CryoGAN: A New Reconstruction Paradigm for Single-Particle Cryo-EM *via* Deep Adversarial Learning

Harshit Gupta[1,*,†], Michael T. McCann[2,†], Laurène Donati[1], and Michael Unser[1]

[1]Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

[2]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI, USA; formerly with [1].

We present CryoGAN, a new paradigm for single-particle cryo-EM reconstruction based on unsupervised deep adversarial learning. The major challenge in single-particle cryo-EM is that the imaged particles have unknown poses. Current reconstruction techniques are based on a marginalized maximum-likelihood formulation that requires calculations over the set of all possible poses for each projection image, a computationally demanding procedure. CryoGAN sidesteps this problem by using a generative adversarial network (GAN) to learn the 3D structure that has simulated projections that most closely match the real data in a distributional sense. The architecture of CryoGAN resembles that of standard GAN, with the twist that the generator network is replaced by a model of the cryo-EM image acquisition process. CryoGAN is an unsupervised algorithm that only demands projection images and an estimate of the contrast transfer function parameters. No initial volume estimate or prior training is needed. Moreover, CryoGAN requires minimal user interaction and can provide reconstructions in a matter of hours on a high-end GPU. In addition, we provide sound mathematical guarantees on the recovery of the correct structure. CryoGAN currently achieves a 8.6 Å resolution on a realistic synthetic dataset. Preliminary results on real β-galactosidase data demonstrate CryoGAN's ability to exploit data statistics under standard experimental imaging conditions. We believe that this paradigm opens the door to a family of novel likelihood-free algorithms for cryo-EM reconstruction.

Single-particle cryo-electron microscopy (cryo-EM) is a powerful method to determine the atomic structure of macro-molecules by imaging them with electron rays at cryogenic temperatures [1–3].

[*]Corresponding author.

[†]Equal contribution.

Its popularity has rocketed in recent years, culminating in 2017 with the Nobel Prizes of Jacques Dubochet, Richard Henderson, and Joachim Frank. There exists a multitude of software packages to produce high-resolution 3D structure(s) from 2D measurements [4–11]. These sophisticated algorithms, which include projection-matching approaches and maximum-likelihood optimization frameworks, enable the determination of structures with unprecedented atomic resolution.

Yet reconstruction procedures in single-particle cryo-EM still face complex obstacles. The task involves a high-dimensional, nonconvex optimization problem with numerous local minima. Hence, the outcome of the global process is predicated on the quality of the initial reconstruction [12, 13]. Moreover, one still often relies on the input of an expert user for appropriate processing decisions and parameter tuning [14]. Even for more automated methods, the risk of outputting incorrect and misleading 3D reconstructions is ever-present. A key reason behind such complexity is that the imaged particles have unknown poses. To handle this, most software packages rely on a marginalized maximum-likelihood (ML) formulation [15] that is solved through an expectation-maximization algorithm [9, 11]. The latter involves calculations over the discretized space of poses for each projection, a computationally demanding procedure.

To bypass these limitations, we introduce CryoGAN, an unsupervised reconstruction algorithm for single-particle cryo-EM that exploits the remarkable ability of generative adversarial networks (GANs) to capture data distributions [16]. Similar to GANs, CryoGAN is driven by the competitive training of two entities: one that tries to capture the distribution of real data, and another that discriminates between generated samples and samples from the real dataset (Figure 1). In a classical GAN, the two entities are each a convolutional neural network (CNN). They are known as the generator and the discriminator and are trained simultaneously using backpropagation. The important twist with CryoGAN is that we replace the generator network by a cryo-EM physics simulator. By doing so, CryoGAN learns the 3D density map whose simulated projections are the most consistent with a given dataset of 2D measurements in a distributional sense (see *Online Methods - Mathematical Framework*).

The CryoGAN architecture represents a complete change of paradigm for single-particle cryo-EM reconstruction. No estimation of the poses is attempted during the learning procedure; rather, the reconstruction is obtained through distributional matching performed in a likelihood-free manner. Hence, CryoGAN sidesteps many of the computational drawbacks associated with likelihood-based methods.

In practice, CryoGAN requires no prior knowledge of the 3D structure; its learning process is purely unsupervised. The user needs only to feed the particle images and estimates of the contrast transfer function (CTF) parameters. No initial volume is needed: the algorithm starts with a volume

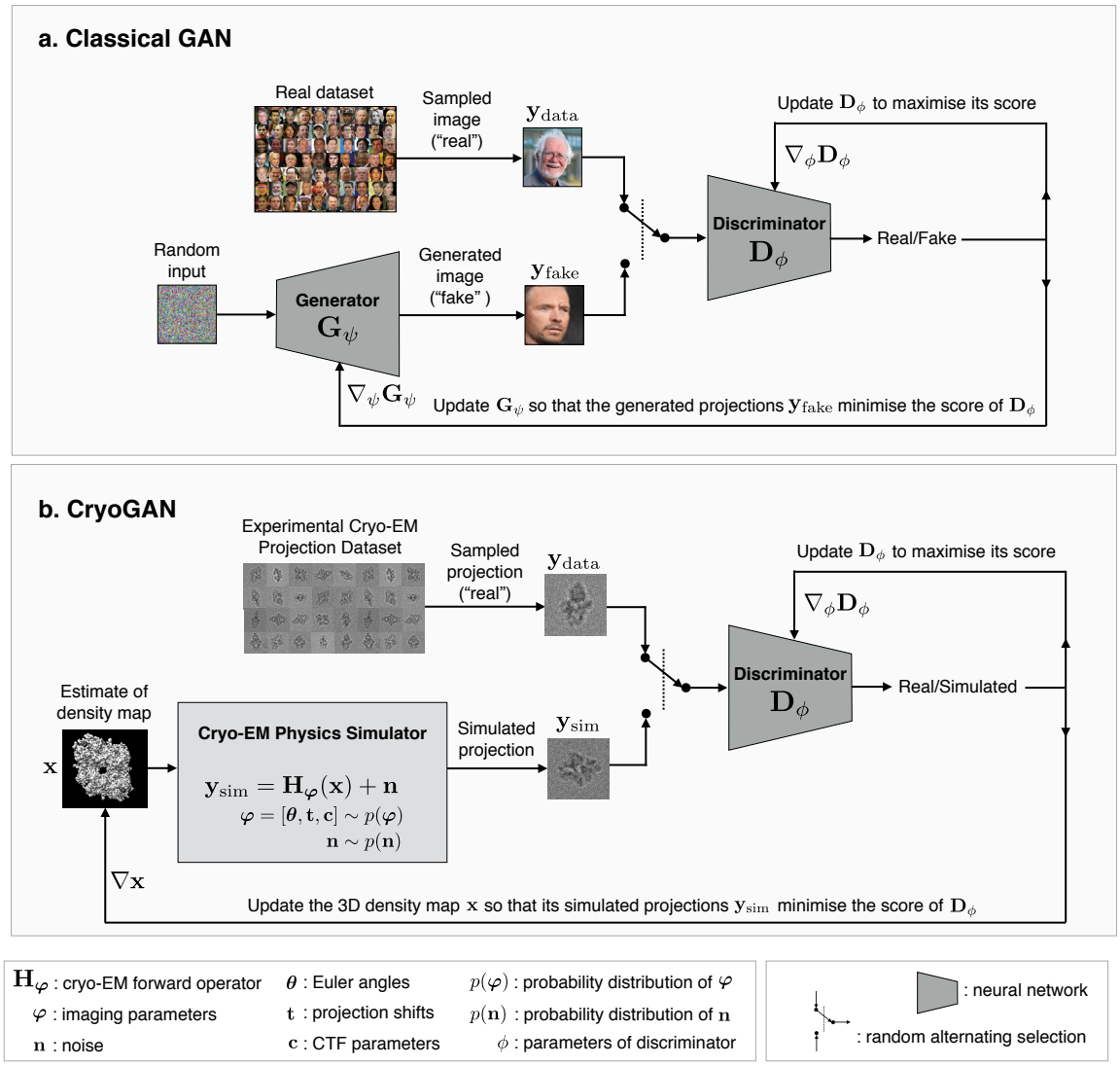


Figure 1: A schematic comparison between **(a)** a classical GAN architecture and **(b)** the CryoGAN architecture. Both frameworks rely on a deep adversarial learning scheme to capture the distribution of the real data. CryoGAN exploits this ability to look for the volume whose simulated projections have a distribution that matches the real data distribution. This is achieved by adding a "cryo-EM physics simulator" that produces measurements following a mathematical model of the cryo-EM imaging procedure. Importantly, CryoGAN does not rely on a first low-resolution volume estimate, but is initialized with a zero-valued volume. Note that, for both architectures, the updates involve backpropagating through the neural networks; those actions are not indicated here for the sake of clarity.

initialized with zeros. The CryoGAN framework is backed up by a comprehensive mathematical framework that provides guarantees on the recovery of the correct structure under a given set of assumptions that are often met in practice, at least to some degree of approximation.

We first assessed the performance and stability of CryoGAN on a synthetic β-galactosidase dataset, where we generated noisy projections *in silico*. The results demonstrate that our unsupervised reconstruction paradigm permits the recovery of a 8.6 Å structure (Figure 2). We then deployed CryoGAN on a real experimental β-galactosidase dataset (EMPIAR-10061) [17], reaching a resolution of 12.1 Å in 150 minutes in far more challenging conditions (Figure 3). These preliminary results provide a strong indication of the suitability of the CryoGAN framework for the reconstruction of real structures. On the implementation side, we expect to be able to improve the resolution of the reconstructions by taking advantage of the many recent technical developments and advances in the area of GANs. In the meantime, the preliminary results obtained with CryoGAN are encouraging and demonstrate the potential of adversarial-learning schemes in image reconstruction. The proposed paradigm opens many new perspectives in single-particle cryo-EM and paves the way for more applications beyond the present one.

RESULTS

The CryoGAN Algorithm

CryoGAN is like a standard GAN, except that the generator network is replaced by a cryo-EM physics simulator (Figure 1b). This simulator implements a mathematical model of the imaging procedure to produce a simulated measurement based on 1) the current volume estimate and 2) a random projection orientation. This image-formation model considers that the cryo-EM 2D measurement is the projection of the volume at that orientation, modulated by microscopy-related effects and corrupted by substantial additive noise.

The cryo-EM physics simulator is paired with a discriminator network whose architecture is similar to those used in standard GANs. The role of this discriminator in CryoGAN is to encourage the simulator to learn the 3D volume $\mathbf{x}_{\text{rec}}$ whose simulated dataset distribution $p(\mathbf{y}|\mathbf{x}_{\text{rec}})$ matches that of the real dataset, $p_{\text{data}}(\mathbf{y})$. Mathematically, this equates to

$$\mathbf{x}_{\text{rec}} = \arg\min_{\mathbf{x}} D\big(p(\mathbf{y}|\mathbf{x}), p_{\text{data}}(\mathbf{y})\big), \tag{1}$$

where D is an appropriate measure of distance between distributions, which the discriminator allows us to compute; in our implementation, D is the Wasserstein distance [18]. This formulation is based

on a sound mathematical framework that provides guarantees on the recovery of the volume. In Theorem 1 (*Supplementary Note: Theoretical Guarantee of Recovery*), we show that, under certain constraints, the reconstructed volume is the same as the true volume up to rotation and reflection.

To perform the minimization (1), the CryoGAN algorithm alternates between updates of the discriminator and of the volume with stochastic gradient descent (SGD). The complete mathematical and algorithmic description of CryoGAN is given in *Online Methods*. The framework was implemented in Python using the PyTorch [19] package.

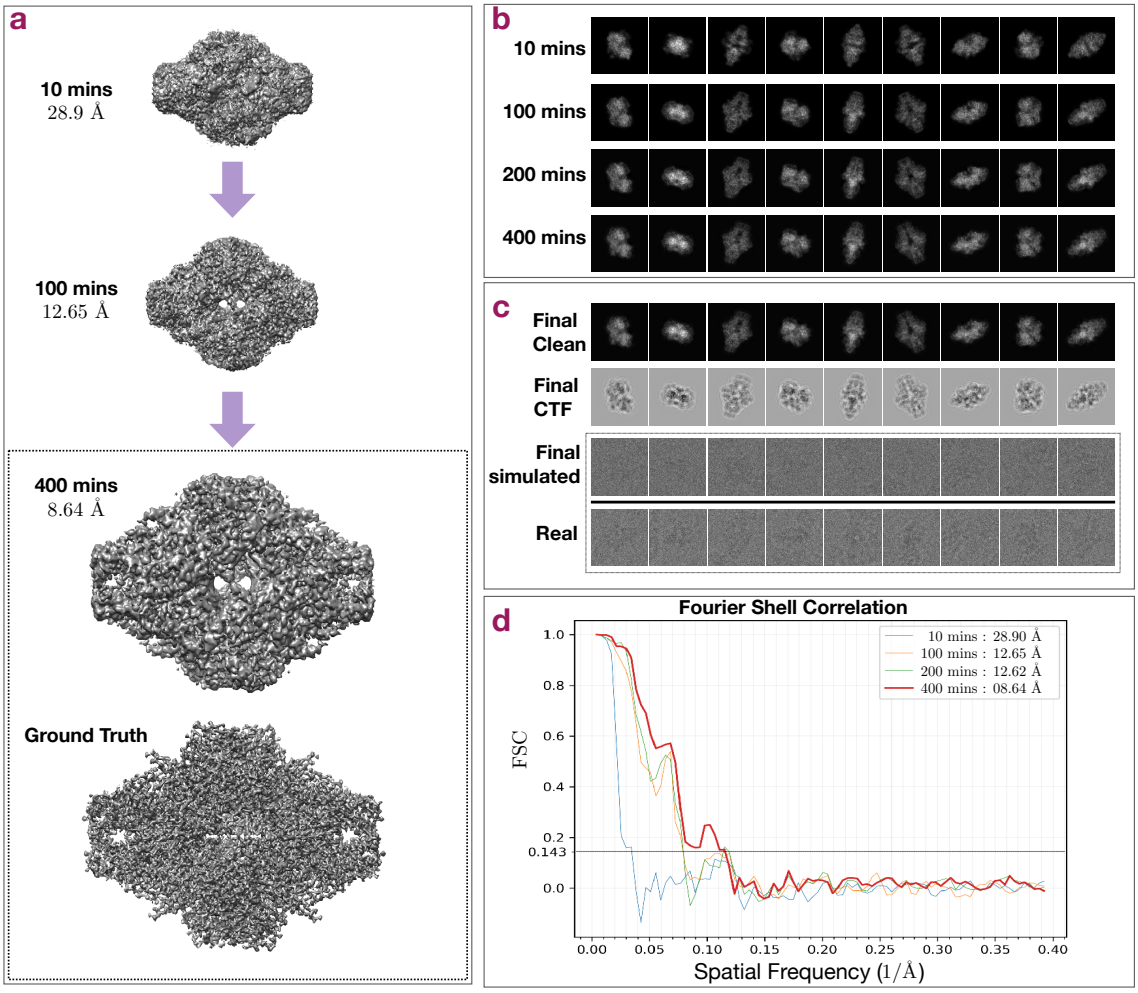


Figure 2: CryoGAN is applied on a synthetic projection dataset generated from a 2.5Å β-galactosidase volume. We refer to these synthetic projections as "real," in contrast to the projections coming from CryoGAN, which we term "simulated." **(a)** The volume is initialized with zeros and is progressively updated to produce projections whose distribution matches that of the real projections. **(b)** Evolution during training of some clean projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. **(c)** *Row 1*: Clean, simulated projections (before CTF and noise) generated at the final stage of training. *Row 2*: CTF-modulated simulated projections (before noise) generated at the final stage of training. *Row 4*: Real projections, for comparison. **(d)** FSC curves between the two reconstructed half-maps at different points during training.

Results on Synthetic Data

We first assessed the viability and performance of CryoGAN on a synthetic dataset that consists of 41,000 β-galactosidase projections, designed to mimic the EMPIAR-10061 data [17] in terms of noise level and CTF parameters. To create this dataset, we generated a 2.5 Å-resolution density map from the PDB entry (*5a1a*) of the protein and applied the forward model described in *Online Methods* to obtain projections modulated by CTF effects and corrupted by noise. We then randomly divided this dataset in two and applied the CryoGAN algorithm separately on both halves to generate half-maps. In the context of this experiment, we refer to these synthetic projections as "real," in contrast to the projections coming from CryoGAN, which we term "simulated." The details of the experimental setup are given in *Supplementary Note: Details of Experiments*.

We ran the CryoGAN algorithm for 400 minutes on an NVIDIA V100 GPU and obtained a reconstruction with a resolution of 8.64 Å (Figure 2a). Starting from a zero-valued volume, CryoGAN progressively updates the 3D structure so that its simulated projections (Figure 2b) reach a distribution that matches that of the real projections. These gradual updates are at the core of the deep adversarial learning scheme of CryoGAN. At each iteration of the algorithm, the gradients from the discriminator carry information about the current difference between the real projections and the simulated projections. These gradients are used by the cryo-EM physics simulator to update the volume so as to improve the fidelity of the simulated projections. Hence, at the end of its run, the volume learned by CryoGAN has simulated projections (Figure 2.c, Rows 1-3) that are similar to the real projections (Figure 2.c, Row 4) in a *distributional* sense. The evolution of the Fourier-shell correlation (FSC) between the reconstructed half-maps (Figure 2.d) testifies to the progressive increase in resolution that derives from this adversarial learning scheme.

We also evaluated the performance of CryoGAN on two additional synthetic datasets: one with a lower noise level and one with off-centered projections. Results were similar to those in Figure 2; further details are given in *Supplementary Discussion: Additional Results on Synthetic Data*.

Results on Experimental Data (EMPIAR-10061)

We then deployed CryoGAN on 41,123 β-galactosidase projections (obtained from EMPIAR-10061 [17]) to assess its capacity to reconstruct real, experimental data. Here as well, we randomly divided the dataset in two and applied CryoGAN separately on both halves. The details of this experimental setup are given in *Supplementary Note: Details of Experiments*.

We ran CryoGAN for 150 minutes to obtain a 12.1 Å-resolution reconstruction using an NVIDIA V100 GPU. The results are displayed in Figure 3. The flexible architecture of CryoGAN permits the straightforward injection of prior knowledge on this specific imaging procedure into the reconstruc-

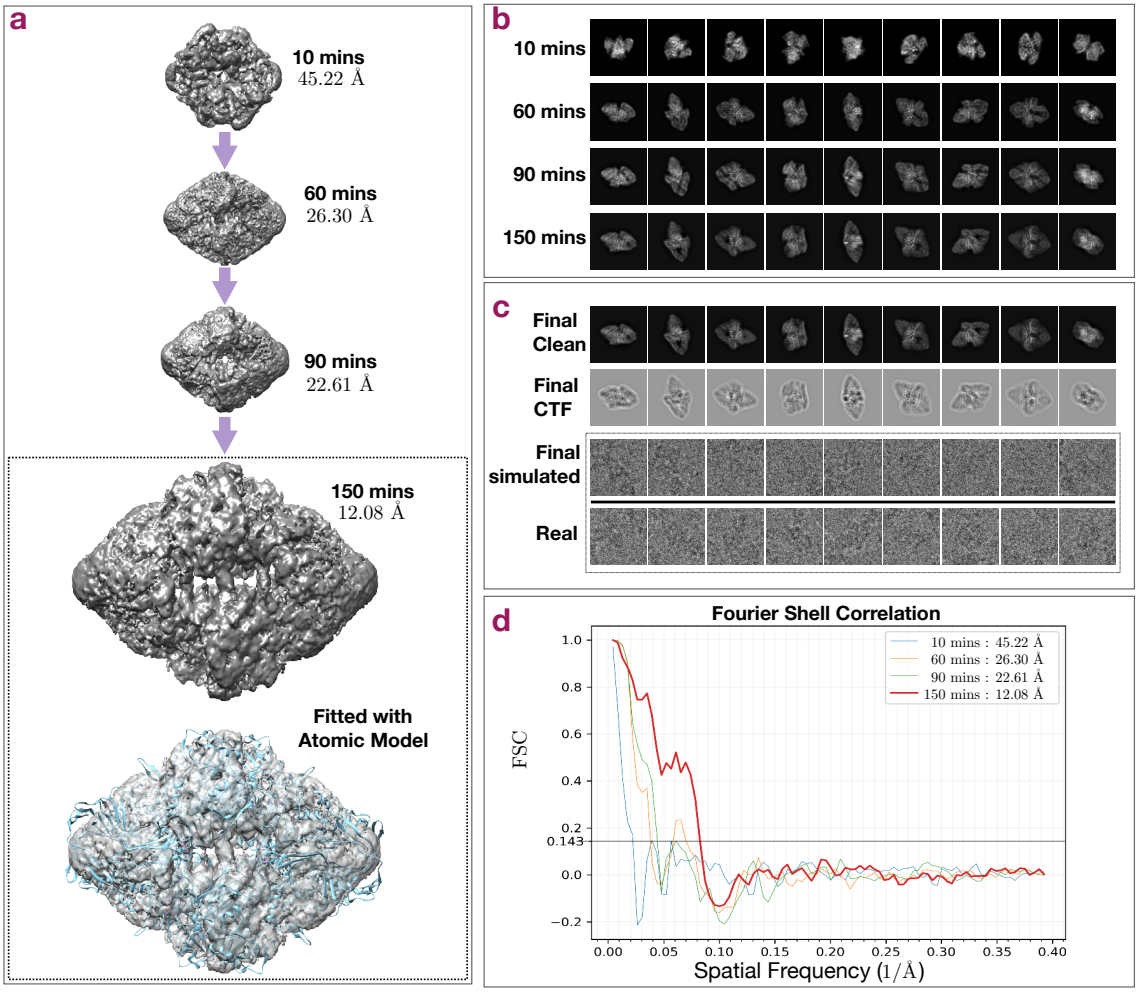


Figure 3: Evolution of CryoGAN while reconstructing the experimental β-galactosidase dataset (EMPIAR-10061) from [17]. **(a)** The volume is initialized with zeros and is progressively updated to produce projections whose distribution matches that of the experimental dataset. **(b)** Evolution during the training of the clean projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. **(c)** *Row 1*: Clean, simulated projections generated at the final stage of training. *Row 2*: CTF-modulated, simulated projections (before noise) generated at the final stage of training. *Row 3*: Simulated projections (with CTF and noise) generated at the final stage of training. *Row 4*: Real projections, for comparison. **(d)** FSC curves between the two reconstructed half-maps at different points during training.

tion pipeline (*e.g.*, the assumption of uniform pose distribution). Using this prior knowledge and its adversarial learning scheme, CryoGAN converges toward the reconstruction that best explains the statistics of the dataset (Figure 3a). As with the synthetic experiments, this is achieved by exploiting the gradients of the discriminator to update the simulator and the current volume estimate, so that, at later iterations, the simulated projections (Figure 3b) follow a distribution that better approaches that of the real dataset. Higher-resolution details are thus progressively introduced in the estimated volume throughout the run, as illustrated by the evolution of the FSC curves between the reconstructed half-maps (Figure 3d). This resulted in a 12.08 Å β-galactosidase structure whose simulated projections closely resemble the real ones (Figure 3c).

DISCUSSION

We demonstrated the ability of CryoGAN to autonomously reconstruct 3D density maps through a purely data-driven adversarial learning scheme, which represents a complete change of paradigm for single-particle cryo-EM reconstruction. Capitalizing on the ability of deep-learning models to capture data distributions, the CryoGAN algorithm looks for the reconstruction that is most consistent with the measurements in a *distributional* sense. Hence, it is able to avoid the angular-assignment procedure altogether by directly exploiting the statistics of the provided dataset. CryoGAN is a purely unsupervised algorithm that requires minimal prior information and user input. It is backed up by a sound mathematical framework that gives guarantees on the recovery of the volume, provided that the image-formation model is valid (see *Online Methods* and *Supplementary Note: Theoretical Guarantee of Recovery*).

An important point is that CryoGAN is likelihood-free, which is in contrast to the iterative-refinement approaches used in most software packages [6, 9, 11]. This permits CryoGAN to bypass marginalization over the angles, a task that is inherent in these likelihood-based approaches, but that is cumbersome because it requires some form of angular assignment and/or the approximation of integrals by sums. CryoGAN also sidesteps many inconvenient processing steps, such as 2D alignment or classification, which further reduces the need for user-dependent inputs.

Our synthetic experiments demonstrate the ability of CryoGAN to resolve a structure so that its simulated projection distribution approaches that of the experimental particles. These results validate the CryoGAN paradigm and the viability of its current implementation. Indeed, without any prior training and starting from a zero-valued volume, the algorithm is able to autonomously capture the relevant statistical information from the dataset of noise-corrupted, CTF-modulated projections, and to progressively learn the volume that best explains these statistics. The results on the real β-galactosidase dataset further demonstrate the capacity of CryoGAN to perform reconstruction in

challenging real imaging conditions.

Roadmap for Future Work

Our implementation of CryoGAN is at the proof-of-concept stage and is bound to further improve. Beyond simple engineering tweaks (*e.g.*, tuning the number of layers in the discriminator, testing different optimization strategies, or using Fourier methods to accelerate projection), we expect that several interesting developmental steps lie ahead.

A promising direction of research is the use of a coarse-to-fine strategy to reconstruct the volume progressively as the resolution improves. The motivation is that increased robustness during the low-resolution regime tends to have a positive impact on the convergence of the higher-resolution steps. Several GAN architectures rely on such approaches, such as the progressive GANs [20] and the styleGANs [21]. The benefits of multi-scale reconstruction could be considerable for CryoGAN, given the extremely difficult imaging conditions that prevail in single-particle cryo-EM and that make the convergence of optimization algorithms to good solutions particularly challenging. The core idea here would be to have the discriminator learn to differentiate between real and simulated distributions at a low resolution first, and then at successively higher ones. The impact on CryoGAN could be as important as the one it had on GANs, which progressed in just a few years from generating blurry facial images [16] to simulated images indistinguishable from real facial images [20, 21]. More generally, the upcoming tools and extensions in GAN architectures could bring significant gain in resolution to the CryoGAN implementation.

The performance of the cryo-EM physics simulator should also improve hand-in-hand with our ability to precisely model the physics behind single-particle cryo-EM with computationally tractable entities. At the moment, CryoGAN assumes that the noise is additive in its image-formation model. One could alternatively consider a Poisson-noise-based forward model [22, 23]. This would however require backpropagating through a Poisson distribution, a nontrivial operation.

Another interesting extension of the simulator would be to directly simulate the patches of nonaligned micrographs/frames (rather than the individual projections) and match their distribution to that of the raw dataset. Doing so would allow CryoGAN to bypass additional preprocessing tasks, in particular particle picking.

Similar to likelihood-based methods, the CryoGAN algorithm requires the specification of the distribution of poses. In the case of CryoGAN, one could also parameterize this distribution and learn its parameters during the reconstruction procedure, along the lines of [24]. The same approach could be used to calibrate the distribution of the translations of the projections.

On the theoretical side, we currently have mathematical guarantees on the recovery of volumes

for which the assumed distribution of poses (be it uniform or not) matches the distribution of the real data. We have prior mathematical indications that this can also be achieved when there is a certain mismatch between the assumed distribution of poses and the actual one, given that an appropriate GAN loss is used.

Like all reconstruction algorithms, CryoGAN can fail if the dataset contains too many corrupted particle images, *e.g.*, those with broken structures or strong optical aberrations. Several solutions could be deployed to handle excessive outliers in the data distribution. One approach would be to include a step that automatically spots and discards corrupted data so that the discriminator never sees them. Recent deep-learning-based approaches able to track outliers in data could prove useful in this regard [25].

Finally, the flexible architecture of CryoGAN should also facilitate its extension to the handling of structures with a continuum of conformational states, a feat often perceived as the greatest technical challenge ahead for single-particle cryo-EM [26]. A route of study we are currently exploring is to parametrize the different 3D conformations as outputs of an appropriate neural network, similar to the strategy used in ambient-GANs [27]. Hence, CryoGAN would learn the set of conformations whose projections are the most consistent with the acquired dataset, in a distributional sense here as well. Importantly, this would not only permit the recovery of the different conformations, but also give a probability on the structure of interest being in any given state in its conformational landscape, which could provide important biological insights.

While the spatial resolution of the CryoGAN reconstructions from real data is not yet competitive with the state-of-the-art, the algorithm is already able to steadily perform the hardest part of the job, which is to obtain a reasonable structure by using nothing save the particle dataset and CTF estimations. We believe that the aforementioned developments will help to bring the CryoGAN algorithm to the stage where it becomes a relevant contributor for high-resolution reconstruction in single-particle cryo-EM. We have laid out a roadmap for future improvements that should get us to this stage, and may eventually help us reconstruct dynamic structures.

Acknowledgements

The authors would like to warmly thank Dr. Ricardo Righetto (University of Basel), Dr. Ricardo Adaixo (University of Basel), Prof. Henning Stahlberg (University of Basel, EPFL), and Dr. Sergey Nazarov (EPFL) for insightful discussions on single-particle cryo-EM. They are also thankful to Shayan Aziznejad (EPFL) and Dr. Quentin Denoyelle (EPFL) for useful feedback on mathematical developments.

This research was supported by the European Research Council (ERC) under the European

Union's Horizon 2020 research and innovation programme, Grant Agreement No. 692726 GlobalBioIm: Global integrative framework for computational bio-imaging.

References

[1] Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* **21**, 129–228 (1988).

[2] Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of molecular biology* **213**, 899–929 (1990).

[3] Frank, J. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state* (Oxford University Press, 2006).

[4] Frank, J., Shimkin, B. & Dowse, H. Spider—a modular software system for electron image processing. *Ultramicroscopy* **6**, 343–357 (1981).

[5] Sorzano, C. *et al.* Xmipp: A new generation of an open-source image processing package for electron microscopy. *Journal of Structural Biology* **148**, 194–204 (2004).

[6] Tang, G. *et al.* Eman2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology* **157**, 38–46 (2007).

[7] Grigorieff, N. Frealign: High-resolution refinement of single particle structures. *Journal of Structural Biology* **157**, 117–125 (2007).

[8] Hohn, M. *et al.* Sparx, a new environment for cryo-em image processing. *Journal of Structural Biology* **157**, 47–55 (2007).

[9] Scheres, S. H. Relion: Implementation of a bayesian approach to cryo-em structure determination. *Journal of Structural Biology* **180**, 519–530 (2012).

[10] de la Rosa-Trevín, J. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology* **195**, 93 – 99 (2016).

[11] Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296 (2017).

[12] Sorzano, C. O. S., Marabini, R., Pascual-Montano, A., Scheres, S. H. & Carazo, J. M. Optimization problems in electron microscopy of single particles. *Annals of Operations Research* **148**, 133–165 (2006).

[13] Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).

[14] Bendory, T., Bartesaghi, A. & Singer, A. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Processing Magazine* **37**, 58–76 (2020).

[15] Singer, A. & Sigworth, F. J. Computational methods for single-particle electron cryomicroscopy. *Annual Review of Biomedical Data Science* **3** (2020).

[16] Goodfellow, I. *et al.* Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680 (2014).

[17] Bartesaghi, A. *et al.* 2.2 å resolution cryo-em structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147–1151 (2015).

[18] Villani, C. *Optimal transport: old and new*, vol. 338 (Springer Science & Business Media, 2008).

[19] Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 8024–8035 (2019).

[20] Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv:1710.10196* (2017).

[21] Karras, T., Laine, S. & Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4401–4410 (2019).

[22] Vulović, M. *et al.* Image formation modeling in cryo-electron microscopy. *Journal of structural biology* **183**, 19–32 (2013).

[23] Rullgård, H., Öfverstedt, L.-G., Masich, S., Daneholt, B. & Öktem, O. Simulation of transmission electron microscope images of biological specimens. *Journal of microscopy* **243**, 234–256 (2011).

[24] Sharon, N., Kileel, J., Khoo, Y., Landa, B. & Singer, A. Method of moments for 3-d single particle ab initio modeling with non-uniform distribution of viewing angles. *Inverse Problems* (2019).

[25] Miolane, N., Poitevin, F., Li, Y.-T. & Holmes, S. Estimation of orientation and camera parameters from cryo-electron microscopy images with variational autoencoders and generative adversarial networks. *arXiv:1911.08121* (2019).

[26] Frank, J. New opportunities created by single-particle cryo-em: The mapping of conformational space. *Biochemistry* **57**, 888–888 (2018).

[27] Bora, A., Price, E. & Dimakis, A. G. AmbientGAN: Generative models from lossy measurements. *ICLR* **2**, 5 (2018).

[28] Frank, J. *Electron tomography: methods for three-dimensional visualization of structures in the cell* (Springer Science & Business Media, 2008).

[29] Natterer, F. *The mathematics of computerized tomography* (Society for Industrial and Applied Mathematics, 2001).

[30] Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223 (2017).

[31] Peyré, G., Cuturi, M. *et al.* Computational optimal transport. *Foundations and Trends® in Machine Learning* **11**, 355–607 (2019).

[32] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. Improved training of wasserstein gans. In *Advances in neural information processing systems*, 5767–5777 (2017).

[33] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[34] van Aarle, W. *et al.* The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **157**, 35–47 (2015).

[35] Penczek, P. A., Grassucci, R. A. & Frank, J. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* **53**, 251 – 270 (1994).

[36] Baker, T. & Cheng, R. A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *Journal of Structural Biology* **116**, 120–130 (1996).

[37] Sigworth, F. J. A maximum-likelihood approach to single-particle image refinement. *Journal of structural biology* **122**, 328–339 (1998).

[38] Sigworth, F. J., Doerschuk, P. C., Carazo, J.-M. & Scheres, S. H. An introduction to maximum-likelihood methods in cryo-em. In *Methods in enzymology*, vol. 482, 263–294 (Elsevier, 2010).

[39] Kam, Z. The reconstruction of structure from electron micrographs of randomly oriented particles. In *Electron Microscopy at Molecular Dimensions*, 270–277 (Springer, 1980).

[40] Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical image analysis* **42**, 60–88 (2017).

[41] McCann, M. T., Jin, K. H. & Unser, M. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine* **34**, 85–95 (2017).

[42] Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).

[43] Bepler, T., Noble, A. J. & Berger, B. Topaz-denoise: general deep denoising models for cryoem. *bioRxiv* 838920 (2019).

[44] Wang, F. *et al.* Deeppicker: a deep learning approach for fully automated particle picking in cryo-em. *Journal of structural biology* **195**, 325–336 (2016).

[45] Zhu, Y., Ouyang, Q. & Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC bioinformatics* **18**, 348 (2017).

[46] Tegunov, D. & Cramer, P. Real-time cryo-em data pre-processing with warp. *BioRxiv* 338558 (2018).

[47] Wagner, T. *et al.* Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em. *Communications Biology* **2**, 218 (2019).

[48] Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods* 1–8 (2019).

[49] Zhong, E. D., Bepler, T., Davis, J. H. & Berger, B. Reconstructing continuous distributions of 3D protein structure from cryo-em images. In *International Conference on Learning Representations* (2020).

[50] Jin, K. H., McCann, M. T., Froustey, E. & Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing* **26**, 4509–4522 (2017).

[51] Adler, J. & Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems* **33**, 124007 (2017).

[52] Gupta, H., Jin, K. H., Nguyen, H. Q., McCann, M. T. & Unser, M. CNN-based projected gradient descent for consistent ct image reconstruction. *IEEE transactions on medical imaging* **37**, 1440–1453 (2018).

[53] Tewari, A. *et al.* State of the art on neural rendering. *arXiv preprint arXiv:2004.03805* (2020).

[54] Tulsiani, S., Efros, A. A. & Malik, J. Multi-view consistency as supervisory signal for learning shape and pose prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2897–2905 (2018).

[55] Gadelha, M., Maji, S. & Wang, R. 3D shape induction from 2D views of multiple objects. In *2017 International Conference on 3D Vision (3DV)*, 402–411 (IEEE, 2017).

[56] Mohamed, S. & Lakshminarayanan, B. Learning in implicit generative models. *arXiv preprint arXiv:1610.03483* (2016).

[57] Vulović, M. *et al.* Image formation modeling in cryo-electron microscopy. *Journal of Structural Biology* **183**, 19–32 (2013).

[58] Unser, M. Sampling—50 years after Shannon. *Proceedings IEEE* **88**, 569–587 (2000).

[59] Panaretos, V. M. *et al.* On random tomography with unobservable projection angles. *The Annals of Statistics* **37**, 3272–3306 (2009).

[60] Helgason, S. *The radon transform*, vol. 2 (Springer, 1980).

[61] Pettersen, E. F. *et al.* Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605–1612 (2004).

[62] Rohou, A. & Grigorieff, N. Ctffind4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* **192**, 216–221 (2015).

[63] Biyani, N. *et al.* Focus: The interface between data collection and data processing in cryo-em. *Journal of structural biology* **198**, 124–133 (2017).

[64] Righetto, R. D., Biyani, N., Kowal, J., Chami, M. & Stahlberg, H. Retrieving high-resolution information from disordered 2d crystals by single-particle cryo-em. *Nature communications* **10**, 1–10 (2019).

ONLINE METHODS

Image-Formation Model in Single-Particle Cryo-EM

We detail here the cryo-EM image formation model used in our implementation of the CryoGAN physics simulator.

We use the standard model for the single-particle cryo-EM imaging procedure [28]

$$\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^M$ is a (vectorized) 2D projection; $\mathbf{x} \in \mathbb{R}^V$ is the (vectorized) 3D density; $\mathbf{H}_{\boldsymbol{\varphi}} \in \mathbb{R}^{M\times V}$ is the forward operator with parameters $\boldsymbol{\varphi}$; and $\mathbf{n} \in \mathbb{R}^M$ is an additive noise following a distribution $p_{\mathbf{n}}$. The imaging parameters $\boldsymbol{\varphi}$ comprise the projection (Euler) angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, the projection shifts $\mathbf{t} = (t_1, t_2)$, and the CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$, where d_1 is the defocus-major, d_2 is the defocus-minor, and α_{ast} is the angle of astigmatism.

The forward operator $\mathbf{H}_{\boldsymbol{\varphi}}$ is given by

$$\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C}_{\mathbf{c}}\mathbf{S}_{\mathbf{t}}\mathbf{P}_{\theta}, \tag{3}$$

where $\mathbf{P}_{\boldsymbol{\theta}} : \mathbb{R}^V \to \mathbb{R}^M$ is a projection operator (mathematically speaking, the X-ray transform [29]), $\mathbf{S}_{\mathbf{t}} : \mathbb{R}^M \to \mathbb{R}^M$ is a shift operator, and $\mathbf{C}_{\mathbf{c}} : \mathbb{R}^M \to \mathbb{R}^M$ is a convolution operator. A more detailed description of the physics behind this image formation model $\mathbf{H}_{\boldsymbol{\varphi}}$ is given in *Supplementary Note: Image Formation*.

Mathematical Framework of CryoGAN

The goal of single-particle cryo-EM reconstruction is to estimate a 3D density map $\mathbf{x}_{\text{rec}}$ whose projections are consistent with the observed projections (data) of the true density map $\mathbf{x}_{\text{true}}$.[1]

We write the conditional probability density function of a measurement $\mathbf{y}$ given a volume $\mathbf{x}$ by marginalizing over the imaging parameters $\boldsymbol{\varphi}$

$$p(\mathbf{y}|\mathbf{x}) = \int_{\boldsymbol{\varphi}\in\Phi} p_{\mathbf{n}}(\mathbf{y} - \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x})p_{\boldsymbol{\varphi}}(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi}, \tag{4}$$

where $p_{\boldsymbol{\varphi}}$ is the distribution of the imaging parameters $\boldsymbol{\varphi}$ and Φ is the set of all their possible values. We denote a noiseless projection as $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$. In our formulation, the projections in the real dataset are samples of a distribution p_{data}; hence, $p(\mathbf{y}|\mathbf{x}_{\text{true}}) = p_{\text{data}}(\mathbf{y})$ assuming that the forward

[1] We assume here conformational homogeneity of the underlying structure. The possible extension of CryoGAN to multiple conformations is discussed in *Discussion*.

model is correct.

We demonstrate in Theorem 1 in *Supplementary Note: Theoretical Guarantee of Recovery* that two 3D volumes have identical conditional distributions if and only if they are identical, up to rotation and reflection. Hence, Theorem 1 implies that, for the reconstruction $\mathbf{x}_{\text{rec}}$ to satisfy $\mathbf{x}_{\text{rec}} = \mathbf{x}_{\text{true}}$, it must also satisfy $p(\mathbf{y}|\mathbf{x}_{\text{rec}}) = p(\mathbf{y}|\mathbf{x}_{\text{true}})$. Thus, we can formulate the reconstruction task as the minimization problem

$$\mathbf{x}_{\text{rec}} = \arg\min_{\mathbf{x}} D\big(p(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\mathbf{x}_{\text{true}})\big), \tag{5}$$

where D is some distance between two distributions. In essence, (5) states that the appropriate reconstruction is the 3D density map whose projection distribution is the most similar to the real dataset in a distributional sense. For the sake of conciseness, we shall henceforth use the notation $p(\mathbf{y}|\mathbf{x}) = p_{\mathbf{x}}(\mathbf{y})$.

As distance in (5), we use the Wasserstein distance defined as

$$D(p_1, p_2) = \inf_{\gamma \in \Pi(p_1, p_2)} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma}[\|\mathbf{y}_1 - \mathbf{y}_2\|], \tag{6}$$

where $\Pi(p_1, p_2)$ is the set of all the joint distributions $\gamma(\mathbf{y}_1, \mathbf{y}_2)$ whose marginals are p_1 and p_2, respectively. Our choice is driven by works demonstrating that the Wasserstein distance is more amenable to minimization than other popular distances (*e.g.*, total-variation or Kullback-Leibler divergence) for this kind of application [30]. Using (6), the minimization problem (5) expands as

$$\mathbf{x}_{\text{rec}} = \operatorname*{argmin}_{\mathbf{x}} \inf_{\gamma \in \Pi(p_{\mathbf{x}}, p_{\text{data}})} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma}[\|\mathbf{y}_1 - \mathbf{y}_2\|]. \tag{7}$$

By using the formalism of [18, 30, 31], this minimization problem can also be stated in its dual form

$$\mathbf{x}_{\text{rec}} = \operatorname*{argmin}_{\mathbf{x}} \max_{f : \|f\|_L < 1} \Big(\mathbb{E}_{\mathbf{y} \sim p_{\text{data}}}[f(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}}[f(\mathbf{y})]\Big), \tag{8}$$

where $\|f\|_L$ denotes the Lipschitz constant of the function $f : \mathbb{R}^M \to \mathbb{R}$.

Connection with Wasserstein GANs

Equation (8) falls under the framework of the generative adversarial networks (GANs) [16] called Wasserstein GANs (WGANs) [30]. In the standard WGAN representation, the function f is parameterized by a neural network $\mathbf{D}_\phi$ with parameters ϕ that is called the discriminator. The task of this discriminator is to learn to differentiate between real samples (typically coming from an experimental dataset) and fake samples. The latter are produced by another neural network, called the generator, which aims at producing samples that are realistic enough to fool the discriminator.

This adversarial-learning scheme progressively drives the WGAN to capture the distribution of the experimental data.

In CryoGAN, we exploit this capability of adversarial schemes to learn the volume $\mathbf{x}$ whose simulated projections follow the captured real-data distribution. To do so, we rely on a cryo-EM physics simulator, whose role is to produce projections of a volume estimate $\mathbf{x}$ using (2). These simulated projections then follow a distribution $\mathbf{y} \sim p_{\mathbf{x}}$. Hence, (8) translates into

$$\mathbf{x}_{\text{rec}} = \underset{\mathbf{x}}{\text{argmin}} \max_{\mathbf{D}_\phi : \|\mathbf{D}_\phi\|_L \leq 1} \Big(\mathbb{E}_{\mathbf{y}\sim p_{\text{data}}}[\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\mathbf{y}\sim p_{\mathbf{x}}}[\mathbf{D}_\phi(\mathbf{y})] \Big). \tag{9}$$

As proposed in [32], the Lipschitz constraint $\|\mathbf{D}_\phi\|_L \leq 1$ can be enforced by penalizing the norm of the gradient of $\mathbf{D}_\phi$ with respect to its input. This gives the final formulation of our reconstruction problem as

$$\mathbf{x}_{\text{rec}} = \underset{\mathbf{x}}{\text{argmin}} \max_{\mathbf{D}_\phi} \Big(\mathbb{E}_{\mathbf{y}\sim p_{\text{data}}}[\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\mathbf{y}\sim p_{\mathbf{x}}}[\mathbf{D}_\phi(\mathbf{y})] - \lambda \cdot \mathbb{E}_{\mathbf{y}\sim p_{\text{int}}}[(\|\nabla_{\mathbf{y}} \mathbf{D}_\phi(\mathbf{y})\| - 1)^2] \Big). \tag{10}$$

Here, p_{int} denotes the uniform distribution along the straight line between points sampled from p_{data} and $p_{\mathbf{x}}$, while $\lambda \in \mathbb{R}_+$ is an appropriate penalty coefficient (see [32], Section 4).

The CryoGAN Algorithm

Equation (10) is a min-max optimization problem. By replacing the expected values with their empirical counterparts (sums) [32], we reformulate it as the minimization of

$$L_S(\mathbf{x}, \mathbf{D}_\phi) = \sum_{n\in S} \mathbf{D}_\phi(\mathbf{y}^n_{\text{data}}) - \sum_{n\in S} \mathbf{D}_\phi(\mathbf{y}^n_{\text{sim}}) - \lambda \sum_{n\in S} (\|\nabla_{\mathbf{y}} \mathbf{D}_\phi(\mathbf{y}^n_{\text{int}})\| - 1)^2), \tag{11}$$

where S consists of either the full dataset $S_{\text{full}} = \{1, \ldots, N_{\text{tot}}\}$ or a batch $B \subseteq S_{\text{full}}$; $\mathbf{y}^n_{\text{data}}$ is a projection sampled from the acquired experimental dataset; $\mathbf{y}^n_{\text{sim}} \sim p_{\mathbf{x}}$ is a projection of the current estimate $\mathbf{x}$ generated by the cryo-EM physics simulator; and $\mathbf{y}^n_{\text{int}} = \alpha_n \cdot \mathbf{y}^n_{\text{data}} + (1 - \alpha_n) \cdot \mathbf{y}_{\text{sim}}$, where α_n is sampled from a uniform distribution between 0 and 1.

In practice, we minimize (11) through stochastic gradient descents (SGD) using batches. We alternatively update the discriminator $\mathbf{D}_\phi$ (for n_{discr} iterations) using an Adam optimizer [33] with gradient

$$\nabla_\phi L_B(\mathbf{x}, \mathbf{D}_\phi) = \nabla_\phi \left(\sum_{n=1}^{N} \mathbf{D}_\phi(\mathbf{y}^n_{\text{batch}}) - \sum_{n=1}^{N} \mathbf{D}_\phi(\mathbf{y}^n_{\text{sim}}) - \lambda \sum_{n=1}^{N} (\|\nabla_{\mathbf{y}} \mathbf{D}_\phi(\mathbf{y}^n_{\text{int}})\| - 1)^2 \right), \tag{12}$$

and the volume $\mathbf{x}$ (for 1 iteration) using the batch gradient,

$$\nabla_{\mathbf{x}} L_B(\mathbf{x}, \mathbf{D}_\phi) = \nabla_{\mathbf{x}} \left(-\sum_{n=1}^{N} \mathbf{D}_\phi(\mathbf{y}_{\text{sim}}^n) \right). \tag{13}$$

The pseudocode and a schematic view of the CryoGAN algorithm are given in Algorithm 1 and Figure 1b, respectively. We provide further details of the CryoGAN physics simulator and discriminator network in the next two sections.

Algorithm 1 Pseudocode for CryoGAN

Parameters: number of training iterations, n_{train}; number of iterations of the discriminator per training iteration, n_{discr}; size of the batches used for SGD, N; penalty parameter, λ

```
for n_train do
    for n_discr do
        sample real projections: {y^1_batch, ..., y^N_batch} = {y^n_data}_{n∈B}
        sample projections simulated from current x: {y^1_sim, ..., y^N_sim} ~ p_x (see Algorithm 2)
        sample {α_1, ..., α_n} ~ U[0, 1]
        for all n ∈ {1, ..., N}, compute y^n_int = α_n · y^n_batch + (1 − α_n) · y^n_sim
        update the parameters φ of the discriminator D_φ using (12)
    sample {y^1_sim, ..., y_sim^N} ~ p_x
    update the volume x using (13)
```

The Cryo-EM Physics Simulator

The goal of the physics simulator is to sample $\mathbf{y}_{\text{sim}} \sim p_{\mathbf{x}}(\mathbf{y})$; this is done in three steps. First, we sample the imaging parameters $\boldsymbol{\varphi}$ from the distribution $p_{\boldsymbol{\varphi}}$: $\boldsymbol{\varphi} \sim p_{\boldsymbol{\varphi}}$. Second, we generate noiseless CTF-modulated and shifted projections from the current volume estimate with $\mathbf{H}_{\boldsymbol{\varphi}}(\mathbf{x})$. Third, we sample the noise model to simulate noisy projections $\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}(\mathbf{x}) + \mathbf{n}$, where $\mathbf{n} \sim p_{\mathbf{n}}$. We detail these steps in the following paragraphs, and a pseudocode of this cryo-EM physics simulator is given in Algorithm 2.

Algorithm 2 Pseudocode for Cryo-EM Physics Simulator

Parameters: current volume estimate, $\mathbf{x}$

```
sample the Euler angles θ = (θ_1, θ_2, θ_3) ~ p_θ.
sample the 2D shifts t = (t_1, t_2) ~ p_t.
sample the CTF parameters c = (d_1, d_2, α_ast) ~ p_c.
generate a synthetic noiseless projection based on (3), with y_noiseless = H_φ x.
sample the noise n ~ p_n. Add to the projection as y_sim = y_noiseless + n.
```

Recall that the set of imaging parameters is given by $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$. We first sample the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ from a distribution $p_{\boldsymbol{\theta}}$ decided *a priori* based on the acquired dataset. Similarly, the projection shifts $\mathbf{t} = (t_1, t_2)$ are sampled from the prior distribution $p_{\mathbf{t}}$. The CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$ are sampled from the prior distribution $p_{\mathbf{c}}$. In practice, we exploit

the fact that the CTF parameters can often be efficiently estimated for all micrographs. We then uniformly sample from the whole set of extracted CTF parameters.

We generate noiseless projections $\mathbf{y}_{\text{noiseless}}$ by applying $\mathbf{H}_{\boldsymbol{\varphi}}$ to the current volume estimate $\mathbf{x}$. The projection operator $\mathbf{P}_{\boldsymbol{\theta}}$ in (3) is implemented using the ASTRA toolbox [34].

The precise modeling of the noise is a particularly challenging feat in single-particle cryo-EM. To produce noise realizations that are as realistic as possible, we extract random background patches directly from the micrographs themselves, at locations where particles do not appear. For consistency, the noise patch added to a given noiseless projection is taken from the same micrograph that was used to estimate the CTF parameters previously applied to that specific projection. Additional details for this implementation are given in *Supplementary Note: Details of Experiments*.

The CryoGAN Discriminator Network

The role of the discriminator is to differentiate between projections from the experimental dataset and projections generated by the cryo-EM physics simulator. The gradients from the discriminator (see (13) in Algorithm 1) carry information on the difference between the real and simulated projections at a given run-time. Those gradients are used by the simulator to update itself, thus improving on the realism of the simulated projections.

The discriminator network takes an image as input and outputs a scalar value. Its architecture is illustrated in Figure 4. It is composed of 8 layers: 6 convolutional blocks followed by 2 fully connected (FC) layers. Each convolutional block is made up of a convolutional layer followed by a max-pooling and a leaky ReLU (with negative slope of 0.1). The number of channels in each convolutional layer is 96, 192, 384, 768, 1536, and 3072, respectively. The filters in these layers are of size 3, and the padding size is 1. The max-pooling layer uses a kernel of size 2 with a stride of 2. This leads to a downsampling by a factor of 2. The output of the final convolutional block is then reshaped, fed into the FC layer with 10 neurons, and finally processed by a leaky ReLU. The resulting activations are fed to the last FC layer to output a scalar.

Related Works

Related works fall into two main categories: current cryo-EM reconstruction methods and deep learning techniques that may apply to the cryo-EM pipeline; we now discuss each of these.

Cryo-EM Reconstruction

The main challenge in cryo-EM reconstruction is that every particle has an unknown pose in its micrograph—if the poses were known, maximum-likelihood (ML) or maximum *a posteriori* (MAP)

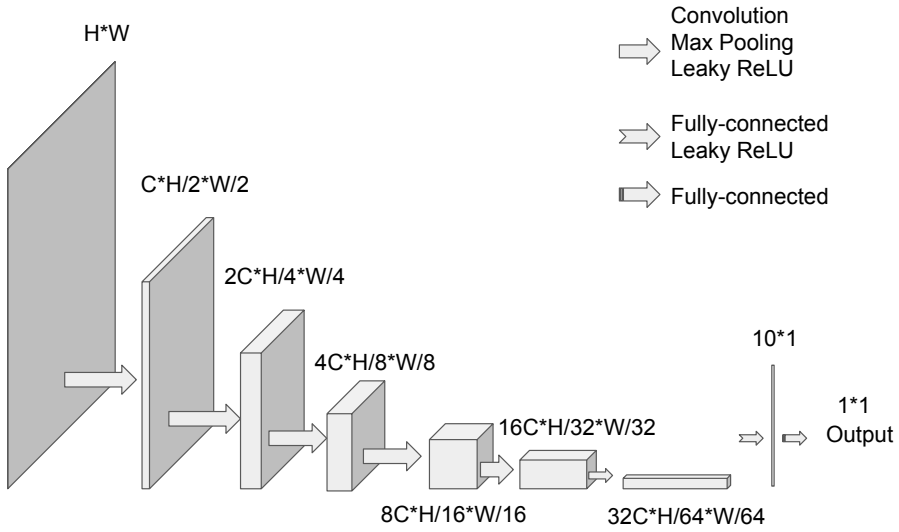


Figure 4: Architecture of the discriminator. The parameter for the channel size is $C = 96$ in every experiment. The input image with size $H \times W$ is successively processed and downsampled to output a scalar.

estimation of the volume could be performed by solving a standard linear inverse problem, where robustness would result from the large number of measurements which would counteract the low SNR of each measurement. One approach is to attempt to estimate the unknown poses iteratively. Pose estimation can be achieved with a variety of strategies, including the popular projection-matching approach [35, 36]. Whatever the method used, pose estimation is challenging because the SNR of individual projection images is extremely low. It also requires the estimation of additional parameters and the projection of the current reconstructed volume at a large number of poses and at every iteration of the reconstruction pipeline; ultimately, this is very computationally demanding.

Another approach is to formulate the reconstruction as a ML (or MAP) estimation problem in which the unknown poses are marginalized away [9, 11, 37]. This is attractive in that no extra parameter need to be estimated. The problem can then be solved using the expectation-maximization algorithm (e.g., [9,11]), where marginalization over poses during the so-called E-step is computationally expensive. Alternatively, the problem can be minimized using stochastic gradient descent (e.g., during the *ab initio* phase of [11]); here, the challenge is that the involved gradients require computations over all poses. For a more in-depth discussion, see [14, 15, 38]. For additional mathematical details on the relationship between likelihood-based methods and CryoGAN, see *Supplementary Note: CryoGAN vs. Likelihood-based Methods.*

Likelihood-free methods for cryo-EM reconstruction are relatively few. An early approach is [39], which proposes to reconstruct an *ab initio* structure such that the first few moments of the

distribution of its theoretical cryo-EM measurements match the ones of the particles. However, the method assumes that the poses of the particles have a uniform distribution. This moment-matching technique has been recently extended in [24] to reconstruct an *ab initio* structure in the case of nonuniform pose distributions.

By contrast, our CryoGAN framework proposes to match the distribution of the theoretical cryo-EM measurements to that of the real projections, by which we mean all the moments and not just the first few. Moreover, our method works for any pose distribution of the particles provided it is known beforehand. Alternatively, one could rely on a parametric model of the pose distribution and use the backpropagation mechanism of neural networks to learn its parameters during the CryoGAN run; this technique is explored in [24].

Deep Learning for Cryo-EM

Deep learning has already had a profound impact in a wide range of image-reconstruction applications [40–42]; however, their current utilization in cryo-EM is mostly restricted to preprocessing steps such as micrograph denoising [43] or particle picking [44–48]. A recent work uses neural networks to model continuous generative factors of structural heterogeneity [49]. However, the algorithm necessitates a pose-estimation procedure that relies on a merely conventional approach. Another recent work [25] uses a variational autoencoder trained using a discriminator-based objective to find a low-dimensional latent representation of the particles. These representations are then used to estimate the poses.

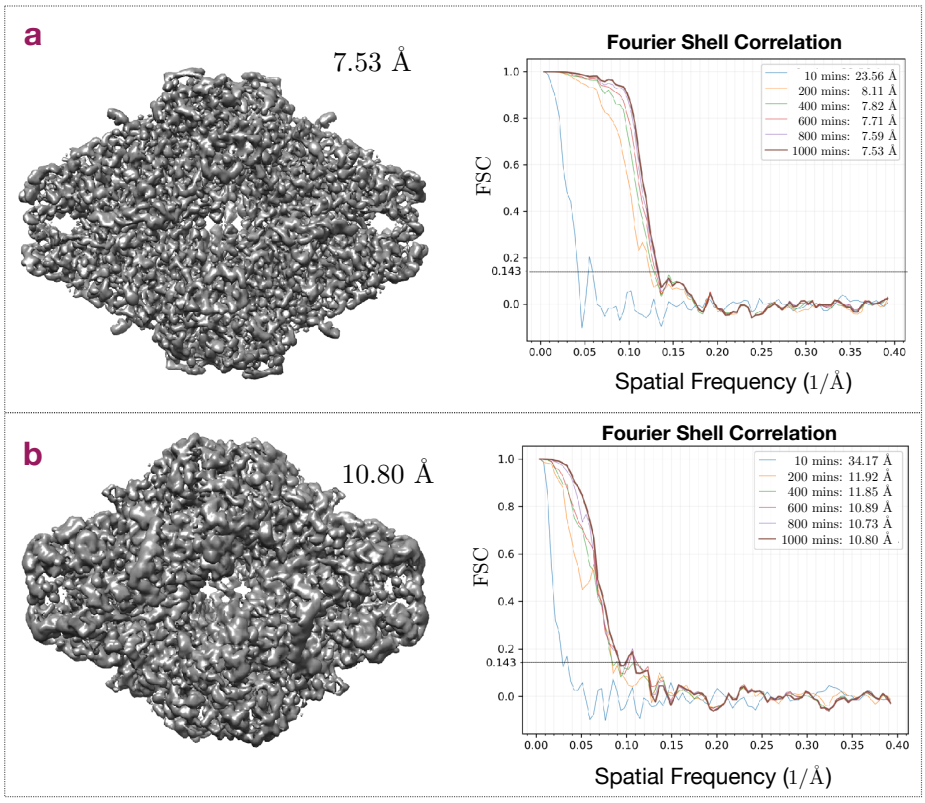
Deep learning is now extensively used to solve inverse problems in imaging [41, 50–52]. However, most methods are based on supervised learning and thus rely on training data. A GAN-based scheme that recovers the underlying distribution of the data from its noisy partial observations through a forward model was recently proposed in [27]. Finally, the reconstruction of a 3D structure (implicitly or explicitly) from its 2D viewpoints (and not projections) is an important problem in computer vision [53]. Many recent deep-learning algorithms have been used in this regard [54, 55]. While this problem is ostensibly similar to cryo-EM reconstruction, the measurement model for these problems is much less complicated than it is for cryo-EM and bares no relation to this modality.

SUPPLEMENTARY MATERIALS

Supplementary Discussion: Additional Results on Synthetic Data

The main synthetic experiment given in Figure 2 considers imaging conditions with a realistic noise level (-20 dB SNR). We performed additional experiments on synthetic β-galactosidase datasets to understand the effect of different imaging conditions on the quality of CryoGAN reconstruction. More precisely, we considered 1) the case with a lower noise level (-5.2 dB SNR) and 2) the case with a realistic noise level (-20 dB SNR) and translations (3% of the image size) in the projections. All the other conditions are identical to the main synthetic experiment. More details on these two additional experiments are provided in *Supplementary Note: Details of Experiments*.

The reconstructions obtained in the two cases reach a resolution of 7.53 Å and 10.8 Å, respectively, as shown in Supplementary Figure 1. As expected, the decrease of the noise level from -20 dB to -5.2 dB improves the reconstruction quality (from 8.64 Å to 7.53 Å). For the second case, the challenging presence of translations results in a slightly lower resolution. We believe that using a discriminator architecture that is invariant to shifts in the input image would improve the result in this case. Indeed, the discriminator would then be blind to such shifts in the real and simulated data, which would yield a reconstruction with a quality similar to the no-shift case.



Supplementary Figure 1: Additional CryoGAN reconstructions for synthetic datasets with different imaging conditions. **(a)** Reconstruction for a low noise case (-5.2 dB SNR). The corresponding evolution of the FSC curves with time is shown on the right. **(b)** Reconstruction for a realistic noise level (-20 dB SNR) and with translations (3% of the image size) in the data. The corresponding evolution of the FSC curves with time is shown on the right.

Supplementary Note: CryoGAN vs. Likelihood-based Methods

As discussed in [15], most reconstruction approaches in SPA currently rely a on maximum-likelihood (ML) formulation [37] written as

$$\mathbf{x}_{\text{rec}} = \arg\max_{\mathbf{x}} \sum_{n=1}^{N} \log p(\mathbf{y}^n_{\text{data}}|\mathbf{x}). \tag{1}$$

This formulation can be solved by using an expectation-maximization (ML-EM) algorithm or gradient descent. The former is preferred for iterative refinement [9], while a stochastic version of the latter is used for generating initial volume estimates in [11]. These techniques all involve the likelihood-estimation for each projection given the current volume estimate, which requires marginalizing the joint distribution $p(\mathbf{y}, \theta|\mathbf{x})$ over the space of poses. In ML-EM, this is side-stepped in the expectation step (E-step) by computing the conditional distribution on the poses for each projection given the volume estimate. This conditional distribution is then used to update the volume in the marginalization step (M-step). Hence, all ML techniques require, either implicitly or explicitly, computations over a large number of poses for each projection.

For another perspective, ML techniques can be viewed as distribution-matching approaches. Specifically, (1) minimizes an empirical estimate of the Kullback-Leibler (KL) divergence between the real distribution $p_{\text{data}}(\mathbf{y})$ and the simulated distribution $p(\mathbf{y}|\mathbf{x})$, such that

$$\begin{aligned}
\mathbf{x}_{\text{rec}} &= \arg\min_{\mathbf{x}} \mathrm{KL}(\mathrm{p}_{\text{data}}(\mathbf{y})||\mathrm{p}(\mathbf{y}|\mathbf{x})) && (2)\\
&= \arg\min_{\mathbf{x}} \mathbb{E}_{\mathbf{y}\sim p_{\text{data}}}\left[\log \frac{p_{\text{data}}(\mathbf{y})}{p(\mathbf{y}|\mathbf{x})}\right] && (3)\\
&= \arg\min_{\mathbf{x}} \mathbb{E}_{\mathbf{y}\sim p_{\text{data}}}\left[-\log p(\mathbf{y}|\mathbf{x})\right] && (4)\\
&\approx \arg\max_{\mathbf{x}} \sum_{n=1}^{N} \log p(\mathbf{y}^n_{\text{data}}|\mathbf{x}). && (5)
\end{aligned}$$

Hence, ML methods aim at finding the reconstruction whose simulated projection distribution matches that of the real data. In practice, this specific goal can be achieved by minimizing any suitable distance between these two distributions. By changing the distance, one can avoid the challenging likelihood computations that are inherent to the current ML methods, while preserving the theoretical guarantees that come with distribution matching (*e.g.*, Theorem 1).

This is precisely the philosophy behind CryoGAN, which relies on the Wasserstein distance to

formulate the distribution-matching task as

$$\mathbf{x}_{\text{rec}} = \underset{\mathbf{x}}{\text{argmin}} \max_{\mathbf{D}_\phi : \|\mathbf{D}_\phi\|_L \leq 1} \Big(\mathbb{E}_{\mathbf{y} \sim p_{\text{data}}}[\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}}[\mathbf{D}_\phi(\mathbf{y})] \Big). \tag{6}$$

In this formulation, the two distributions indirectly interact through a common function: the discriminator network D_ϕ. As a result, likelihood estimation is avoided; only a reliable sampler for each of the two distributions [56] is required. The samples from the real data distribution are readily available in the form of the acquired projection dataset, while the ones from the simulated distribution are generated by the cryo-EM physics simulator. Hence, CryoGAN is a likelihood-free technique with theoretical properties that are at least as good—if not better—than the ML ones. In fact, the Wasserstein distance is often easier to minimize than the KL divergence (*e.g.*, due to the smoothness of the former) [30].

Supplementary Note: Image Formation

For our forward model, we follow the developments in [28] (2.1-2.10), [57], and [23], which model the relationship between the 3D density map and a 2D measurement as linear.

The continuous-domain measurement can be given in the following form

$$y = H_{\boldsymbol{\varphi}}\{f\} + n \tag{7}$$

where $y : \mathbb{R}^2 \to \mathbb{R}$ is the intensity measured on the image plane, $n : \mathbb{R}^2 \to \mathbb{R}$ is the noise, $f : \mathbb{R}^3 \to \mathbb{R}$ is the density map we aim to recover, and $H_{\boldsymbol{\varphi}}$ is the measurement operator dependent on the imaging parameters $\boldsymbol{\varphi}$. This model can be further decomposed as

$$y = C_{\mathbf{c}} S_{\mathbf{t}} P_{\boldsymbol{\theta}}\{f\} + n \tag{8}$$

where $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \mathbf{t}, \mathbf{c})$. We now discuss the three involved operators in more details.

Projection Operator

The projection operation is given by [29]

$$P_{\boldsymbol{\theta}}\{f\}(x_1, x_2) = \int_{-\infty}^{\infty} \mathbf{R}_{\theta}\{f\}(x_1, x_2, x_3)\, \mathrm{d}x_3 \tag{9}$$

where $\mathbf{R}_{\boldsymbol{\theta}}$ is the rotation matrix associated with $\boldsymbol{\theta}$ and $R_{\boldsymbol{\theta}}\{f\}(\mathbf{x}) = f(R_{\boldsymbol{\theta}}^{-1}\mathbf{x})$.

Shift Operator

The projection measurements are picked from the micrographs and can thus be off-centered. This is modelled via the shift operator which, for any $y_s : \mathbb{R}^2 \to \mathbb{R}$, yields

$$S_{\mathbf{t}}\{y_s\}(x_1, x_2) = y(x_1 - t_1, x_2 - t_2) \tag{10}$$

where $\mathbf{t} = (t_1, t_2)$.

Convolution by CTF

The effect of the operator of $C_{\mathbf{c}}$ on any $y_c : \mathbb{R}^2 \to \mathbb{R}$ is given in Fourier domain as

$$\mathcal{F}\{C_{\mathbf{c}}\{y_c\}\}(\boldsymbol{\omega}) = \hat{C}_{\mathbf{c}}(\boldsymbol{\omega}).\hat{y}_c(\boldsymbol{\omega}) \tag{11}$$

where $\mathcal{F}$ is the Fourier transform and $\hat{y}_c = \mathcal{F}\{y_c\}$. Its Fourier transform $\hat{C}_{\mathbf{c}}$ (*i.e.*, the CTF) is given by

$$\hat{C}_{\mathbf{c}}(\boldsymbol{\omega}) = \hat{C}^p_{\mathbf{c}}(\boldsymbol{\omega})\hat{E}(\boldsymbol{\omega})\hat{A}(\boldsymbol{\omega}). \tag{12}$$

There, $\hat{C}^p_{\mathbf{c}} : \mathbb{R}^2 \to \mathbb{R}$ is the phase-contrast transfer function that takes the form

$$\hat{C}^p_{\mathbf{c}}(\boldsymbol{\omega}) = -\sqrt{1 - A^2}\sin(\gamma_{\mathbf{c}}(\boldsymbol{\omega})) - A^2\cos(\gamma_{\mathbf{c}}(\boldsymbol{\omega})), \tag{13}$$

with

$$\gamma_{\mathbf{c}}(\boldsymbol{\omega}) = \pi\lambda\left(d_{\mathbf{c}}(\alpha)\|\boldsymbol{\omega}\|^2 - \frac{1}{4}\lambda^3 c_{\mathrm{s}}\|\boldsymbol{\omega}\|^4\right), \tag{14}$$

where λ is the electron wavelength, c_s is third-order spherical aberration constant, α is the phase of the vector $\boldsymbol{\omega}$, and $d_{\mathbf{c}}(\alpha)$ is the defocus arising at the phase α. This defocus is given as

$$d_{\mathbf{c}}(\alpha) = d_1\cos^2(\alpha - \alpha_{\mathrm{ast}}) + d_2\sin^2(\alpha - \alpha_{\mathrm{ast}}), \tag{15}$$

where d_1 and d_2 are the horizontal and vertical defocus, respectively, and α_{ast} is the reference angle that defines the azimuthal direction of axial astigmatism. The objective aperture function $\hat{A} : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\hat{A}(\boldsymbol{\omega}) = \begin{cases} 1, & \|\boldsymbol{\omega}\| \leq \omega_{\mathrm{cutoff}} \\ 0, & \|\boldsymbol{\omega}\| > \omega_{\mathrm{cutoff}}, \end{cases} \tag{16}$$

where $\omega_{\mathrm{cutoff}} = \frac{2\pi d_{\mathrm{ap}}}{f_0\lambda}$ is the cutoff frequency, f_0 is the focal length of the objective lens, and d_{ap} corresponds to the diameter of the aperture. The spatial and chromatic envelope function $\hat{E} : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\hat{E}(\boldsymbol{\omega}) = \exp\left(-B(\|\boldsymbol{\omega}\|^2)\right), \tag{17}$$

where $B : \mathbb{R}^2 \to \mathbb{R}$ is a function influenced by chromatic aberration and spatial incoherence.

Discretization

The discretization of $H_{\boldsymbol{\varphi}}$ results in $\mathbf{H}_{\boldsymbol{\varphi}}$. This discretized measurement operator is itself decomposed of the discretized projection, shift, and convolution operation which are denoted by $\mathbf{P}_{\boldsymbol{\theta}}$, $\mathbf{S}_{\mathbf{t}}$, and $\mathbf{C}_{\mathbf{c}}$, respectively. The input to the operator $\mathbf{H}_{\boldsymbol{\varphi}}$ is a disretized version of the continous-domain 3D volume. This discretization of the 3D volume is done using a suitable basis function [58].

Supplementary Note: Theoretical Guarantee of Recovery

The CryoGAN paradigm is supported by Theorem 1, which states that perfect recovery is possible from continuous-domain measurements. In the continuous domain, we have $y = H_{\boldsymbol{\varphi}} f + n$ where $y : \mathbb{R}^2 \to \mathbb{R}$ is the 2D measurement obtained from the 3D volume f and n is the noise. Here $H_{\boldsymbol{\varphi}} = C_{\mathbf{c}} S_{\mathbf{t}} P_{\boldsymbol{\theta}}$ is the continuous-domain forward operator, $P_{\boldsymbol{\theta}}$ is the projection operator, $S_{\mathbf{t}}$ is the shift operator, and $C_{\mathbf{c}}$ is the operator for convolution with the CTF (see *Supplementary Note: Image Formation*).

Theorem 1. *Let $y = H_{\boldsymbol{\varphi}} f + n$ as given in (8) with $\boldsymbol{\varphi} = (\boldsymbol{\theta}, \mathbf{t}, \mathbf{c})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ are the projection angles, $\mathbf{t} = (t_1, t_2)$ are the shifts, and $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$ are the CTF parameters (defocus-major, defocus-minor, and angle of astigmatism, respectively), f is the continuous-domain 3D volume, and y, n are continuous-domain 2D images. Let $\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}$, $\mathbf{c} \sim p_{\mathbf{c}}$, $\mathbf{t} \sim p_{\mathbf{t}}$, and $\mathbb{P}_n$ be the probability measure associated with n. Moreover, assume that*

1. *the characteristic function $\hat{\mathbb{P}}_n$ of the noise probability measure $\mathbb{P}_n$ is non-zero everywhere in its domain and n is pointwise defined everywhere in $\mathbb{R}^2$;*
2. *the support of $p_{\mathbf{c}}$ is such that, for any $\mathbf{c}_1, \mathbf{c}_2 \in \text{Support}\{p_{\mathbf{c}}\}$ and $\mathbf{c}_1 \neq \mathbf{c}_2$, the fourier transform $\mathcal{F}\{C_{\mathbf{c}_1} + C_{\mathbf{c}_2}\}$ is non-zero everywhere;*
3. *the volume f is nonnegative everywhere and has a bounded support; and*
4. *the probability distributions $p_{\boldsymbol{\theta}}$, $p_{\mathbf{c}}$, and $p_{\mathbf{t}}$ are bounded.*

Then, it holds that

$$\mathbb{P}(\cdot|f_1) = \mathbb{P}(\cdot|f_2) \Leftrightarrow f_1 = G(f_2), \tag{18}$$

where G is some member of the set of rotation-reflection operations.

Before proving Theorem 1, we first comment on the assumptions. Assumption 1) is true for many noise distributions, including a white Gaussian noise filtered with any kernel of arbitrarily non-zero compact support. Assumption 2) is generally true as well. In fact, it is used to justify the application of Wiener filter to the clustered projections in classical cryo-EM reconstruction pipelines. Assumption 3) is true since the volume represents the density map, which is nonnegative. Also, the biological structures considered in cryo-EM have finite sizes.

Proof of Theorem 1. We denote the probability measure of $y_{\text{noiseless}} = H_{\boldsymbol{\varphi}} f$ with $\mathbb{P}_{\text{noiseless}}(\cdot|f)$. We shall prove the following in sequence:

1. $\mathbb{P}(\cdot|f_1) = \mathbb{P}(\cdot|f_2) \Leftrightarrow \mathbb{P}_{\text{noiseless}}(\cdot|f_1) = \mathbb{P}_{\text{noiseless}}(\cdot|f_2)$,

2. $\mathbb{P}_{\text{noiseless}}(\cdot|f_1) = \mathbb{P}_{\text{noiseless}}(\cdot|\mathbf{f}_2) \Leftrightarrow f_2 = G(f_1)$.

For the first part we progress by noting that $y = y_{\text{noiseless}} + n$. Recall that the characteristic function of the probability measure associated to the sum of two random variables is the product of their characteristic functions. Mathematically,

$$\hat{\mathbb{P}}(\cdot|f) = \hat{\mathbb{P}}_{\text{noiseless}}(\cdot|f)\,\hat{\mathbb{P}}_n. \tag{19}$$

By Assumption (1, Theorem 1), we can now write that

$$\hat{\mathbb{P}}_{\text{noiseless}}(\cdot|f) = \frac{\hat{\mathbb{P}}(\cdot|f)}{\hat{\mathbb{P}}_n}. \tag{20}$$

From (20), it is easy to see that $\mathbb{P}(\cdot|f_1) = \mathbb{P}(\cdot|f_2) \Leftrightarrow \mathbb{P}_{\text{noiseless}}(\cdot|f_1) = \mathbb{P}_{\text{noiseless}}(\cdot|f_2)$. This concludes the first part.

For the second part, we now invoke the result from Theorem 4. It states that if $f_2 \neq G(f_1)$ for any G in the set of rotation and reflection operation, then the corresponding $\mathbb{P}_{\text{noiseless}}(\cdot|f_1)$ and $\mathbb{P}_{\text{noiseless}}(\cdot|f_2)$ are mutually singular. This means that the support of their intersection has zero measure. Since, we have $\mathbb{P}_{\text{noiseless}}(\cdot|f_1) = \mathbb{P}_{\text{noiseless}}(\cdot|f_2)$ this means they are not mutually singular. This implies that $f_2 = G(f_1)$ for some G in the set of rotation and reflection operation. This concludes the proof.

Discrete-domain Extension. In practice, the cryo-EM measurements are acquired on a detector grid and are therefore discrete. CryoGAN reconstructs a voxel-domain $\mathbf{x}_{\text{rec}}$ such that $p(\mathbf{y}|\mathbf{x}_{\text{rec}}) = p(\mathbf{y}|\mathbf{x}_{\text{GT}})$, meaning that the corresponding probability measures are equal as well. Theorem 1 then holds approximately, up to some error that results from discretization of the measurements, forward model, and 3D density map. We leave a more thorough analysis of this error for future work.

In the following two sections, we build up to Theorem 4. First by analyzing the problem in the absence of the CTF and noise, and then by adding the CTF.

Recovery in the Absence of Noise and CTF

In the absence of CTF and shifts the recoverability of $f : \mathbb{R}^3 \to \mathbb{R}$ from its 2D projections obtained at unknown random poses is guaranteed by [59, Theorem 3.1]. We first go through the notations described in [59] before we state the required foundational result. We then extend [59, Theorem 3.1] to the case when the CTF and shifts are present.

Notations and Preliminaries

Let $SO(3)$ be the space of the special orthonormal matrices and $\mathcal{D}$ be the Borel $\sigma-$algebra induced using the standard Riemannian metric on $SO(3)$. Then, $(SO(3), \mathcal{D})$ describes the measurable space of orthonormal matrices. Let $\Delta_N^W = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_2 \leq W\}$ for some $W \in \mathbb{R}^+$. By $(\mathcal{L}_2, \mathcal{B})$, we denote the measurable space of all the square-integrable functions supported in Δ_2^W with Borel $\sigma-$algebra $\mathcal{B}$ induced by the L_2-norm. We denote by $\mathbb{F}$ the set of all the functions supported in Δ_3^W, which are nonnegative and essentially bounded.

For any $f \in \mathbb{F}$ and $\mathbf{A} \in SO(3)$, we denote $y = \mathcal{P}_{\mathbf{A}}\{f\} = \int_{-\infty}^{\infty} \mathbf{A}f(x_1, x_2, x_3)\, \mathrm{d}x_3$ where $\mathbf{A}f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{x})$. Let $p_{\mathbf{A}}$ be a probability density on the space $(SO(3), \mathcal{D})$. Note that there is a bijective mapping from $\boldsymbol{\theta}$ in Theorem 1 and $\mathbf{A}$. In fact, $\mathbf{A}$ represents the rotation matrix associated with the projection angle $\boldsymbol{\theta}$.

We denote by Ψ the normalized Haar measure on $(SO(3), \mathcal{D})$ and by $\Psi_{\mathbf{A}}$ the measure associated with $p_{\mathbf{A}}$ such that $\Psi_{\mathbf{A}}[\cdot] = \int_{(\mathbf{a} \in \cdot)} p_{\mathbf{A}}(\mathbf{a}) \Psi[\,\mathrm{d}\mathbf{a}]$.

For a given $f \in \mathbb{F}$, the density $p_{\mathbf{A}}$ induces a probability measure $\mathbb{P}_{\text{proj}}(\cdot|f)$ on the space $(\mathcal{L}_2, \mathcal{B})$ through the mapping $\mathcal{P}_{\mathbf{A}}\{f\}$ such that

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \Psi_{\mathbf{A}}[\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{A}}\{f\} \in \cdot\}]. \tag{21}$$

When $p_{\mathbf{A}}$ is uniform on $SO(3)$, one has that

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \mathbb{P}_{\text{proj}}(\cdot|\mathbf{R}f), \ \forall f \in \mathbb{F} \text{ and } \mathbf{R} \in O(3), \tag{22}$$

where O_3 is the space of all orthogonal matrices such that $\det \mathbf{A} \in \{-1, 1\}$. The invariance in (22) is true since

$$\begin{aligned}
\mathbb{P}_{\text{proj}}(\cdot|f) &= \Psi[\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{A}}\{f\} = \cdot\}] \\
&= \Psi[\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{R}^{-1}\mathbf{A}}\{\mathbf{R}f\} = \cdot\}] \\
&= \Psi[\{\mathbf{R}\mathbf{A}' \in SO(3) : \mathcal{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\}] \\
&= \Psi[\{\mathbf{A}' \in SO(3) : \mathcal{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\}]
\end{aligned} \tag{23}$$

where $\mathbf{A}' = \mathbf{R}^{-1}\mathbf{A}$ and the last equality follows from the right invariance of Haar measure. We define $G\{\mathbb{F}\} = \{\gamma_A : A \in O_3\}$ such that

$$(\gamma_A f)(\cdot) = f(\mathbf{A}^{-1}\cdot), \forall \mathbf{A} \in O(3), f \in \mathbb{F}. \tag{24}$$

We define the shape $[f]$ as an orbit of f under the influence of G such that $[f] = \{\gamma_{\mathbf{A}} f : \gamma_A \in G\}$. When $p_{\mathbf{A}}$ is uniform, the shape $[f]$ is composed of all the rotations and reflections of f.

We can now restate [59, Theorem 3.1]. We discuss here the sketch of the proof given in [59].

Theorem 2 ([59, Theorem 3.1])**.** *Let $p_{\mathbf{A}}$ be any bounded distribution on $SO(3)$ and let the assumptions of Theorem 1 be true; then, $\forall f, g \in \mathbb{F}$,*

$$[f] \neq [g] \Rightarrow \mathbb{P}_{\text{proj}}(\cdot|f) \perp \mathbb{P}_{\text{proj}}(\cdot|g). \tag{25}$$

Sketch of the Proof. Without loss of generality, we provide the sketch of the proof for the case when $p_{\mathbf{A}}$ is uniform. For the case when $p_{\mathbf{A}}$ is nonuniform the argument remains the same provided that $\Psi_{\mathbf{A}}$ associated with the non-uniform distribution $p_{\mathbf{A}}$ is absolutely continuous with respect to Ψ ($\Psi_{\mathbf{A}} \ll \Psi$). This has been stated in [59]. Since we assume $p_{\mathbf{A}}$ to be bounded, this condition is satisfied. The only difference here with respect to the uniform distribution is that the orbit of f and g are more restricted than O(3).

The proof first uses in [60, Proposition 7.8] which we restate here as Proposition 3.

Proposition 3 ([60, Proposition 7.8])**.** *Let $f \in \mathbb{F}$ and let $S_{\mathbf{A}}$ be an uncountably infinite subset of $SO(3)$, then f is determined by the collection $\{\mathcal{P}_A\{f\}\}_{\mathbf{A} \in S_{\mathbf{A}}}$ ordered with respect to $\mathbf{A} \in S_{\mathbf{A}}$.*

Note that this proposition assumes that the angle of the projections are known. Although in our case the angles are unknown, we shall see that this proposition will be useful.

We now want to determine how different $\mathbb{P}_{\text{proj}}(.|f)$ and $\mathbb{P}_{\text{proj}}(.|g)$ are for any given f and g. For this, we use the equality

$$\text{TV}(\mathbb{P}_1, \mathbb{P}_2) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{(y_1, y_2) \sim \gamma}[\mathbb{1}_{y_1 \neq y_2}], \tag{26}$$

where TV is the total variation distance and $\Pi(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all the joint distributions $\gamma(y_1, y_2)$ whose marginals are $\mathbb{P}_1$ and $\mathbb{P}_2$ [18]. In fact, $\mathbb{E}[\mathbb{1}_{y_1 \neq y_2}]$ is equal to the probability of the event $y_1 \neq y_2$. In our context, this translates into

$$\text{TV}(\mathbb{P}_{\text{proj}}(.|f), \mathbb{P}_{\text{proj}}(.|g)) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_{\text{proj}}(.|f), \mathbb{P}_{\text{proj}}(.|g))} \text{Prob}(y_1 \neq y_2), \text{ where } (y_1, y_2) \sim \gamma. \tag{27}$$

The optimum is achieved at the extremas which are sparse joint distributions and are such that the variable y_2 is a function of y_1. For any arbitrary joint distribution (or coupling) of this form, the proof then assigns a measurable function $h : SO(3) \to SO(3)$ such that $(y_1, y_2) = (\mathcal{P}_{\mathbf{A}}\{f\}, \mathcal{P}_{h(\mathbf{A})}\{g\})$ for $\mathbf{A} \sim p_{\mathbf{A}}$.

We can then write that

$$\Psi[\{\mathbf{A} \in SO(3) : \mathcal{P}_{h(\mathbf{A})}\{g\} \in \cdot\}] = \mathbb{P}_{\text{proj}}(\cdot|g). \tag{28}$$

The task now is to estimate $\text{Prob}(y_1 \neq y_2)$, where $(y_1, y_2) = (\mathcal{P}_{\mathbf{A}}\{f\}, \mathcal{P}_{h(\mathbf{A})}\{g\})$ for $\mathbf{A} \sim p_{\mathbf{A}}$.

(Continuous h). When h is continuous, Proposition 3 implies that, if $[f] \neq [g]$, then

$$\Psi[\{\mathbf{A} \in SO(3) : \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{h(\mathbf{A})}\{g\}\|_2 > 0\}] = 1. \tag{29}$$

(General h). When the function h is discontinuous, the proof uses Lusin's theorem to approximate h by a continuous function. Lusin's theorem states that, for any $\delta > 0$, there exists an h_δ such that $h(\mathbf{A}) = h_\delta(\mathbf{A}), \forall \mathbf{A} \in \mathcal{H}_\delta$ and $\Psi[SO(3)|\mathcal{H}_\delta] < \delta$. This then leads to

$$\begin{aligned}\Psi[\{\mathbf{A} \in SO(3) : \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{h(\mathbf{A})}\{g\}\|_2 > 0\}] &\geq \Psi(H_\delta) \\ &\geq 1 - \delta. \end{aligned} \tag{30}$$

Since δ is arbitrarily small, the event $\{\mathcal{P}_{\mathbf{A}}\{f\} \neq \mathcal{P}_{h(\mathbf{A})}\{g\}\}$ has probability 1.

In conclusion, for any arbitrary coupling, the event $\{\mathcal{P}_{\mathbf{A}}\{f\} \neq \mathcal{P}_{h(\mathbf{A})}\{g\}\}$ has probability 1 if $[f] \neq [g]$. This implies that, when $[f]$ and $[g]$ are not the same, the total-variation distance between $\mathbb{P}_{\text{proj}}(\cdot|f)$ and $\mathbb{P}_{\text{proj}}(\cdot|g)$ is 2. This ensures that the two probability measures are mutually singular meaning that the intersection of their support has zero measure. This concludes the proof.

Recovery in the absence of Noise

We now extend the previous result to the case when the CTF is present. For the sake of simplicity we do not take into account shifts in the forward model. However, it is trivial to generalize the results to them since, shifts don't change the information content in the projections but only their location.

We assume that $\mathbf{c} \sim p_{\mathbf{c}}$ such that the support of $p_{\mathbf{c}}$ is in some bounded region $\mathcal{C} \subset \mathbb{R}^3$. We denote $\Psi_{\mathbf{c}}[\cdot]$ as the measure associated with $p_{\mathbf{c}}$ on the space $\mathcal{C}$.

We denote by $(SO(3) \times \mathcal{C})$ the product space of $SO(3)$ and $\mathcal{C}$, while we denote by $\Psi_{\mathbf{A},\mathbf{c}}$ the measure on this product space. We then define

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|f) = \Psi_{\mathbf{A},\mathbf{c}}[\{(\mathbf{A}, \mathbf{c}) \in (SO(3) \times \mathcal{C}) : C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} \in \cdot\}], \tag{31}$$

where $C_{\mathbf{c}}$ is the space-domain CTF given in (8).

Theorem 4. *Let $p_{\mathbf{A}}$ be a bounded probability distribution on $SO(3)$, $p_{\mathbf{c}}$ be a distribution of the CTF with parameters $\mathbf{c} \in \mathcal{C}$, and let the assumptions of Theorem 1 be true; then, $\forall f, g \in \mathbb{F}$,*

$$[f] \neq [g] \Rightarrow \mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g). \tag{32}$$

Proof. Similarly to the previous proof, we show that the TV distance between $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$ is 2 when $[f]$ and $[g]$ are distinct. For simplification, we assume that $p_{\mathbf{A}}$ is uniform. (When this is not the case the proof essentially remains the same.) We need to show that $\text{Prob}(y_1 \neq y_2) = 1$, where $(y_1, y_2) \sim \gamma$ for any arbitrary coupling γ of $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$. For an arbitrary coupling γ such that $\text{Prob}(y_1 \neq y_2)$ is minimum, we again assign $h : (SO(3) \times \mathcal{C}) \to (SO(3) \times \mathcal{C})$ such that

$$(y_1, y_2) = (C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\}, C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\}), \tag{33}$$

where $\mathbf{A} \sim p_{\mathbf{A}}, \mathbf{c} \sim p_{\mathbf{c}}$ and where $h_0 : (SO(3) \times \mathcal{C}) \to SO(3)$ and $h_1 : (SO(3) \times \mathcal{C}) \to \mathcal{C}$ are such that $h(\mathbf{A}, \mathbf{c}) = (h_0(\mathbf{A}, \mathbf{c}), h_1(\mathbf{A}, \mathbf{c}))$. This implies that

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|g) = \Psi_{\mathbf{A},\mathbf{c}}[\{(\mathbf{A}, \mathbf{c}) \in (SO(3) \times \mathcal{C}) : C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\} \in \cdot\}]. \tag{34}$$

We now show that, for any h, the event $\{y_1 \neq y_2\}$ has probability 1.

(Continuous h). We first assume that h is continuous and use the same kind of technique as in the proof of [59, Theorem 3.1].

Since $SO(3)$ is transitive, we can write that

$$h(\mathbf{A}, \mathbf{c}) = (\mathbf{A}\Gamma_{\mathbf{A},\mathbf{c}}, h_1(\mathbf{A}, \mathbf{c})). \tag{35}$$

As h is continuous, so is $\Gamma_{\mathbf{A},\mathbf{c}}$. Let $\{\mathcal{A}_n^m \times \mathcal{C}_n^m\}_{m=1}^n$ be a collection of n disjoint sets which creates the partition of $(SO(3) \times \mathcal{C})$. These partitions are such that for any m, there exists a k_m such that $\{\mathcal{A}_{n+1}^m \times \mathcal{C}_{n+1}^m\} \subset \{\mathcal{A}_n^{k_m} \times \mathcal{C}_n^{k_m}\}$. This means that, as n increases, the partitions become finer. We now define

$$h_n(\mathbf{A}, \mathbf{c}) = (\mathbf{A}\Gamma_n^m, h_{n,1}^m(\mathbf{A}, \mathbf{c})) \ \forall \ (\mathbf{A}, \mathbf{c}) \in \{\mathcal{A}_n^m \times \mathcal{C}_n^m\}, \tag{36}$$

such that

$$\Gamma_n^m = \underset{\Gamma \in \{\Gamma_{\mathbf{A},\mathbf{c}} : (\mathbf{A},\mathbf{c}) \in \{\bar{\mathcal{A}}_n^m \times \bar{\mathcal{C}}_n^m\}\}}{\arg\min} \ \min_{(\mathbf{A},\mathbf{c}) \in \{\bar{\mathcal{A}}_n^m \times \bar{\mathcal{C}}_n^m\}} \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{\mathbf{A}\Gamma}\{g\}\|, \tag{37}$$

where $\bar{\mathcal{A}}_n^m$ and $\bar{\mathcal{C}}_n^m$ are the closures of $\mathcal{A}_n^m$ and $\mathcal{C}_n^m$, respectively. The sequence h_n converge to h as $n \to \infty$. We denote

$$\mathcal{K} = \{(\mathbf{A}, \mathbf{c}) \in (SO(3) \times \mathcal{C}) : \|C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}\Gamma_{\mathbf{A}}}\{g\}\| > 0\}, \tag{38}$$

$$\mathcal{K}_n = \{(\mathbf{A}, \mathbf{c}) \in (\mathcal{A}_n^m \times \mathcal{C}_n^m) : \|C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma_n^m g\}\| > 0\}. \tag{39}$$

Similarly to [59, Theorem 3.1], we can then show that

$$\Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}] = \lim_{n\to\infty} \sum_{m=1}^{m=n} \Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}_m]. \tag{40}$$

We invoke Proposition 5, which gives that $\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}_n] = \Psi_{\mathbf{A},\mathbf{c}}[(\mathcal{A}_n^m \times \mathcal{C}_n^m)]$. Therefore, $\Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}] = \Psi_{\mathbf{A},\mathbf{c}}[(SO(3) \times \mathcal{C})] = 1$. This means that, when h is continuous, the event $\{y_1 \neq y_2\}$ has probability 1 if $[f] \neq [g]$.

(General h). When h is discontinuous, we can invoke Lusin's theorem to claim the same, similarly to Theorem 2. This means that, for any h, if $[f] \neq [g]$, then the probability of the event $\{y_1 \neq y_2\}$ is 1. Therefore, the TV distance between $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$ is 2, yielding that $\mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g)$. This concludes the proof.

Proposition 5. *Let $f, g \in \mathbb{F}$, $\mathcal{A}' \subseteq SO(3)$, $\mathcal{C}' \subseteq \mathcal{C}$, $\Gamma \in SO(3)$, and*

$$\mathcal{K}' = \{(\mathbf{A}, \mathbf{c}) \in (\mathcal{A}' \times \mathcal{C}') : \|C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma g\}\| > 0\}. \tag{41}$$

Let the assumptions from Theorem 1 be true. Then, if $[f] \neq [g]$, it holds that

$$\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}'] = \Psi_{\mathbf{A},\mathbf{c}}[(\mathcal{A}' \times \mathcal{C}')]. \tag{42}$$

Proof. We show that $\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}'^c] = 0$, where $(\mathcal{K}'^c \cup \mathcal{K}') = (\mathcal{A}' \times \mathcal{C}')$. We define the set $S_{\mathbf{A}} = \{\mathbf{c} \in \mathcal{C}' : \|C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma g\}\| = 0\}$. We define $S_{\mathcal{A}''} = \cup_{\mathbf{A}\in\mathcal{A}''} S_{\mathbf{A}}$ for any $\mathcal{A}'' \subseteq \mathcal{A}'$. We define

$$\mathcal{A}_1' = \{\mathbf{A} \in \mathcal{A}' : S_{\mathbf{A}} \text{ is an uncountable set}\}, \tag{43}$$

$$\mathcal{A}_2' = \{\mathbf{A} \in \mathcal{A}' : S_{\mathbf{A}} \text{ is a countable non-empty set}\}. \tag{44}$$

Note that $\mathcal{K}'^c = \cup_{k=1}^2 \cup_{\mathbf{A}\in\mathcal{A}_k'} (\mathbf{A} \times S_{\mathbf{A}})$. Then,

$$\Psi_{\mathbf{A},\mathbf{c}}[\{\mathcal{K}'^c] = \sum_{k=1}^{2} \Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}_k'} (\mathbf{A} \times S_{\mathbf{A}})\}] \tag{45}$$

We now look at the two cases.

- *(When $S_{\mathbf{A}}$ is uncountable).* For this case, we show that $\Psi[\mathcal{A}_1'] = 0$. The main argument is that if this is not true, then it contradicts $[f] \neq [g]$.

  For the sake of conciseness, we denote $\mathcal{P}_{\mathbf{A}}\{f\}$ by I_f and $\mathcal{P}_{\mathbf{A}}\{\Gamma g\}$ by I_g. For any $\mathbf{A} \in \mathcal{A}_1'$, it holds that

$$C_{\mathbf{c}} * I_f = C_{h_1(\mathbf{A},\mathbf{c})} * I_g, \ \forall \mathbf{c} \in S_{\mathbf{A}}, \tag{46}$$

$$\widehat{C}_{\mathbf{c}} \cdot \hat{I}_f = \widehat{C}_{h_1(\mathbf{A},\mathbf{c})} \cdot \hat{I}_g, \ \forall \mathbf{c} \in S_{\mathbf{A}}, \tag{47}$$

  where $\widehat{C}, \hat{I}_f$, and $\hat{I}_g$ are the Fourier transforms of $\mathbf{C}, I_f$, and I_g, respectively.

  We define $ze(\hat{I}) = \{\omega \in \mathbb{R}^2 : \widehat{I}(\omega) = 0\}$, $\omega_\alpha = \{[(r\cos\alpha, r\sin\alpha)] : r > 0\}$, and $ze_\alpha(\hat{I}) = ze(\hat{I}) \cap \omega_\alpha$. From (47), we can write that

$$ze(\widehat{C}_{\mathbf{c}}) \cup ze(\hat{I}_f) = ze(\widehat{C}_{h_1(\mathbf{A},\mathbf{c})}) \cup ze(\hat{I}_g), \ \forall \mathbf{c} \in S_{\mathbf{A}}. \tag{48}$$

  Two remarks are in order. Firstly, by assumption 2 of Theorem 1, $ze(\widehat{C}_{\mathbf{c}_1}) \cap ze(\widehat{C}_{\mathbf{c}_2}) = \emptyset$ for $\mathbf{c}_1 \neq \mathbf{c}_2$. (Remember that $ze_\alpha(\widehat{C}_{\mathbf{c}})$ for any $\alpha \in [0, \pi]$ is nonempty (see "Image Formation Theory).) Secondly, by assumption 3 of Theorem 1, the supports of f and g are compact and nontrivial, so are the supports of I_f and I_g. This means that their Fourier transforms $\hat{I}_f$ and $\hat{I}_g$ are analytic functions, which implies that there are infinitely many α such that the cardinality of the sets $ze_\alpha(\hat{I}_f)$ and $ze_\alpha(\hat{I}_f)$ is countable. We call the set of such α as $\mathcal{S}_\alpha$. Now, we have that

$$\begin{aligned}
&ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap (ze_\alpha(\widehat{C}_{\mathbf{c}}) \cup ze_\alpha(\hat{I}_f)) = ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap (ze_\alpha(\widehat{C}_{h_1(\mathbf{A},\mathbf{c})}) \cup ze_\alpha(\hat{I}_g)),\\
&ze_\alpha(\widehat{C}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = (ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\widehat{C}_{h_1(\mathbf{A},\mathbf{c})})) \cup (ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g)),\\
&ze_\alpha(\widehat{C}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g)
\end{aligned} \tag{49}$$

  for all $\mathbf{c} \in S_{\mathbf{A}}$ and $\alpha \in [0, \pi]$.

  We can now write that

$$\cup_{\mathbf{c} \in S_{\mathbf{A}}} ze_\alpha(\widehat{C}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = \cup_{\mathbf{c} \in S_{\mathbf{A}}} ze_\alpha(\widehat{C}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g). \tag{50}$$

  for any $\alpha \in S_\alpha$. The set on the left hand side of (50) has an uncountably infinite cardinality since there are uncountably many $\mathbf{c} \in \mathcal{S}_{\mathbf{A}}$ and for each $\mathbf{c}$ there are distinct $ze_\alpha(\widehat{C}_{\mathbf{c}})$. In return, the set in the right hand side of (50) is countable for a given $\alpha \in S_\alpha$. Therefore, for any $\alpha \in S_\alpha$, the two sets have different cardinality, which raises a contradiction. The only possible scenario

in which (48) is true is when $h_1(\mathbf{A}, \mathbf{c}) = \mathbf{c}$. Using (47), we infer that $\mathcal{P}_{\mathbf{A}}\{f\} = \mathcal{P}_{\mathbf{A}}\{\Gamma g\}$. Therefore, for any $\mathbf{A} \in \mathcal{A}_1'$, $\mathcal{P}_{\mathbf{A}}\{f\} = \mathcal{P}_{\mathbf{A}}\{\Gamma g\}$. However, $\Psi[\mathcal{A}_1'] = 0$ since, if this is not true, then $[f] = [g]$ by Proposition 3.

Now note that

$$\begin{aligned}\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}_1'}(\mathbf{A} \times S_{\mathbf{A}})] &\leq \underbrace{\Psi[\mathcal{A}_1']}_{0}\underbrace{\Psi_{\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}_1'} S_{\mathbf{A}}]}_{\text{finite}} \\ &= 0. \end{aligned} \tag{51}$$

- *(When $S_{\mathbf{A}}$ is countable and nonempty).* Since $S_{\mathbf{A}}$ is a countable set in this case, its elements have a bijection with natural numbers. We denote this bijection by $b : \mathbb{Z} \times \mathcal{A}_2' \to S_{\mathbf{A}}$. We denote by $q(z) = \cup_{\mathbf{A}\in\mathcal{A}_2'}(\mathbf{A}, b_{\mathbf{A}}(z))$, $\forall z \in \mathbb{Z}$. Note that $q(z)$ is a graph of the function $b(z, \cdot)$. Since it is a graph, $\Psi_{\mathbf{A},\mathbf{c}}[q(z)] = 0$.

  We also have that $\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}_2'}(\mathbf{A} \times S_{\mathbf{A}})] = \Psi_{\mathbf{A},\mathbf{c}}[\sum_{z\in\mathbb{Z}} q(z)]$. The latter vanishes since it is the measure of a countable addition of sets of measure zero. Hence, $\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}_2'}(\mathbf{A} \times S_{\mathbf{A}})] = 0$.

This gives that $\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}'^c] = \sum_{k=1}^{2} \Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}_k'}(\mathbf{A} \times S_{\mathbf{A}})] = 0$, which concludes the proof.

Supplementary Note: Details of Experiments

Synthetic Data (Figure 2)

Dataset. We construct a synthetic cryo-EM dataset that mimics the experimental β-galactosidase dataset (EMPIAR-10061) from [17]. We generate 41,000 synthetic β-galactosidase projections of size 192×192 using our cryo-EM image-formation model (see *Online Methods*). For the ground-truth volume, we generate a 2.5 Å density map from PDB-*5a1a* atomic model using Chimera [61]. This gives a volume of size $(302 \times 233 \times 163)$ with voxel size 0.637 Å. The volume is then padded, averaged, and downsampled to a size $(180 \times 180 \times 180)$ with voxel size 1.274 Å. This corresponds to a Nyquist resolution of 2.548 Å for the reconstructed volume.

The projection poses are sampled from a uniform distribution over $SO(3)$, where $SO(3)$ is the group of 3D rotations around the origin of $\mathbb{R}^3$.

In order to apply random CTFs and noise, we randomly pick a micrograph in the EMPIAR-10061 dataset. We extract its CTF paramters using CTFFIND4 [62] and apply them to a clean projection. The parameter B of the envelope function of the CTF (see (17)) is chosen such that it decays to 0.2 at the Nyquist frequency. We then randomly select a background patch from the same micrograph to simulate noise. The noise is downsampled to size 192×192, and normalized to zero-mean, scaled and added to the projection. The scaling is such that the ratio of the energy of the signal to the energy of the noise (SNR) is kept at 0.01, which is equivalent to -20 dB. The dataset is then randomly divided into two halves. The algorithm is applied separately on both halves to generate the half-maps. The FSC between the two half maps is then reported using FOCUS [63,64].

Generator Settings. We reconstruct a volume of size $(184 \times 184 \times 184)$ voxels for each half dataset. The pixel size is 1.274 Å. The volumes are initialized with zeros. The D2 symmetry of β-galactosidase is enforced during reconstruction.

We use our image-formation model to simulate realistic projections from the current volume estimate at every CryoGAN iteration. The distribution of the imaging parameters is identical to the one used to generate the dataset. To add the noise on the CTF-modulated projections, we also keep the same approach as the one used to generate the dataset. However, we assume that the final SNR of each projection is unknown, leading us to learn the scaling parameter that controls the ratio between the projections and the noise patches.

We apply a binary spherical mask of size $(184 \times 184 \times 184)$ on the learned volume. To handle the sharp transitions at the mask borders, we restrict the voxel values to lie above a certain value. This value changes as a function of position and iteration number: it increases linearly with the

distance from the center of the volume to the border of the mask, from $V_{\min}$ to 0, and the value $V_{\min}$ changes as a function of iteration number, starting at 0 and decreasing to -2% of the current maximum voxel value. This promotes nonnegativity during the initial phases of reconstruction and increases the stability of the algorithm.

Discriminator Architecture. The architecture of the discriminator network is detailed in *Online Methods*. The discriminator is initialized identically for the two half datasets. All projections (*i.e.*, the real projections and the ones generated by the simulator) are normalized to zero-mean and a unit standard-deviation before being fed to the discriminator.

General Settings. The adversarial learning scheme is implemented in Pytorch [19]. For the optimization, we use [33] ($\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$) with a learning rate of 10^{-3} and a batch size of 8. The algorithm is run for 40 epochs and the learning rate decreases by 1% at every epoch. The parameter for the gradient-penalty term (see (10)) is kept to $\lambda = 0.001$. The discriminator is trained 4 times for every training of the generator ($n_{\text{discr}} = 4$ in Algorithm 1).

For the back-propagations, the norm of the gradients for the discriminator are clipped to a maximal value of 10^8. For the generator, the gradients for each pixel are clipped to a maximal value of 10^4. The clipping values increase linearly from zero to those maxima in the first two epochs. Doing so improves on the stability of the adversarial learning scheme in the start, in particular, on that of the discriminator. All parameters are tuned for a fixed value range that follows from the normalization of all projections.

Computational Resources. The reconstruction is run on a Nvidia V100 GPU with 18GB memory. Each epoch lasts 10 minutes. The algorithm is run for 40 epochs which, in the current implementation, takes 400 minutes.

Additional Synthetic Data (Supplementary Figure 1)

Synthetic dataset. The data is generated similarly to the main synthetic experiment, with the exception of some changes to the imaging conditions. In a first case, the noise level is set to a SNR of -5.2 dB. This corresponds to the energy of the noise being almost four times that of the signal. In the second case, the SNR is kept at -20 dB (the same as for the main experiment), but the projections are also translated. The translation (both horizontal and vertical) for each projection is sampled from a zero-mean symmetric triangular distribution whose total width is 6% of the image size from the centre. This corresponds to a shift of maximum 5 pixels from the centre in each direction.

Reconstruction Settings. The reconstruction settings for both cases are the same than the ones used in the main experiment, except for the few following differences. For the second case, the translations are also imposed, and the translation distribution is kept the same as the one used for generating the dataset. Furthermore, in both cases, the lower bound of the clipping value at the centre reaches -5% of the maximum voxel value of the volume. Finally, the algorithm is run for 100 epochs for both experiments.

Experimental Data (Figure 3)

Dataset. The dataset consists of 41,123 β-galactosidase (EMPIAR-10061) projections extracted from 1,539 micrographs [17]. The projections of size (384×384) are downsampled to (192×192), with a pixel size of 1.274 Å. This corresponds to a Nyquist resolution of 2.548 Å for a reconstructed volume of size $(180 \times 180 \times 180)$.

The dataset is randomly divided into two halves. The algorithm is applied separately on both halves to generate the half-maps. The defocus and astigmatism parameters of the CTF are estimated from each micrograph using Relion [9].

Generator Settings. We reconstruct a volume of size $(180 \times 180 \times 180)$ voxels for each half dataset. The pixel size is 1.274 Å. The volumes are initialized with zeros. The D2 symmetry of β-galactosidase is enforced during reconstruction. A uniform distribution is assumed for the poses. The CTF parameters estimated with CTFFIND4 [62] are used in the forward model of the cryo-EM physics simulator. The parameter B of the envelope function of the CTF (see (17)) decays to 0.4 at the Nyquist frequency. The translations are set to zero.

To handle the noise, we randomly extract (prior to the learning procedure) 41,123 patches of size (384×384) from the background of the micrographs at locations where particles do not appear; this is done by identifying patches with the lowest variance. We extract as many noise patches per micrograph as we have particle images. Each noise patch is then downsampled to size (192×192) and normalized. Then, during run-time, the noise patches are sampled from this collection, scaled, and added to the simulated projections. For consistency, the noise patch added to a given simulated projection is taken from the same micrograph as the one that was used to estimate the CTF parameters previously applied to that specific projection. The scaling operation weights the contribution of the noise with respect to the projection signal. This is handled by multiplying the pixel values of the noise patches and the projections by two scalars that are learnt throughout the procedure. These two scalar values are the same for every pair of noise/projection images, so that

the same amount of extracted noise is added to every simulated projection.

We apply a binary spherical mask of size $(171 \times 171 \times 171)$ on the learned volume. To handle the sharp transitions at the mask borders, we enforce constraints on the masked volume. These are similar to those used in the synthetic experiment, with difference that the lower bound, $V_{\min}$ decreases to -5% of the maximum voxel value.

Discriminator Architecture. The architecture of the discriminator network is detailed in *Online Methods*. The discriminator is initialized identically for the two half datasets. The projection images (real and simulated) are smoothed with a Gaussian kernel before being fed to the discriminator. The standard deviation of the kernel is initially set at 2 and changes in every iteration so that it decreases by a total of 2% in each epoch.

General Settings. The adversarial learning scheme is implemented in Pytorch [19]. For the optimization, we use [33] ($\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$) with a learning rate of 10^{-3} and a batch size of 8. The algorithm is run for 15 epochs and the learning rate decreases by 8% at every epoch. The parameter for the gradient-penalty term is kept to $\lambda = 1$ (see (10)). The discriminator is trained 4 times for every training of the generator ($n_{\text{discr}} = 4$ in Algorithm 1).

For the back-propagations, the norm of the gradients for the discriminator are clipped to a maximal value of 10^7. For the generator, the gradients for each pixel are clipped to a maximal value of 10^3. The clipping values increase linearly from zero to those maxima in the first two epochs. Doing so improves on the stability of the adversarial learning scheme in the start, in particular, on that of the discriminator. The gradients that correspond to the learning of the scaling ratios between the noise and projection images are clipped to a value of 10.

Computational Resources. The reconstruction is run on a Nvidia V100 GPU with 18GB memory. Each epoch lasts 10 minutes. The algorithm is run for 15 epochs, which takes 150 minutes.