driven posterior sampling for inverse problems, 2025.

[12] Melody G. Campbell, David Veesler, Anchi Cheng, Clinton S. Potter, and Bridget Carragher. 2.8 å resolution reconstruction of the Thermoplasma acidophilum 20s proteasome using cryo-electron microscopy. *eLife*, 4:e06380, 2015.