

CryoFlow: Amortized pose inference for Cryo-EM using a normalizing flow-based representation

Andy Cai

Abstract—Cryo-electron microscopy (Cryo-EM) is a technique for determining the 3D structure of molecules, especially macromolecules like proteins. The main computational task is to reconstruct the 3D density of a molecular structure given a series of 2D tomographic projections at random poses. Recent work has applied machine learning to this problem via amortized inference, but has faced issues with non-convergence due to hard to escape local minima. Prior work has analyzed the problem from the perspective of the amortization gap, but in a manner agnostic to the underlying geometric structure of the latent space. We instead analyze the problem as a result of inherent obstructions caused by auto-encoding into a non-trivial topology, and propose a solution using a novel multimodal normalizing flow based representation of pose space. We demonstrate that our representation escapes obstructions at a higher rate compared to a state-of-the-art amortized inference model (CryoAI) in simplified scenarios, and has comparable reconstruction performance on a synthetic dataset.

Index Terms—Cryo-EM, Normalizing Flow, Amortized Inference

1 INTRODUCTION

1.1 Background

Cryo-electron microscopy (Cryo-EM) is a recently popularized technique for determining the 3D structure of molecules, particularly macromolecules like proteins. Compared to X-ray crystallography (the traditional structural imaging technique), it requires less sample material and can does not need significant preparation like crystallization. In addition, recent improvements on both the reconstruction algorithm and microscopy hardware front have allowed the technique to produce 3D structures with unprecedented resolution, which has fueled growing interest [2].

The Cryo-EM technique begins by freezing target molecules in a thin layer of vitreous ice. The layer is then imaged using a transmission electron microscope (TEM), resulting in an image containing many (10^4 - 10^7) 2D tomographic projections (micrographs) of the molecule at random orientations, which are then individually separated. The final and most computationally challenging task is to reconstruct the separated micrographs into a 3D potential density; this is the main focus of our paper.

There are many factors that make this task inherently challenging. For one, the noise levels in real Cryo-EM images are often extremely high, as the strength of the electron beam must be kept low to avoid damaging the molecule. Another challenge is that unlike a typical tomographic reconstruction task, the orientations for each image are unknown and must be estimated. [3]

Traditional state-of-the-art techniques (RELION [4], cryoSPARC [5]) utilize iterative refinement in the form of maximum likelihood expectation-maximization (ML-EM). These require an expensive pose posterior estimation step for each image in the dataset, making them difficult to scale. Recent research tackles this problem using machine learning via amortized inference (cryoAI [6], cryoPoseNet [7]),

replacing the per-image pose estimation component with a global, model-based estimator. However, these methods face challenges in consistently converging to the correct structure, instead sometimes becoming stuck in spuriously symmetric local minima.

Shekarforoush et al. [8] analyze the cause of these local minima as a result of the amortization gap, where the model-based estimator fails to predict the true mode of the posterior. They solve the problem by using a multi-headed encoder that predicts multiple pose candidates, as well as by introducing a 2-stage training process where poses are refined per-image after a certain point. They demonstrate faster and more consistent convergence compared to cryoAI, but don't incorporate geometric inductive bias (e.g. equivariance), which could allow for further improvement.

1.2 Contributions

Building upon prior work from Esmaeili et al. [9] and de Ruijter et al. [10], we analyze the convergence problem as a result of topological obstructions, which occur intrinsically when encoding into a non-trivial topology. In the case of amortized inference, this topology is the group of poses, $SO(3) \times \mathbb{R}^2$. For simplicity, we ignore the translation component of the pose for the rest of the paper, and only work with the group $SO(3)$.

We focus on degree obstructions that occur as a result of the non-simply connected nature of the latent topology. We propose an experiment to measure the ability of models to escape these degree obstructions in a simplified scenario, both when they occur naturally through random initialization or are introduced artificially through adversarial training.

We then propose a novel modification to the cryoAI encoder, where the predicted pose is instead represented

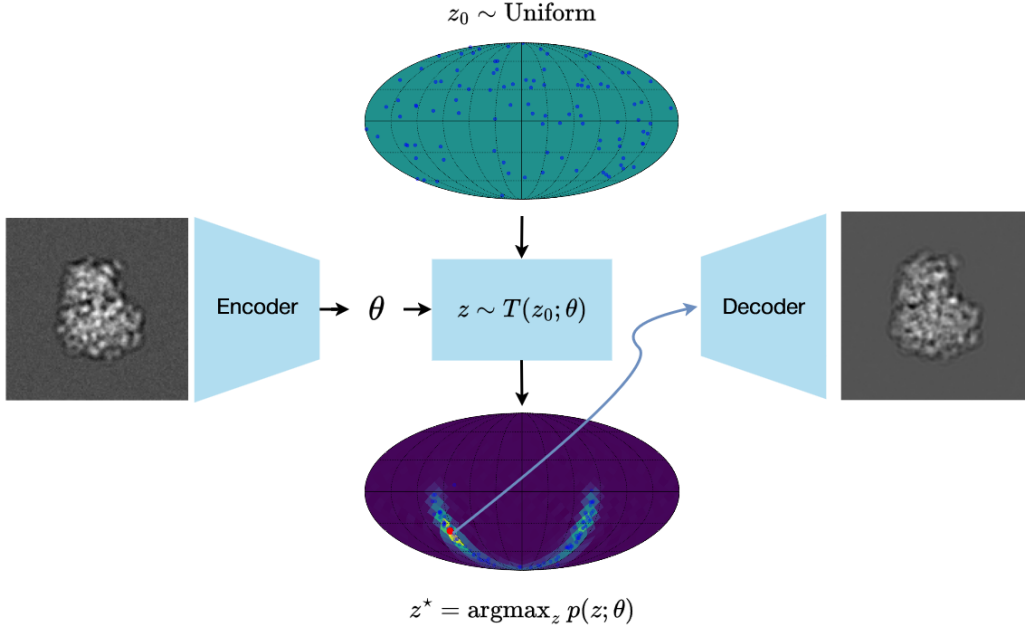


Fig. 1. An overview of our proposed model; the encoder now produces a latent vector that parameterizes a distribution over $\text{SO}(3)$ using normalizing flow, from which we take the mode as our final pose prediction.

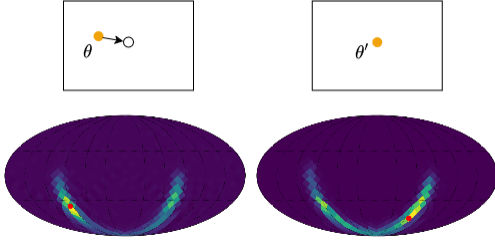


Fig. 2. A visualization of how the mode of the distribution can make a discrete jump with a small change in latent representation. The distributions seen here are marginalized over the view direction and visualized using HEALpix [1].

as the mode of a parameterized distribution, similar to the GroupFlow-VAE architecture proposed by Esmaeili et al. Using normalizing flow, the parameterized distribution can be multi-modal, allowing discrete jumps in the global mode and thus predicted pose to occur (see Fig. 2).

In our simplified experimental scenario, we demonstrate improved ability to escape degree obstructions over cryoAI. We also demonstrate comparable reconstruction performance and similar rates of convergence on a synthetic pre-catalytic spliceosome dataset.

2 RELATED WORK

2.1 Amortized Inference for Cryo-EM

Recent work [6] [7] has applied machine learning to the problem of cryo-EM reconstruction in the form of amortized inference; a network is trained to predict optimal pose of the volume given the image, replacing the expensive posterior estimation step in previous ML-EM methods. This technique can be thought of as analogous to training an auto-encoder

or variational auto-encoder; the pose predictor acts as the encoder, encoding images into the latent space of poses, and the volume estimate acts as the decoder, reconstructing poses back into predicted projections.

Although more efficient, these methods have been observed to exhibit convergence issues, sometimes becoming stuck in spuriously symmetric local minima. Levy et al. hypothesize this is due to the inherent mirror ambiguity of the reconstruction problem, and attempt to mitigate this issue using a “symmetric” loss, which takes the minimal reconstruction loss between the image and its 180° rotation. Although they demonstrate improved convergence, they are unable to fully eliminate the problem of local minima, especially for noisy datasets.

Shekarforoush et al. propose that the problem can be explained by the amortization gap, i.e. the pose predictor failing to accurately predict the true mode of the posterior implied by the decoder’s volume estimate. From this perspective, the symmetric loss introduced by cryoAI is essentially a 2-headed estimator, where the image and its 180° rotation each explore a different region of the posterior, explaining its ability to improve performance. They introduce CryoSPIN, which extends this to an arbitrary number of heads. They also introduce a semi-amortized training scheme, where at some point the model switches to an auto-decoding stage that refines existing pose predictions per-image. However, they lose the geometric inductive bias (in-plane flip equivariance) found in the cryoAI symmetric loss.

de Ruijter et al. instead address the problem by introducing more geometric inductive bias in the form of stronger in-plane rotation equivariance. They use equivariant neural networks to incorporate the bias directly into the encoder, and extend it from flip equivariance to the C_4 (i.e. 90° rotations) and D_4 (i.e. $90^\circ + \text{flip}$) groups. Although they

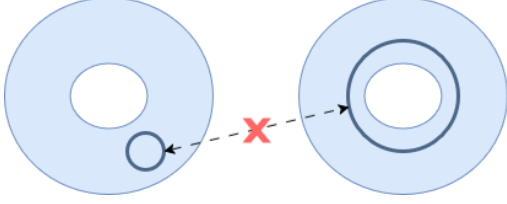


Fig. 3. The blue domain is an example of a non-simply connected space. There is no way to continuously transform the loop on the left to the loop on the right while remaining on the blue domain.

demonstrate improved convergence speed and consistency compared to cryoAI, they still observe convergence problems in some scenarios.

2.2 Topological Obstructions

Falorsi et al. [11] introduce the “homeomorphic encoder”, which is a model that learns a homeomorphic mapping of an input space to a topologically equivalent latent space. Amortized inference in cryo-EM can be thought of as an instance of this, where the latent space topology is described by the group of poses, $SO(3)$.

Esmaili et al. demonstrate that training homeomorphic encoders is inherently difficult due to obstructions at initialization that are inescapable through continuous optimization. Of particular relevance to us are degree obstructions, which occur as a result of the non-simply connected nature of the latent space.

For some intuition about why these obstructions occur, recall that the definition of a non-simply connected space is the presence of distinct classes of loops that can not be continuously transformed between each other (see Fig. 3). In the ideal homeomorphic encoder, loops of one class in the input space would have to be mapped to loops of the same class in latent space.

The problem arises when the encoder incidentally maps a loop from one class into a different class at initialization. Assuming the encoder is continuous in both its parameters and inputs, a continuous path in parameter space that brings the mapping back to the correct class would imply the existence of a continuous deformation between loops of different classes, which by definition can not exist. Therefore, under gradient-flow dynamics, the model becomes trapped in a local minimum.

While the real training process is not actually continuous due to non-zero learning rate, the necessity of discrete jumps to escape the defect adds a degree of difficulty, potentially leading to the convergence problems we see in practice.

To solve this issue, they propose encoding into a latent vector that parameterizes a distribution on the desired latent space using normalizing flow. The latent representation is then chosen to be the mode of the distribution, which they propose is capable of making discontinuous jumps when the underlying distribution is multimodal, allowing the model to escape local minima. They demonstrate improved performance on synthetic datasets with latent space in $SO(2)$ and the torus $SO(2) \times SO(2)$.

Our method extends their technique to $SO(3)$ and applies it to amortized pose inference, to the best of our knowledge is novel.

3 PROPOSED METHOD

3.1 Image Formation Model

The task is to reconstruct the 3D electrostatic potential of the molecule, which we represent as a density map $V : \mathbb{R}^3 \rightarrow \mathbb{R}$. We are given a series of micrographs I_i , formed by electron scattering when the beam hit the molecule at a random pose, which we represent as $R_i \in SO(3)$.

Under the weak-phase object model, the micrograph can be approximately modeled as the 2D tomographic projection of the density. Since the image is formed using a microscope, we also have to account its associated PSF, h_i . We also model noise as an additive Gaussian noise variable ϵ_i . For the sake of simplicity, we have ignored the effect of small translations of the particle in this paper.

$$P_{R_i}[V] : (x, y) \rightarrow \int V(R_i^\top [x \ y \ z]^\top) dz \quad (1)$$

$$I_i(x, y) = (h_i * P_{R_i}[V])(x, y) + \epsilon_i(x, y) \quad (2)$$

3.2 General Pipeline

Much of our architecture is shared with cryoAI, with the main change being to the latent space representation.

As a brief overview, a VGG-based encoder extracts the features from a micrograph image, which are then passed to a fully connected network to produce a latent representation vector. This is then converted into an element of $SO(3)$ using our process, which acts as the pose estimate. For the decoding step, this estimate is fed to an implicit decoder representation which models the Fourier coefficients of the volume as a function from $\mathbb{R}^3 \rightarrow \mathbb{C}$. Using the Fourier slice theorem, we can query this decoder at a central slice to get the Fourier transform of the simulated projection. We then apply the PSF of the microscope to match the image formation process, giving our reconstruction.

3.3 Normalizing Flow Representation

Our representation is the mode of a parameterized distribution $p(z; \theta)$, as suggested by Esmaili et al.:

$$z^*(\theta) = \arg \max_z p(z; \theta) \quad (3)$$

, where $z \in SO(3)$.

Since finding the mode explicitly is expensive, we use a Monte-Carlo estimate, sampling K poses and taking the one with the highest probability. We found $K = 200$ to be a good choice for our experiments, although we did not do an explicit hyperparameter search to determine the best value.

To parameterize the distribution, we use normalizing flow. In essence, we apply a parameterized, invertible transformation T to a base density (in our case, the uniform distribution on $SO(3)$) to shape it into our desired distribution. We can formally describe the process of sampling a value z and its associated log probability $\log p(z)$ from the transformed distribution as follows:

$$z_0 \sim \text{Uniform} \in SO(3) \quad (4)$$

$$z = T(z_0; \theta) \quad (5)$$

$$\log p(z; \theta) = \log p(z_0) - \log |J_{T(\cdot; \theta)}(z_0)| \quad (6)$$

, where $J_{T(\cdot; \theta)}$ is the Jacobian of the transform $T(\cdot; \theta)$.

We utilize the convex gradient transform proposed by Kohler et al. [12]. They work over the space of quaternions for convenience, but ensure that the transformation is flip equivariant (and thus that the distribution is flip invariant), avoiding the need for additional marginalization when converting back to $\text{SO}(3)$.

The transformation is described as:

$$\Phi_{\text{CG}}(\mathbf{x}; \theta) = \frac{\nabla_{\mathbf{x}} \phi(\mathbf{x}; \theta)}{\|\nabla_{\mathbf{x}} \phi(\mathbf{x}; \theta)\|} \quad (7)$$

$$\phi(\mathbf{x}; \mathbf{W}, \mathbf{u}, \mathbf{b}, c) = \mathbf{u}^\top \log(\mathbf{b} + \cosh(\mathbf{W}\mathbf{x})) + c\mathbf{x}^\top \mathbf{x} \quad (8)$$

, where $\mathbf{W} \in \mathbb{R}^{4 \times H}$, $\mathbf{u} \in \mathbb{R}_{>0}^H$, $\mathbf{b} \in \mathbb{R}_{>0}^H$, and $c \in \mathbb{R}_{>0}$.

The total parameterization of the function can then be represented as a concatenation of these subparameters, $\theta \in \mathbb{R}^{6H+1}$. The positive constraint for parameters $\mathbf{u}, \mathbf{b}, c$ is ensured by applying the softmax function to the latter $2H + 1$ components in θ .

H is a hyperparameter that describes how expressive the transform can be. The authors demonstrate that $H = 32$ is already capable of modeling highly multi-modal distributions, so we use that for all of our experiments.

Finally, we compute the on-manifold volume change, $|J_{T(\cdot; \theta)}(z_0)|$, using the procedure from their paper. The off-manifold Jacobian matrix is computed automatically using Pytorch, then projected onto the manifold using the tangent spaces of the input and output vectors. We then compute its determinant, giving us the volume change.

$$|J_{T(\cdot; \theta)}| = \det \left| E_p^\top J_{T(\cdot; \theta)}(p) E_{p'} \right| \quad (9)$$

3.4 Prior Loss

We also experiment with the addition of a prior matching component to the loss to encourage multi-modal behaviour in the parameterized distribution:

$$L_{\text{prior}} = \beta \log p(z^*; \theta) \quad (10)$$

This term penalizes modes that are too concentrated, and encourages the predicted distribution to be closer to the uniform distribution.

4 EXPERIMENTAL RESULTS

4.1 Degree Obstruction

4.1.1 Experiment Setup

Since the goal of these experiments is to explore the behaviour of the encoder, we replace the decoder with a fixed volume. In particular, we use the density of the pre-catalytic spliceosome [13] to match the volume used in our later reconstruction experiments. For simplicity, we also ignore the effect of the microscope's PSF. The task of this simplified encoder is then to learn the homeomorphism from 2D projections to poses:

$$\phi : P_R[V] \rightarrow R \quad (11)$$

For training the encoder, we use the same reconstruction loss objective as the actual model.

To explore degree obstructions, we use easy to generate classes of loops in $\text{SO}(3)$ with known degree. For degree

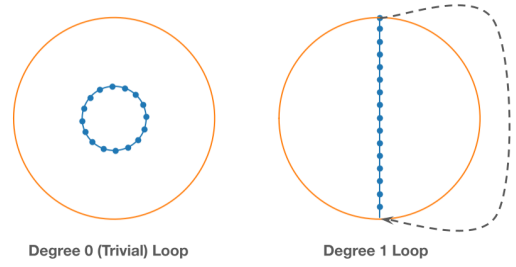


Fig. 4. Degree 0 and degree 1 loops. The loops are visualized in a slice of the axis-angle representation sphere, where the orange circle represents where rotations wrap around (i.e. $\pi \rightarrow -\pi$)

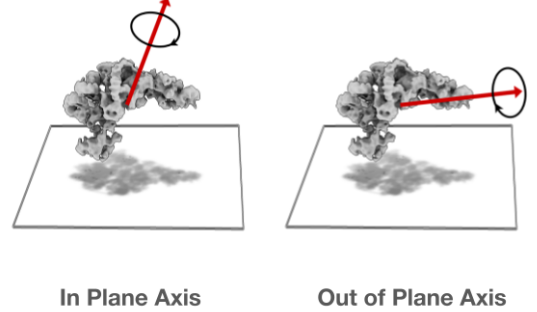


Fig. 5. Visualization of in plane and out of plane degree 1 loops.

0, we use the class of loops defined by circles of radius 1 centered at the origin in axis-angle space. For degree 1, we use the class of loops generated by a 2π rotation around a fixed axis (i.e. a line in axis-angle space). We represent these loops discretely as a sequence of evenly spaced rotations, $\mathcal{L} = \{R_i\}_{i=1}^n$. We use $n = 127$ for our experiments.

We also define a method for measuring the degree of an arbitrary loop. We first define a simple algorithm to map the loop onto the space of quaternions: iterate through each rotation R_i , converting it to a quaternion q_i . This conversion is ambiguous due to the double cover. If this is the first rotation, make an arbitrary choice. Otherwise, pick the choice from $\{q_i, -q_i\}$ that is closest to q_{i-1} .

Then, the degree of a loop can be found by checking whether q_n or $-q_n$ is closest to q_1 . If it is q_n , then the loop has degree 0, otherwise, the loop has degree 1. This is a result of the simply-connected nature of the space of quaternions.

Finally, we define a notion of loop continuity similar to the continuity metric by Esmaeili et al.:

$$d_i = \|q_i - q_{((i-2) \bmod n) + 1}\| \quad (12)$$

$$M = \max_i d_i \quad (13)$$

$$P_\alpha = \alpha\text{th percentile of } \{d_i\}_{i=1}^N \quad (14)$$

$$L_{\text{cont}} = \frac{M}{P_{90}} \quad (15)$$

For all experiments, we test with loops of degree 1 in input space. We generate two main loops; one using an axis with a large in-plane component (i.e. z component), and another with a small component. This is to test the

hypothesis that the in-plane flip equivariance enforced by the symmetric loss helps the model escape obstructions with in-plane rotations. To produce our training data for a given loop, we simply project the fixed volume at the pose defined by its discrete rotations:

$$\{I_i\}_{i=1}^n = \{P_{R_i}[V]\}_{i=1}^n \quad (16)$$

We use two main methods to generate degree obstructions:

- **Random Initialization** Randomly initialize the encoder until it encodes a degree 0 loop for the degree 1 input. This is closer to how defects would occur in training, but estimating the degree can be inconsistent if the model is not smooth at initialization.
- **Induced via Training** Train the encoder directly to produce a degree 0 loop given the degree 1 input by minimizing the MSE of rotation matrices.

After producing a degree obstruction, the encoder is trained for 150 epochs using Adam [14] with a learning rate of 1×10^{-4} . We define the model to have converged if the encoded loop is of degree 1 and has a loop continuity metric below 10. We repeat this process for 16 trials to determine the average convergence rate.

4.1.2 Results

Randomly Initialized Defects Summarized results can be seen in Tab. 1.

For out-of-plane rotations, our representation escapes the defects at a significantly higher rate to cryoAI. A minor improvement is also observed when a component of prior matching loss is added. Removing the symmetric loss from cryoAI also greatly degrades performance, as seen by the complete lack of convergence, suggesting that it plays an important role in allowing cryoAI to escape from defects.

If the defect is resolved, cryoAI seems to converge to a lower L_{cont} than our model, implying better training behaviour when defects are not present.

For in-plane rotations, we observe a significantly higher rate of convergence and lower L_{cont} from cryoAI if symmetric loss is used. However, performance remains poor if symmetric loss is not used, supporting our hypothesis that in-plane equivariance is most relevant for defect loops with a large in-plane component.

Induced Defects Summarized results can be seen in Tab. 2.

For out-of-plane rotations, our representation escapes defects at a higher rate than cryoAI, but the difference is less pronounced than with random initialization. For in-plane rotations, we observe a similar improvement with cryoAI if symmetry loss is used, but also a small but noticeable improvement without symmetry loss. This suggests that the in-plane problem may be inherently easier to escape, but we still observe a much larger improvement when symmetry loss is used, supporting our hypothesis.

4.2 Reconstruction

4.2.1 Experiment Setup

Dataset We use a synthetic dataset of 100 000 128×128 simulated micrographs with SNR = 1. We generate this from

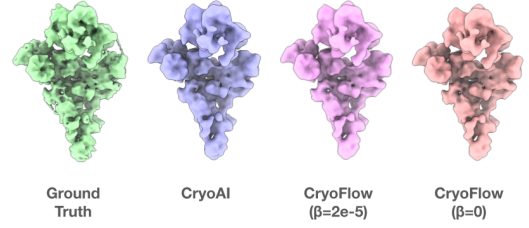


Fig. 6. Reconstructions of each model compared to the ground truth structure. Visualized using ChimeraX

a pre-catalytic spliceosome density map computed from a ground truth structure [13].

Training Hyperparameters All experiments use Adam with a learning rate of 1×10^{-4} for training the encoder and decoder.

Convergence To measure the tendency of the model to become stuck in local minima, we train each model for 3 epochs using 8 random initializations and report the number of runs that reach convergence. Following de Ruijter et al., we define convergence as achieving an MSE below 0.1, where the MSE is computed between the aligned predicted rotation matrices and the ground truth.

Reconstruction Quality We train each model for 60 epochs to measure its performance at convergence. We use a standard metric for reconstruction quality, the Fourier Shell Correlation (FSC) [15], which is a measure of correlation between spherical shells of the reconstruction and ground truth in Fourier space. We find the resolution of the reconstructions using the standard FSC cutoff at 0.5.

4.2.2 Results

Convergence Summarized results can be seen in Tab. 3.

Without a prior matching loss, our model performs worse than cryoAI, converging less often and producing worse MSE at convergence. With the addition of a prior loss, the model performance improves, achieving results similar to cryoAI, but with a slightly worse MSE at convergence.

This suggests that our explanation of convergence failures purely in terms of topological defects is incomplete, as if it was the dominant cause, we might expect a more pronounced improvement in line with our results on our toy example.

Reconstruction Quality Reconstructions are visualized using ChimeraX in Fig. 6. The FSC curves are shown in Fig. 7.

CryoAI produces a higher quality reconstruction, with better resolution and higher correlations in the FSC at all frequency shells. This is somewhat in agreement with the results in our toy experiment, as we observed that our model converged slower than cryoAI once the topological obstruction was resolved. We also found that adding a component of prior matching loss slightly improves reconstruction quality for our model.

5 CONCLUSION

We introduce a new pose representation defined as the mode of a parameterized normalizing flow, and show that

TABLE 1
Results on Randomly Initialized Obstructions

Axis Type	Models	Conv. / Total	Mean L_{cont} @ Conv.
Out of Plane	CryoFlow ($\beta = 0$)	8/16	1.87
	CryoFlow ($\beta = 3 \times 10^{-5}$)	15/16	1.91
	CryoAI	2/16	1.38
	CryoAI (no L_{sym})	0/16	–
In Plane	CryoFlow ($\beta = 0$)	15/16	1.43
	CryoFlow ($\beta = 3 \times 10^{-5}$)	13/16	1.61
	CryoAI	14/16	1.04
	CryoAI (no L_{sym})	2/16	5.62

TABLE 2
Results on Induced Obstructions

Axis Type	Models	Conv. / Total	Mean L_{cont} @ Conv.
Out of Plane	CryoFlow ($\beta = 0$)	12/16	2.12
	CryoFlow ($\beta = 3 \times 10^{-5}$)	14/16	1.87
	CryoAI	8/16	1.08
	CryoAI (no L_{sym})	0/16	–
In Plane	CryoFlow ($\beta = 0$)	14/16	1.50
	CryoFlow ($\beta = 3 \times 10^{-5}$)	16/16	1.47
	CryoAI	14/16	1.12
	CryoAI (no L_{sym})	3/16	1.04

TABLE 3
Convergence Results on Spliceosome Dataset

Models	Conv. / Total	Avg. MSE @ Conv.
CryoFlow ($\beta = 0$)	3/8	0.0327
CryoFlow ($\beta = 2 \times 10^{-5}$)	6/8	0.0209
CryoAI	6/8	0.0148

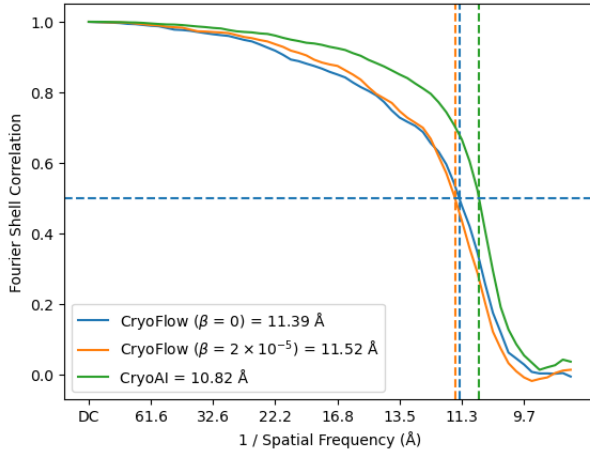


Fig. 7. Fourier Shell Correlation curves for trained models. The resolutions are reported using the 0.5 cutoff.

this representation improves a model’s ability to escape topological defects in a controlled toy setting. We further demonstrate that our method can reconstruct a 3D density from a synthetic dataset with quality comparable to cryoAI, although we do not observe improved convergence behavior.

ior.

Future work should better characterize how topological defects arise and evolve in realistic training scenarios. This could be done by detecting when defects occur in model training, or by extending our toy setup with a trainable decoder to better approximate actual behavior.

REFERENCES

- [1] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, “Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere,” *The Astrophysical Journal*, vol. 622, no. 2, p. 759, 2005.
- [2] T. Bendory, A. Bartesaghi, and A. Singer, “Single-Particle Cryo-Electron Microscopy: Mathematical Theory, Computational Challenges, and Opportunities,” *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 58–76, Mar. 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/9016106/>
- [3] Y. Cheng, N. Grigorieff, P. Penczek, and T. Walz, “A Primer to Single-Particle Cryo-Electron Microscopy,” *Cell*, vol. 161, no. 3, pp. 438–449, Apr. 2015. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0092867415003700>
- [4] S. H. W. Scheres, “RELION: Implementation of a Bayesian approach to cryo-EM structure determination,” *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1047847712002481>

- [5] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination," *Nature Methods*, vol. 14, no. 3, pp. 290–296, Mar. 2017. [Online]. Available: <https://doi.org/10.1038/nmeth.4169>
- [6] A. Levy, F. Poitevin, J. Martel, Y. Nashed, A. Peck, N. Miolane, D. Ratner, M. Dunne, and G. Wetzstein, "CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images," Aug. 2022, arXiv:2203.08138 [cs]. [Online]. Available: <http://arxiv.org/abs/2203.08138>
- [7] Y. S. G. Nashed, F. Poitevin, H. Gupta, G. Woollard, M. Kagan, C. H. Yoon, and D. Ratner, "CryoPoseNet: End-to-End Simultaneous Learning of Single-particle Orientation and 3D Map Reconstruction from Cryo-electron Microscopy Data," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021, pp. 4049–4059, iSSN: 2473-9944. [Online]. Available: <https://ieeexplore.ieee.org/document/9607470>
- [8] S. Shekarforoush, D. B. Lindell, M. A. Brubaker, and D. J. Fleet, "Cryospin: improving ab-initio cryo-em reconstruction with semi-amortized pose inference," *Advances in Neural Information Processing Systems*, vol. 37, pp. 55 785–55 809, 2024.
- [9] B. Esmaeili, R. Walters, H. Zimmermann, and J.-W. v. d. Meent, "Topological Obstructions and How to Avoid Them," Dec. 2023, arXiv:2312.07529 [cs]. [Online]. Available: <http://arxiv.org/abs/2312.07529>
- [10] L. d. Ruijter and G. Cesa, "Equivariant amortized inference of poses for cryo-EM," Jun. 2024, arXiv:2406.01630 [q-bio]. [Online]. Available: <http://arxiv.org/abs/2406.01630>
- [11] L. Falorsi, P. d. Haan, T. R. Davidson, N. D. Cao, M. Weiler, P. Forré, and T. S. Cohen, "Explorations in Homeomorphic Variational Auto-Encoding," Jul. 2018, arXiv:1807.04689 [stat]. [Online]. Available: <http://arxiv.org/abs/1807.04689>
- [12] J. Köhler, M. Invernizzi, P. De Haan, and F. Noé, "Rigid body flows for sampling molecular crystal structures," in *International Conference on Machine Learning*. PMLR, 2023, pp. 17 301–17 326.
- [13] C. Plaschka, P.-C. Lin, and K. Nagai, "Structure of a pre-catalytic spliceosome," *Nature*, vol. 546, no. 7660, pp. 617–621, Jun. 2017. [Online]. Available: <https://doi.org/10.1038/nature22799>
- [14] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [15] M. v. Heel and M. Schatz, "Fourier shell correlation threshold criteria," *Journal of Structural Biology*, vol. 151, no. 3, pp. 250–262, 2005. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1047847705001292>