CryoGAN: A New Reconstruction Paradigm for Single-particle Cryo-EM Via Deep Adversarial Learning

Harshit Gupta[1], Michael T. McCann[1], Laurène Donati, Michael Unser

Biomedical Imaging Group, EPFL, Switzerland

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 measured particles have unknown poses. Current reconstruction techniques either estimate the poses or marginalize them away—steps that are computationally challenging. CryoGAN sidesteps this problem by using a generative adversarial network (GAN) to learn the 3D structure whose simulated projections 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 cryo-EM physics simulator. CryoGAN is an unsupervised algorithm that only demands picked particle images and CTF estimation as inputs; no initial volume estimate or prior training are needed. Moreover, it requires minimal user interaction and can provide reconstructions in a matter of hours on a high-end GPU. Experiments on synthetic datasets confirm that CryoGAN can reconstruct a high-resolution volume with its adversarial learning scheme. Preliminary results on real β-galactosidase data demonstrate its ability to capture and exploit real data statistics in more challenging imaging conditions. Given the plurality of technical developments ahead in GAN architectures, further gain in resolution is expected in the near-future for this new paradigm.

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

[1]Equal contributions

the acquired 2D measurements [4–11]. These sets of sophisticated algorithms, which include projection matching approaches, maximum likelihood optimization frameworks or regularized methods, enable the determination of structures with unprecedented atomic resolutions.

The reconstruction procedure in single-particle cryo-EM remains nonetheless riddled with complex obstacles. The task equates to a high-dimensional nonconvex optimization problem with numerous local minima, and the outcome of the global process usually depends 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 measured particles have unknown poses. To handle this, current methods either estimate the poses [9] or marginalize them in likelihood-based optimization procedures [11]. Both are computationally challenging approaches that can demand large resources or rely on approximations.

To overcome 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 [15]. Similar to GANs, CryoGAN is driven by the competitive training of two entities: one that captures the real data distribution, and another that discriminates between generated samples and samples from the real dataset. In a classical GAN, the two entities are convolutional neural networks (CNNs)—respectively known as the generator and the discriminator—that are trained simultaneously using backpropagation (Figure 1a). The important twist with CryoGAN is that we replace the generator network by a cryo-EM physics simulator (Figure 1b). By doing so, CryoGAN learns the 3D density map whose simulated projections are the most consistent with a given 2D measurement set 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. Thanks to this innovative setting, CryoGAN sidesteps many cumbersome processing steps, such as 2D alignment and 2D/3D classification. It also avoids many of the computational drawbacks associated with likelihood-based methods, in particular the need to marginalise over all poses via numerical integration.

In practice, CryoGAN requires no prior knowledge of the 3D structure; its learning process is purely unsupervised and data-driven. At a minimum, the user needs only to feed the picked particles and CTF estimations to the algorithm. No initial estimate of the volume is needed: the

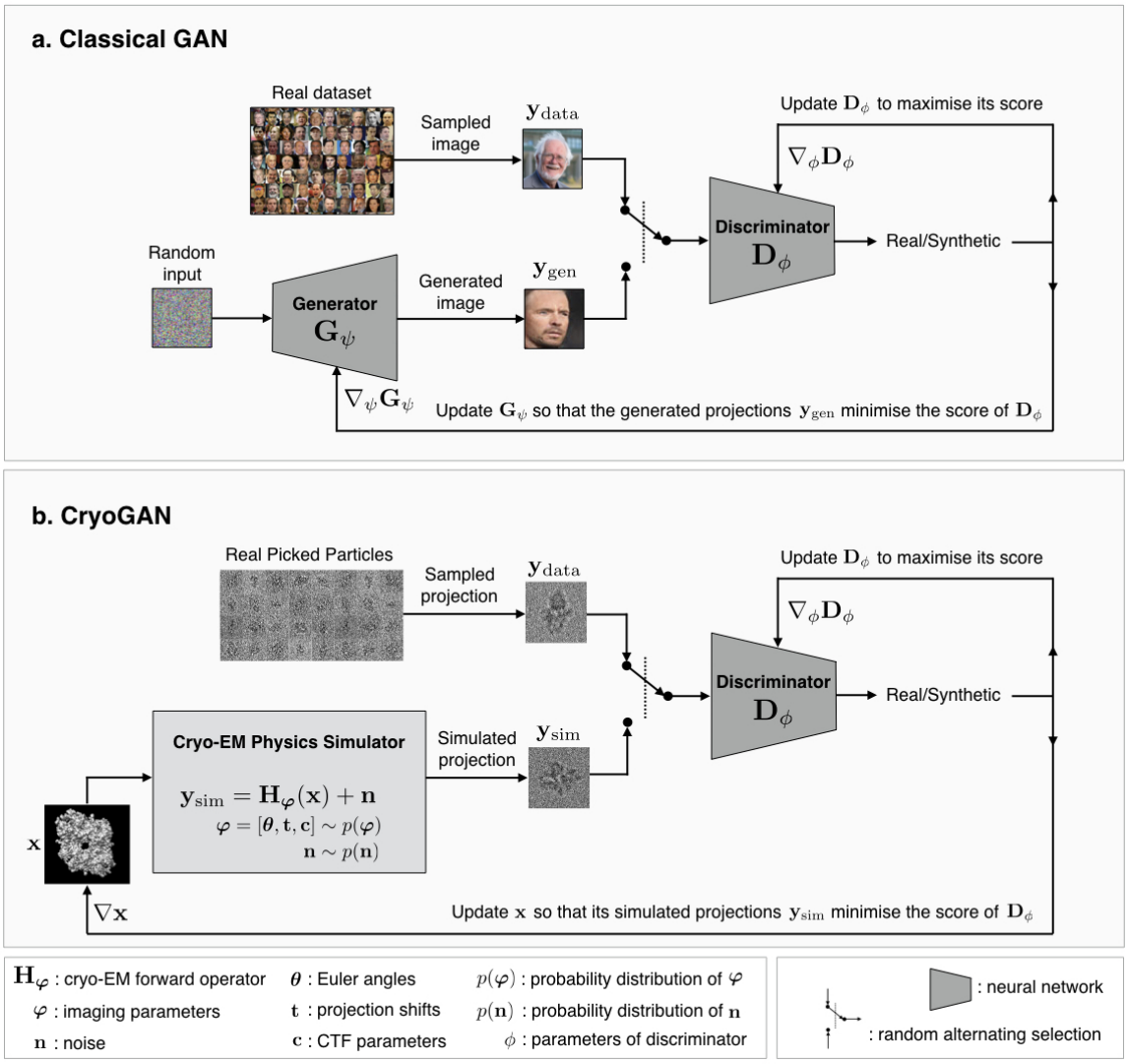

Figure 1: A schematic comparison between **(a)** a classical GAN architecture and **(b)** the CryoGAN architecture. Both frameworks rely on a deep adversarial learning scheme to capture the distribution of real data. CryoGAN exploits this ability to look for the volume whose simulated measurements have a distribution that matches this real data distribution. This is achieved by adding a "cryo-EM physics simulator" that produces synthetic 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.

algorithm starts with a volume initialized with zeros. The CryoGAN framework is backed up by a comprehensive mathematical framework that provides guarantees on the recovery of the volume under a given set of assumptions 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 via computer simulation. The results demonstrate that our unsupervised reconstruction paradigm permits accurate recovery of a high-resolution 3D structure (Figure 2). We then deployed CryoGAN on a real β-galactosidase dataset [16], reaching

a resolution of 7.99Å in under 200 minutes in far more challenging conditions (Figure 3). These preliminary results are a strong indication of the viability 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 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 scheme in image reconstruction. The proposed paradigm opens many new perspectives in single-particle cryo-EM reconstruction and paves the way for more applications beyond the present one.

RESULTS

The CryoGAN Algorithm

CryoGAN is like a classical 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 synthetic measurement based on 1) the current volume-estimate and 2) a given 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 that of standard GANs. The role of the discriminator in CryoGAN is to encourage the simulator to learn the volume whose simulated dataset distribution matches that of the real dataset, while it simultaneously gets better at evaluating the simulated projections for authenticity. Thanks to this novel adversarial-learning scheme, CryoGAN is able to output the volume that best explains the statistics of a provided set of particle images.

CryoGAN is based on a sound mathematical framework that provides guarantees on the recovery of the volume that best explains the measurements, under a given set of assumptions. Its adversarial learning scheme falls under the framework of Wasserstein GANs (WGANs), with the key architectural difference mentioned above. The algorithm alternates between updates of the discriminator and the volume with stochastic gradient descents, and is implemented in PyTorch [17]. The complete mathematical and algorithmic descriptions of CryoGAN are given in the *Online Methods*.

Performance on a Synthetic Dataset

We first assessed the viability and performance of CryoGAN on a synthetic dataset consisting of 41,000 β-galactosidase particles. To generate this dataset of "picked particles", we fitted the protein's

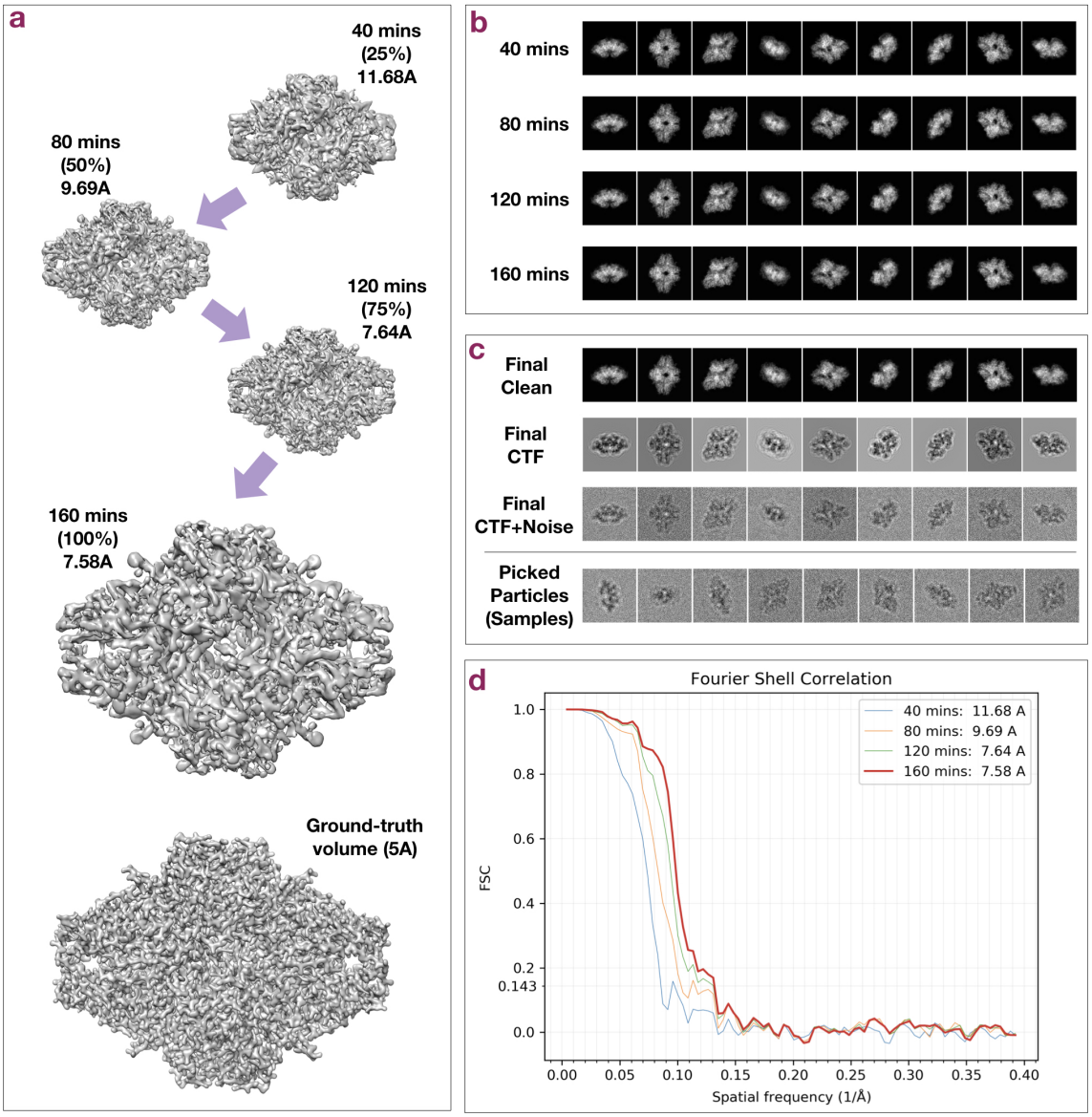

Figure 2: CryoGAN is applied on a synthetic dataset (dubbed "picked particles") generated from a 5Å β-galactosidase volume. **(a)** Starting with zero-values, the volume is progressively updated to produce projections whose distribution matches that of the picked particles. **(b)** Evolution during training of some "clean" projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. **(c)** *Row 1*: Clean projections (before CTF and noise) generated at the final stage of training. *Row 2*: CTF-modulated projections (before noise) generated at the final stage of training. *Row 3*: Realistic projections (with CTF and noise) generated at the final stage of training. *Row 4*: Samples of picked particles, for comparison. **(d)** FSC curves between the two reconstructed half-maps at different training times.

PDB entry (*5a1a*) with a 5Å-resolution density map, and applied the forward model described in the *Online Methods* to obtain thousands of 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. The complete details behind the experimental conditions are given in the *Supplementary Materials*.

We ran the CryoGAN algorithm for 160 minutes on a NVIDIA P100 GPU to obtain a reconstruction of 7.58Å resolution (Figure 2.a). Starting from a zero-valued volume, CryoGAN progressively updates the 3D structure so that its simulated projections—generated by the cryo-EM physics simulator and displayed in Figure 2.b—reach a distribution that matches that of the particles dataset. These gradual updates are at the core of the deep adversarial learning scheme of CryoGAN. At each algorithm iteration, the gradients from the discriminator (see Figure 1.b) carry information about the current difference between the picked particles ("real data") and the generated projections ("fake data"). Those gradients are used by the cryo-EM physics simulator to update itself and learn a volume that yields more realistic projections. Hence, at the end of its run, the volume learned by CryoGAN has projections (Figure 2.c, rows 1-3) that are similar to the picked particles (Figure 2.c, row 4) in a *distributional* sense. The evolution of the Fourier-shell correlation (FSC) curves between the reconstructed half-maps (Figure 2.d) shows the progressive increase in resolution that derives from this adversarial learning scheme.

Results on Real Data

We then deployed CryoGAN on 41,123 β-galactosidase particles (EMPIAR-10061) to assess its capacity to reconstruct real, experimental data. The dataset obviously represents a much more challenging test-case for CryoGAN, whose adversarial learning scheme relies on our ability to faithfully model the imaging physics of cryo-EM. Here as well, we randomly divided the dataset in two and applied CryoGAN separately on both halves. The complete details behind this experiment are given in the *Supplementary Materials*.

We ran CryoGAN for 160 minutes to obtain a 3D reconstruction wit 7.99Å resolution using a NVIDIA P100 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 reconstruction pipeline (*e.g.*, the assumption of uniform pose distribution). Using this prior knowledge and its adversarial learning scheme, CryoGAN progressively converges toward the reconstruction that best explains the statistics of the dataset (Figure 3.a). As for the synthetic experiment, this is achieved by exploiting the gradients of the discriminator to update the simulator and the current volume estimate, so that the projections generated at later iterations (Figure 3.b)

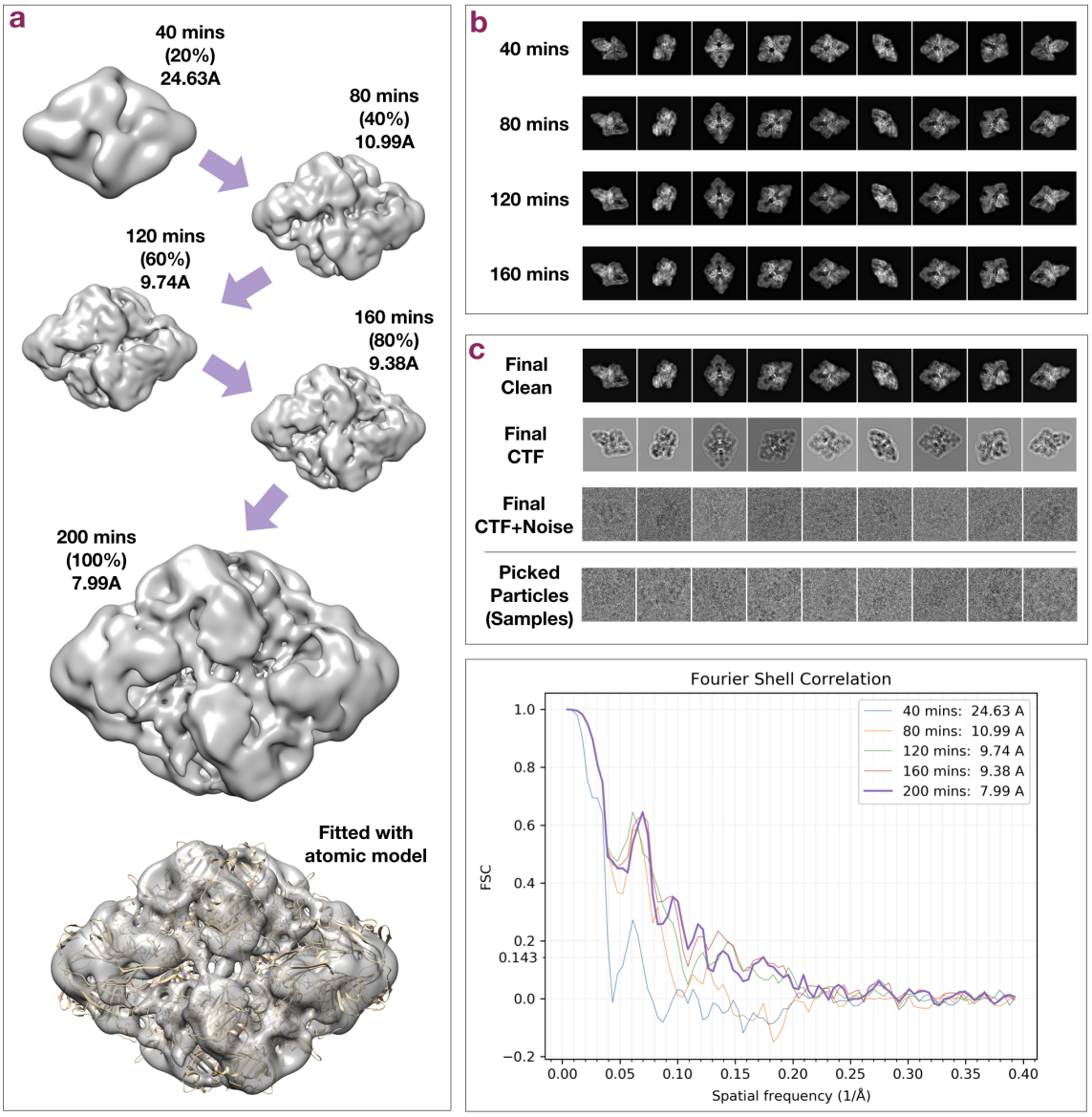

Figure 3: Results of CryoGAN on the real β-galactosidase dataset from [16]. **(a)** Starting with zero-values, the volume is progressively updated to produce projections whose distribution matches that of the real dataset. **(b)** Evolution during training of some "clean" projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. **(c)** *Row 1*: Clean projections (before CTF and noise) generated at the final stage of training. *Row 2*: CTF-modulated projections (before noise) generated at the final stage of training. *Row 3*: Realistic projections (with CTF and noise) generated at the final stage of training. *Row 4*: Samples of picked particles, for comparison. **(d)** FSC curves of the two reconstructed half-maps at different training times.

gradually follow a distribution that approaches that of the real dataset. Higher-resolution details are thus progressively introduced in the estimated volume throughout the run, as illustrated by the FSC curves between successive reconstructed half-maps (Figure 3.d). For this particular run, this resulted in a 7.99Å β-galactosidase structure whose synthetic projections closely resemble the real picked particles, both visually (Figure 3.c) and—more importantly—statistically.

DISCUSSION

We demonstrated the ability of CryoGAN to autonomously reconstruct 3D density maps through its 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 distribution, the CryoGAN algorithm looks for the reconstruction most consistent with the measurements in a *distributional* sense. Hence, it is able to avoid the whole angular-assignment procedure by directly exploiting the statistics of the provided dataset. CryoGAN is a completely 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 under specific assumptions (*i.e.*, the validity of image formation model). When these assumptions are met, our main theorem (see *Supplementary Materials*) asserts that CryoGAN samples the proper probability distribution and recovers the correct 3D volume.

An important point is that CryoGAN bypasses angular-assignment in a likelihood-free manner, which is in contrast with likelihood-based approaches, used for example in CryoSPARC [11]. This permits CryoGAN to avoid marginalizing over the angles, a complex but necessary task in likelihood-based approaches that requires the approximation of integrals by sums. CryoGan also sidesteps many cumbersome processing steps, *e.g.*, 2D alignment or 2D/3D classification, which further reduces the need for user-dependent inputs.

Our synthetic experiments demonstrate the ability of CryoGAN to gradually resolve a structure so that its simulated projections distribution progressively matches that of the experimental picked particles. These results validate the CryoGAN paradigm and the viability of its current implementation: 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 particles, and to learn the volume that best explains these statistics.

The results on the real β-galactosidase dataset demonstrate the capacity of CryoGAN to perform reconstruction in challenging real imaging conditions. The implementation of the CryoGAN algorithm is bound to further improve, and several interesting developmental steps still lie ahead. In particular, we expect the ongoing progresses in deep-learning architectures to help enrich the

fast-evolving CryoGAN algorithm so that additional gain in resolution can be obtained in the near-future.

Roadmap for Future Work

The current implementation of CryoGAN is at the "proof-of-concept" stage and could benefit from several algorithmic refinements. For example, we expect the speed could be increased by using a fast Fourier transform-based projector. A fine tuning of the global CryoGAN architecture (*e.g.*, number of convolutional layers in the discriminator) could further improve its performance, leading to faster convergence. This would consequently increase the resolution that can be obtained in a given amount of time. The use of a larger discriminator and/or dataset size—when GPUs permit—would likely have the same effect.

Like most reconstruction algorithms, CryoGAN can fail if the provided dataset contains a disproportionate amount of corrupted particle images (*e.g.* 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 in the CryoGAN learning scheme a step that automatically spots and discards corrupted data so that the discriminator never sees them. Another interesting option is to directly simulate the patches of non-aligned micrographs/frames (rather than picked particles), and match their distribution to that of the raw dataset. Doing so would also allow cryoGAN to bypass additional preprocessing tasks, in particular particle picking. Finally, recent DL-based approaches able to track outliers in data could prove useful [18].

Similar to the likelihood-based methods, the CryoGAN algorithm requires the specification of the distribution of poses. One could also parametrize the pose distribution and learn its parameters during the reconstruction procedure [19]. A similar approach could be used to calibrate the distribution of the projection translations.

On the theoretical side, we currently have mathematical guarantees on the recovery of volumes for which the assumed pose distribution (be it uniform or not) matches the distribution of the real data. Moreover, we have prior mathematical indication that this can also be achieved when there is a certain mismatch between the assumed pose distribution and the actual one, given that an appropriate GAN loss is used.

The performance of the cryo-EM physics simulator should 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 relies on an additive noise model in its image formation model. One could go even further and consider a Poisson-noise-based forward model [20, 21]. This would, however, require backpropagating through a Poisson distribution, a non-trivial operation at this

stage of development.

Another promising direction of research is the use of a coarse-to-fine strategy to reconstruct the volume progressively at higher and higher resolutions. The motivation is that an increased robustness during the low-resolution regime tends to positively impact the convergence of all the subsequent higher-resolution steps. Several GAN architectures rely on such frameworks, such as the progressive GANs [22] and the styleGANs [23]. The benefits of multi-scale refinement could be considerable for CryoGAN given the extremely challenging imaging conditions faced in single-particle cryo-EM, which make the convergence of optimization algorithms non-trivial. The core idea here would be to have the discriminator learn to differentiate between real and synthetic distributions at a low resolution first, and then at successively higher ones. The impact on CryoGAN could be as important as of the one it had on GANs, which progressed from generating blurry facial images [15] to synthetic images non-distinguishable from real facial images [22, 23] in just a few years. More generally, the new upcoming tools and extensions in GAN architectures could bring significant gain in resolution to the CryoGAN implementation.

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 harder part of the job, which is to obtain a reasonable structure by using only 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. Moreover, we have laid out a roadmap of future improvements that should get us to this stage. Our hope is that this new take on GANs will foster developments beyond the present application in cryo-EM.

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 Global-BioIm: Global integrative framework for computational bio-imaging.

References

[1] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz, "Cryo-electron microscopy of vitrified specimens," *Quarterly Reviews of Biophysics*, vol. 21, no. 2, pp. 129–228, 1988.

[2] R. Henderson, J. M. Baldwin, T. Ceska, F. Zemlin, E. a. Beckmann, and K. H. Downing, "Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy," *Journal of molecular biology*, vol. 213, no. 4, pp. 899–929, 1990.

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

[4] J. Frank, B. Shimkin, and H. Dowse, "Spider—a modular software system for electron image processing," *Ultramicroscopy*, vol. 6, no. 4, pp. 343–357, 1981.

[5] C. Sorzano, R. Marabini, J. Velázquez-Muriel, J. R. Bilbao-Castro, S. H. Scheres, J. M. Carazo, and A. Pascual-Montano, "Xmipp: A new generation of an open-source image processing package for electron microscopy," *Journal of Structural Biology*, vol. 148, no. 2, pp. 194–204, 2004.

[6] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "Eman2: An extensible image processing suite for electron microscopy," *Journal of Structural Biology*, vol. 157, no. 1, pp. 38–46, 2007.

[7] N. Grigorieff, "Frealign: High-resolution refinement of single particle structures," *Journal of Structural Biology*, vol. 157, no. 1, pp. 117–125, 2007.

[8] M. Hohn, G. Tang, G. Goodyear, P. R. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams, and S. J. Ludtke, "Sparx, a new environment for cryo-em image processing," *Journal of Structural Biology*, vol. 157, no. 1, pp. 47–55, 2007.

[9] S. H. Scheres, "Relion: Implementation of a bayesian approach to cryo-em structure determination," *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.

[10] J. de la Rosa-Trevín, A. Quintana, L. del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, J. Gómez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Otón, G. Sharov, J. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C. Sorzano, and J. Carazo, "Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy," *Journal of Structural Biology*, vol. 195, no. 1, pp. 93 – 99, 2016.

[11] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination," *Nature Methods*, vol. 14, no. 3, pp. 290–296, Feb. 2017.

[12] C. O. S. Sorzano, R. Marabini, A. Pascual-Montano, S. H. Scheres, and J. M. Carazo, "Optimization problems in electron microscopy of single particles," *Annals of Operations Research*, vol. 148, no. 1, pp. 133–165, 2006.

[13] R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes, J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, and C. L. Lawson, "Outcome of the first electron microscopy validation task force meeting," *Structure*, vol. 20, no. 2, pp. 205–214, 2012.

[14] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 58–76, 2020.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[16] A. Bartesaghi, A. Merk, S. Banerjee, D. Matthies, X. Wu, J. L. Milne, and S. Subramaniam, "2.2 å resolution cryo-em structure of β-galactosidase in complex with a cell-permeant inhibitor," *Science*, vol. 348, no. 6239, pp. 1147–1151, 2015.

[17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.

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

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

[20] M. Vulović, R. B. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lücken, H. Rullgård, O. Öktem, and B. Rieger, "Image formation modeling in cryo-electron microscopy," *Journal of structural biology*, vol. 183, no. 1, pp. 19–32, 2013.

[21] H. Rullgård, L.-G. Öfverstedt, S. Masich, B. Daneholt, and O. Öktem, "Simulation of transmission electron microscope images of biological specimens," *Journal of microscopy*, vol. 243, no. 3, pp. 234–256, 2011.

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

[23] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

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

[25] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, 2017, pp. 214–223.

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

[27] G. Peyré, M. Cuturi *et al.*, "Computational optimal transport," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, 2017, pp. 5767–5777.

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

[30] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The astra toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[31] P. A. Penczek, R. A. Grassucci, and J. Frank, "The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles," *Ultramicroscopy*, vol. 53, no. 3, pp. 251 – 270, 1994.

[32] T. Baker and R. Cheng, "A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy," *Journal of Structural Biology*, vol. 116, no. 1, pp. 120–130, 1996.

[33] Z. Kam, "The reconstruction of structure from electron micrographs of randomly oriented particles," in *Electron Microscopy at Molecular Dimensions*. Springer, 1980, pp. 270–277.

[34] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[35] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, Nov. 2017.

[36] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, 2019.

[37] T. Bepler, A. J. Noble, and B. Berger, "Topaz-denoise: general deep denoising models for cryoem," *bioRxiv*, p. 838920, 2019.

[38] F. Wang, H. Gong, G. Liu, M. Li, C. Yan, T. Xia, X. Li, and J. Zeng, "Deeppicker: a deep learning approach for fully automated particle picking in cryo-em," *Journal of structural biology*, vol. 195, no. 3, pp. 325–336, 2016.

[39] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy," *BMC bioinformatics*, vol. 18, no. 1, p. 348, 2017.

[40] D. Tegunov and P. Cramer, "Real-time cryo-em data pre-processing with warp," *BioRxiv*, p. 338558, 2018.

[41] T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel, T. Raisch, D. Prumbaum *et al.*, "Sphire-cryolo is a fast and accurate fully automated particle picker for cryo-em," *Communications Biology*, vol. 2, no. 1, p. 218, 2019.

[42] T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, and B. Berger, "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature methods*, pp. 1–8, 2019.

[43] E. D. Zhong, T. Bepler, J. H. Davis, and B. Berger, "Reconstructing continuous distributions of 3D protein structure from cryo-em images," in *International Conference on Learning Representations*, 2020.

[44] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[45] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, p. 124007, 2017.

[46] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent ct image reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1440–1453, 2018.

[47] A. Bora, E. Price, and A. G. Dimakis, "AmbientGAN: Generative models from lossy measurements." *ICLR*, vol. 2, p. 5, 2018.

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

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

[50] M. Vulović, R. B. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lücken, H. Rullgård, O. Öktem, and B. Rieger, "Image formation modeling in cryo-electron microscopy," *Journal of Structural Biology*, vol. 183, no. 1, pp. 19–32, Jul. 2013.

[51] V. M. Panaretos *et al.*, "On random tomography with unobservable projection angles," *The Annals of Statistics*, vol. 37, no. 6A, pp. 3272–3306, 2009.

[52] S. Helgason, *The radon transform*. Springer, 1980, vol. 2.

[53] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera—a visualization system for exploratory research and analysis," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

ONLINE METHODS

Image Formation Model in Single-Particle Cryo-EM

We model the single-particle cryo-EM imaging procedure [24] by the linear relationship

$$\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}, \tag{1}$$

where

- $\mathbf{y} \in \mathbb{R}^M$ is a 2D projection of size $M = M_1 \times M_2$;
- $\mathbf{x} \in \mathbb{R}^V$ is the 3D density map of size $V = V_1 \times V_2 \times V_3$;
- $\mathbf{H}_{\boldsymbol{\varphi}} \in \mathbb{R}^{M \times V}$ denotes the forward operator (see (2) below) with parameters $\boldsymbol{\varphi}$;
- $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$ is the set of imaging parameters. It includes 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;
- $\mathbf{n} \in \mathbb{R}^M$ represents an additive noise following a distribution $p_{\mathbf{n}}$.

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

$$\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C}_{\mathbf{c}}\mathbf{S}_{\mathbf{t}}\mathbf{P}_{\theta}. \tag{2}$$

It is composed of the projection operator $\mathbf{P}_{\boldsymbol{\theta}} : \mathbb{R}^V \to \mathbb{R}^M$, the shift operator $\mathbf{S}_{\mathbf{t}} : \mathbb{R}^M \to \mathbb{R}^M$, and the convolution operator $\mathbf{C}_{\mathbf{c}} : \mathbb{R}^M \to \mathbb{R}^M$. A more detailed description of the physics behind this image formation model $\mathbf{H}_{\boldsymbol{\varphi}}$ is given in *Supplementary Materials*.

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 $\mathbf{x}_{\text{true}}$.

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

$$p(\mathbf{y}|\mathbf{x}) = \int_{\boldsymbol{\varphi}} p_{\mathbf{n}}(\mathbf{y} - \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}) p_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\varphi}, \tag{3}$$

where $p_{\boldsymbol{\varphi}}$ is the distribution of the imaging parameters $\boldsymbol{\varphi}$. We denote $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$.

In our formulations, the projections in the real dataset are samples of a distribution p_{data}. We then make the assumption that the distribution $p(\mathbf{y}|\mathbf{x}_{\text{true}})$ corresponds to the distribution of the real dataset, *i.e.* $p(\mathbf{y}|\mathbf{x}_{\text{true}}) = p_{\text{data}}(\mathbf{y})$, which is reasonable if the image-formation model faithfully mimics the cryo-EM physics.

We demonstrate in Theorem 1 in *Supplementary Materials* that two 3D volumes $\mathbf{x}_1$ and $\mathbf{x}_2$ have identical conditional distributions, *i.e.* $p(\mathbf{y}|\mathbf{x}_1) = p(\mathbf{y}|\mathbf{x}_2)$, if and only if $\mathbf{x}_1$ is equal to $\mathbf{x}_2$ (up to rotation and reflection). Hence, Theorem 1 implies that for the reconstruction $\mathbf{x}_{\text{rec}}$ to be perfect (*i.e.*, $\mathbf{x}_{\text{rec}} = \mathbf{x}_{\text{true}}$), it must satisfy $p(\mathbf{y}|\mathbf{x}_{\text{rec}}) = p(\mathbf{y}|\mathbf{x}_{\text{true}})$.

This is a mathematical result of importance as it means we can formulate the reconstruction task as the minimization problem

$$\mathbf{x}_{\text{rec}} = \arg\min_{\mathbf{x}} D\big(p(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\mathbf{x}_{\text{true}})\big) \tag{4}$$

$$= \arg\min_{\mathbf{x}} D\big(p_{\mathbf{x}}(\mathbf{y}), p_{\text{data}}(\mathbf{y})\big), \tag{5}$$

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

We use for (5) 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 stable than other popular distances (*e.g.*, Total-Variation or Kullback-Leibler) for this kind of applications [25].

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 [25–27], this minimization problem can also be stated in its dual form

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

where the function f belongs to the set of functions with Lipschitz value $\|f\|_L$ less than 1.

CryoGAN and the connection with WGANs

A key observation is that Equation (8) falls under the framework of the generative adversarial networks (GANs) [15] called WGANs (for Wasserstein-GANs) [25].

In the classical 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 the discriminator is to learn to differentiate between samples originating from real data and samples originating from synthetic data. These synthetic data are produced by another neural network called the generator that aims at producing data realistic-enough to "fool" the discriminator. This adversarial learning scheme drives the WGAN to progressively capture the distribution of the real data.

The idea behind CryoGAN is that we learn the volume $\mathbf{x}$ whose simulated projections follow the real-data distribution captured through the adversarial learning scheme. In terms of architecture, the key difference with classical WGANs is that we replace the generator network by a cryo-EM physics simulator (see "The Cryo-EM Physics Simulator" below). Its role is to produce synthetic projections of a volume estimate $\mathbf{x}$ using the image formation model (1). These simulated projections then follow a distribution $\mathbf{y} \sim p_{\mathbf{x}}$.

Hence, Equation (8) translates into

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

As proposed in [28], the Lipschitz constraint $\|\mathbf{D}_\phi\|_L < 1$ is best 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:

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

Here, p_{int} describes the uniform distribution along the straight line between points sampled from p_{data} and $p_{\mathbf{x}}$ and $\lambda \in \mathbb{R}_+$ is an appropriate penalty coefficient (see [28], Section 4).

The CryoGAN Algorithm

Equation (10) is a min-max optimization problem. By replacing the expected values by their empirical counterparts (sums) [28], 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 real projection sampled from the acquired dataset;
- $\mathbf{y}^n_{\text{sim}} \sim p_{\mathbf{x}}$ is a synthetic projection of the current estimate $\mathbf{x}$ generated by the cryo-EM physics simulator;
- $\mathbf{y}^n_{\text{int}} = \alpha_n \cdot \mathbf{y}^n_{\text{data}} + (1 - \alpha_n) \cdot \mathbf{y}_{\text{sim}}$, where α_n is sampled from a uniform distribution between 0 and 1.

In practice, we minimize (11) with stochastic gradient descent (SGD) using batches. We alternatively update with an Adam optimizer [29] the discriminator $\mathbf{D}_\phi$ (in n_{discr} iterations) and the volume $\mathbf{x}$ (in 1 iteration) using their respective gradients of $L_S(\mathbf{x}, \mathbf{D}_\phi)$. A pseudo-code and a schematic view of the CryoGAN algorithm are given in Algorithm 1 and Figure 4, respectively. The architecture of the CryoGAN discriminator is presented below (see "The CryoGAN Discriminator Network").

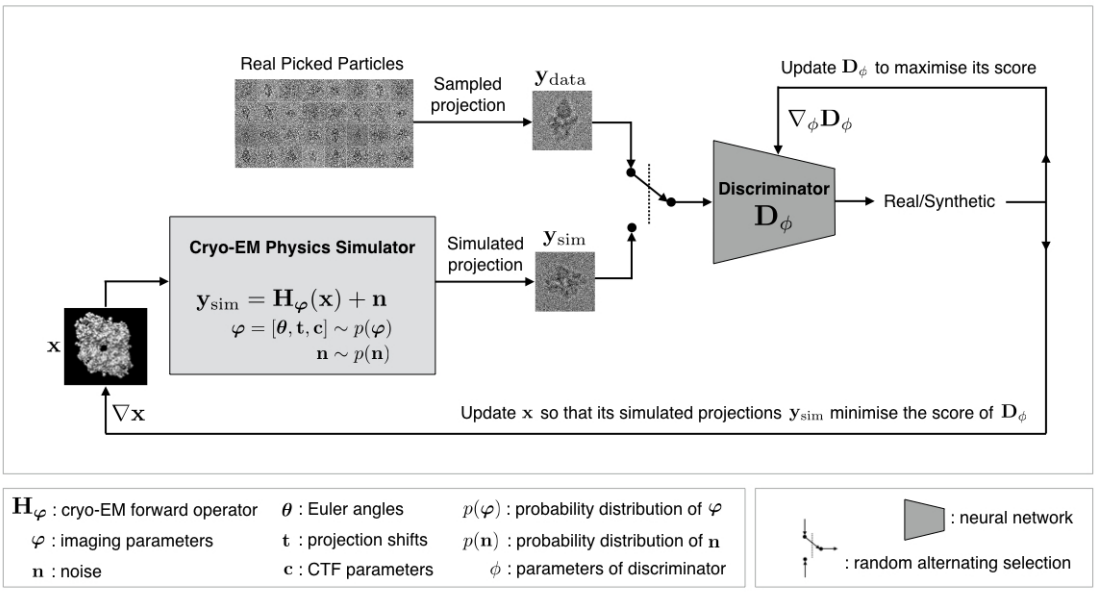

Figure 4: Schematic view of CryoGAN. The algorithm relies on a deep adversarial learning scheme to capture the distribution of real data. CryoGAN exploits this ability to then look for the volume whose simulated measurements have a distribution that matches this real data distribution. This is achieved by adding a "cryo-EM physics simulator" that produces synthetic measurements following a mathematical model of the cryo-EM imaging procedure. Note that the updates involve backpropagating through the neural networks, not indicated here for the sake of clarity.

The Cryo-EM Physics Simulator

The goal of the physics simulator is to sample $\mathbf{y}_{\text{sim}} \sim p_{\mathbf{x}}(\mathbf{y})$. We do this in three steps:

1. Sample the imaging parameters $\boldsymbol{\varphi}$ from the distribution $p_{\boldsymbol{\varphi}}$: $\boldsymbol{\varphi} \sim p_{\boldsymbol{\varphi}}$;

Algorithm 1 CryoGAN

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

for n_{train}**do**

for n_{discr} **do**

- Sample real projections: $\{\mathbf{y}^1_{\text{batch}}, \dots, \mathbf{y}^N_{\text{batch}}\} = \{\mathbf{y}^n_{\text{data}}\}_{n \in B}$;
- Sample projections simulated from current $\mathbf{x}$: $\{\mathbf{y}^1_{\text{sim}}, \dots, \mathbf{y}^N_{\text{sim}}\} \sim p_{\mathbf{x}}$ (see Algorithm 2);
- Sample $\{\alpha_1, \dots, \alpha_n\} \sim U[0, 1]$;
- For all $n \in \{1, \dots, N\}$, compute $\mathbf{y}^n_{\text{int}} = \alpha_n \cdot \mathbf{y}^n_{\text{batch}} + (1 - \alpha_n) \cdot \mathbf{y}^n_{\text{sim}}$;
- Update the parameters ϕ of the discriminator $\mathbf{D}_\phi$ by ascending its stochastic gradient

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

end for

- Sample $\{\mathbf{y}^1_{\text{sim}}, \dots, {\mathbf{y}_{\text{sim}}}^N\} \sim p_{\mathbf{x}}$;
- Update of the volume $\mathbf{x}$ by descending its stochastic gradient

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

end for

2. Generate noiseless CTF-modulated and shifted projections from the current volume estimate $\mathbf{x}$: $\mathbf{H}_{\boldsymbol{\varphi}}(\mathbf{x})$;
3. Sample the noise model to simulate noisy projections: $\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}(\mathbf{x}) + \mathbf{n}$ where $\mathbf{n} \sim p_{\mathbf{n}}$.

A pseudo-code of the cryo-EM Physics Simulator is given in Algorithm 2.

Step 1: Sampling the imaging parameters $\boldsymbol{\varphi}$

We 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 also sampled from a prior distribution $p_{\mathbf{t}}$.

The CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$ can also be sampled from a prior distribution $p_{\mathbf{c}}$. For example, a uniform distribution over the defocus range can be assumed. In practice, we exploit the fact that the CTF parameters can often be efficiently estimated for the micrographs, and we then uniformly sample from the whole set of extracted CTF parameters.

Algorithm 2 Cryo-EM Physics Simulator

Inputs: current volume estimate $\mathbf{x}$

1. Sample the imaging parameters $\boldsymbol{\varphi} = [\boldsymbol{\theta}, \mathbf{t}, \mathbf{c}]$, *i.e.*:
   - Sample the Euler angles: $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \sim p_{\boldsymbol{\theta}}$;
   - Sample the 2D shifts: $\mathbf{t} = (t_1, t_2) \sim p_{\mathbf{t}}$;
   - Sample the CTF parameters: $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}}) \sim p_{\mathbf{c}}$;
2. Generate a synthetic noiseless projection based on (2): $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$;
3. Sample the noise: $\mathbf{n} \sim p_{\mathbf{n}}$. Add to the projection: $\mathbf{y}_{\text{sim}} = \mathbf{y}_{\text{noiseless}} + \mathbf{n}$.

Step 2: Simulation of noiseless cryo-EM projections

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 (2) is implemented using the ASTRA toolbox [30].

Step 3: Sampling of the noise model

The precise modeling of the noise is a particularly challenging feat in cryo-EM. To produce noise realizations 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 in Step 1 to estimate the CTF parameters previously applied to that specific projection. Additional details for this implementation are given in the *Supplementary Materials*.

The CryoGAN Discriminator Network

The role of the discriminator is to learn to differentiate between projections from the real dataset and projections generated by the cryo-EM physics simulator (*i.e.*, "fake" images). The gradients from the discriminator (see Equation (13) in Algorithm 1) carry information on the difference between real and fake images at a given run-time. Those gradients are used by the simulator to update itself and learn a volume that generates more realistic projections.

The discriminator network takes an image as input and outputs a single value. Its architecture is illustrated in Figure 5. 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 are 96, 192, 384, 768, 1536, and 3072, respectively. The filter size 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 then processed by a leaky ReLU. The resulting activations are then fed to the last FC layer to output a scalar.

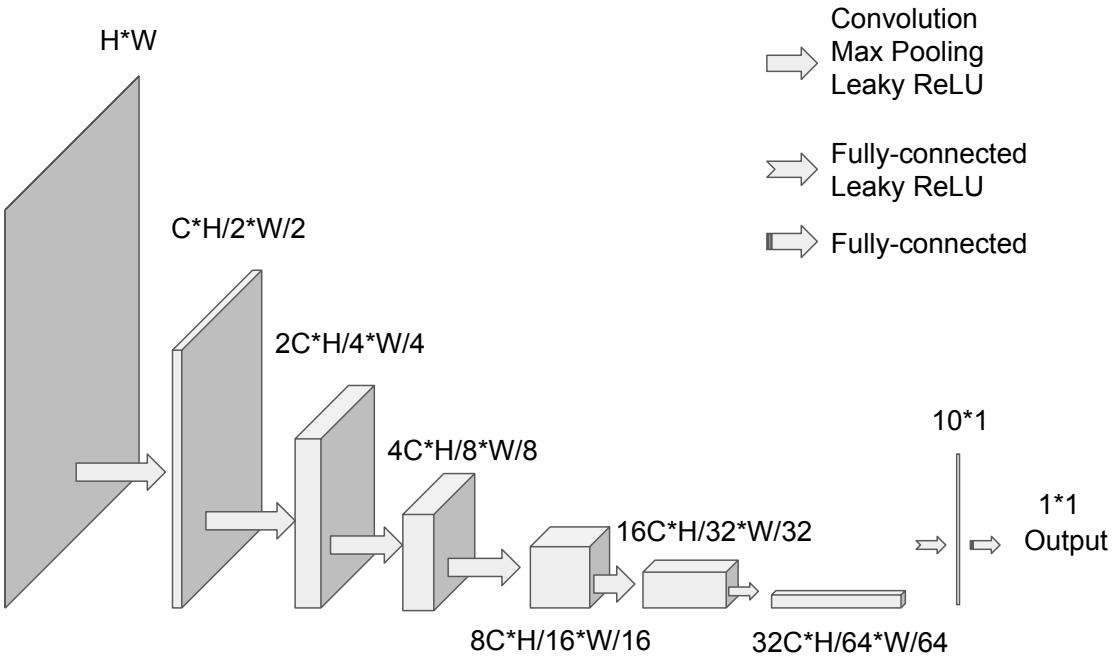

Figure 5: Architecture of the discriminator. It consists of 6 blocks of convolutional layers followed by 2 blocks of fully-connected layers. The parameter for the channel size is $C = 96$ in all the experiments. The input image with size $H \times W$ is successively processed and downsampled to output a scalar.

Related Works

The main challenge in cryo-EM reconstruction is that every particle has an unknown pose in the micrographs. If the poses were known, maximum likelihood (ML) or maximum a posteriori (MAP) estimation of the volume could be performed by solving a standard linear inverse problem, where the large number of measurements would be useful to counteract the low SNR of the measurements. The dominant strategies for cryo-EM reconstruction are likelihood-based; these involve either the estimation of the unknown poses [9] or their marginalization (e.*g.*, first phase of the CryoSPARC package [11]). For a more in-depth discussion of the two approaches, see the review [14].

Cryo-EM Reconstruction With Pose Estimation

Pose estimation can be achieved with a variety of strategies, including the popular projection matching approach [31,32]. 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 param-

eters and the projection of the current reconstructed volume at a large number of angles, at every iteration of the reconstruction pipeline; this is obviously very computationally demanding.

Cryo-EM Reconstruction With Pose Marginalization

Marginalization over the poses, as done in the first phase of [11], is an attractive approach to sidestep many of the shortcomings of pose estimation. In particular, there are no extra parameters to be estimated for every noisy projection image. However, a traditional downside of those methods is their computational cost, as true marginalization requires integration over all poses. Here again this requires projecting the current reconstruction volume at a large number of angles, which inherently involves some degree of approximation. Some ingenuous algorithmic schemes can reduce their computational complexity [11]. Marginalization-based reconstruction approaches usually remain limited to a certain resolution (typically no better than 10Å). In practice, they therefore need to be followed by a pose-estimation-based refinement procedure to obtain a high-resolution volume.

Cryo-EM Reconstruction Without Pose Estimation nor Marginalization

Methods that reconstruct a cryo-EM volume without pose-estimation nor marginalization (hence avoiding the pitfalls of likelihood-based methods) are relatively few, even though a first approach was already proposed in 1980s [33]. This method reconstructs an "ab-initio" structure such that the first few moments of the distribution of its theoretical cryo-EM measurements matches the ones of the picked particles. However, the method assumes that the poses of the picked-particles have a uniform distribution. This moment-matching technique has been recently extended in [19] to reconstruct an "ab-initio" structure in the case of non-uniform pose distribution.

By contrast, our method proposes to *exactly* match the distribution of the theoretical cryo-EM measurements and the distribution of the picked particles, *i.e.*, to match all the moments and not just the first few. Moreover, our method works for any pose distribution of the particles provided the latter 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, similarly to [19].

Deep Learning in Cryo-EM

Deep learning has already had a profound impact in a wide range of image reconstruction applications [34–36]. However, their current utilization in cryo-EM is mostly restricted to pre-processing steps such as micrographs denoising [37] or particle picking [38–42]. A recent work used neural networks to model continuous generative factors of structural heterogeneity [43]. However, the al-

gorithm relies on a pose-estimation procedure that is done using a conventional approach. Another recent work [18] uses a variational autoencoder trained using a discriminator based-objective to find a low dimensional latent representation of the picked particles. These representations are then used for pose estimation.

Unsupervised Deep Learning for Inverse Problems

Deep learning is now extensively used to solve inverse problems in imaging [35, 44–46]. However, most methods are based on supervised learning and thus rely on training data. An unsupervised scheme that needs no training data was recently proposed for general inverse problems [47]. Our CryoGAN method extends it to the harder inverse problem of cryo-EM, both theoretically and experimentally.

Finally, the reconstruction of a 3D structure from its 2D viewpoints (and not projections) is an important problem in computer vision. Many recent deep learning algorithms have been used in this regard [48, 49]. However, the measurement model for these problems is much less complicated than the cryo-EM one, and is thus not straightforwardly applicable to this modality.

SUPPLEMENTARY MATERIALS

Image Formation Theory

For our forward model, we follow the development in [24] (2.1-2.10), [50] and [21], which results in a linear relationship between the 3D Coulomb potential of the molecule and the 3D measurement in the image plane. Specifically, we have in the Fourier domain

$$\mathcal{F}\{y\}(\overrightarrow{\boldsymbol{\omega}}) = \hat{C}(\overrightarrow{\boldsymbol{\omega}})\mathcal{F}\{\mathcal{P}_{\boldsymbol{\theta}}\{f\}\}(\overrightarrow{\boldsymbol{\omega}}) \tag{14}$$

where

- $\mathcal{F}\{\cdot\}$ is the 2D Fourier transform;
- $y : \mathbb{R}^2 \to \mathbb{R}$ is the intensity measured on the image plane;
- $f : \mathbb{R}^3 \to \mathbb{R}$ is the Coulomb potential we aim to recover;
- the transform

$$\mathcal{P}_{\boldsymbol{\theta},\mathbf{r}_0}\{f\}(\mathbf{r}) = \int_{-\infty}^{\infty} f\left(\mathbf{R}_{\boldsymbol{\theta}}\left(\begin{bmatrix}\mathbf{r}\\ z\end{bmatrix} - \begin{bmatrix}\mathbf{r}_0\\ 0\end{bmatrix}\right)\right) \mathrm{d}z \tag{15}$$

  is the X-ray projection of f in a pose specified by the 3D rotation matrix $\mathbf{R}_{\boldsymbol{\theta}}$ and the translation $\mathbf{r}_0$;
- and where $\hat{C} : \mathbb{R}^2 \to \mathbb{R}$ is the Contrast Transfer Function (CTF).

The CTF itself can be written as

$$\hat{C}(\overrightarrow{\boldsymbol{\omega}}) = \hat{C}_p(\overrightarrow{\boldsymbol{\omega}})\hat{E}(\overrightarrow{\boldsymbol{\omega}})\hat{A}(\overrightarrow{\boldsymbol{\omega}}) \tag{16}$$

where

- $\hat{A} : \mathbb{R}^2 \to \mathbb{R}$ is the objective aperture function given by

$$\hat{A}(\overrightarrow{\boldsymbol{\omega}}) = \begin{cases} 1, & \text{for } \|\overrightarrow{\boldsymbol{\omega}}\| \leq \omega_{\text{cutoff}}, \text{ and} \\ 0, & \text{for } \|\overrightarrow{\boldsymbol{\omega}}\| > \omega_{\text{cutoff}}, \end{cases} \tag{17}$$

  where $\omega_{\text{cutoff}} = \frac{2\pi d_{\text{ap}}}{f_l \lambda}$ is the cut-off frequency, f_l is the focal length of the objective lens, and d_{ap} corresponds to the diameter of the aperture.

- $\hat{E} : \mathbb{R}^2 \to \mathbb{R}$ describes the spatial and chromatic envelop function given as

$$\hat{E}(\overrightarrow{\boldsymbol{\omega}}) = \exp\left(-B(\|\overrightarrow{\boldsymbol{\omega}}\|^2)\right), \tag{18}$$

  where $B(\|\overrightarrow{\boldsymbol{\omega}}\|^2)$ is a function influenced by chromatic aberration and spatial incoherence.

- $\hat{C}_p : \mathbb{R}^2 \to \mathbb{R}$ is the phase contrast transfer function that takes the form

$$\hat{C}_p(\overrightarrow{\boldsymbol{\omega}}) = -\sqrt{1 - A^2}\sin(\gamma(\overrightarrow{\boldsymbol{\omega}})) - A^2\cos(\gamma(\overrightarrow{\boldsymbol{\omega}})), \tag{19}$$

  with

$$\gamma(\overrightarrow{\boldsymbol{\omega}}) = \pi\lambda\left(z(\alpha)\|\overrightarrow{\boldsymbol{\omega}}\|^2 - \frac{1}{4}\lambda^3 c_s\|\overrightarrow{\boldsymbol{\omega}}\|^4\right), \tag{20}$$

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

$$z(\alpha) = z_u\cos^2(\alpha - \alpha_0) + z_v\sin^2(\alpha - \alpha_0), \tag{21}$$

  where z_u and z_v are the horizontal and vertical defocus and α_0 is the reference angle defining the azimuthal direction of axial astigmatism. All these parameters are part of the experimental setup.

For more details on the image formation model, we refer to [50] and [21].

Theoretical Recovery Guarantee

The proposed paradigm is supported by Theorem 1 which is also a contribution of this work. Recall from (1) and (2) that $\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}$ is the 2D measurement obtained from a 3D volume $\mathbf{x}$. The operator $\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C}_{\mathbf{c}}\mathbf{S}_{\mathbf{t}}\mathbf{P}_{\theta}$ where $\mathbf{P}_{\theta}$ is the projection operator, $\mathbf{S}_{\mathbf{t}}$ is the shift operator, and $\mathbf{C}_{\mathbf{c}}$ is the convolution operator.

Let $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}$. Then its support is $\text{Support}\{\mathbf{f}\} = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{f}(\mathbf{x}) \neq 0\}$. If $\text{Support}\{\mathbf{f}\} = \mathbb{R}^d$, then $\mathbf{f}$ is said to have a full support.

Theorem 1. *Let $\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}$ as given in (2) with $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$, 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 (respectively, the defocus-major, the defocus-minor and the angle of astigmatism), $\mathbf{x} \in \mathbb{R}^{\mathbf{V}}$ is the vectorized 3D volume, and $\mathbf{y}, \mathbf{n} \in \mathbb{R}^M$ are vectorized 2D images. Let $\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}$, $\mathbf{c} \sim p_{\mathbf{c}}$, $\mathbf{t} \sim p_{\mathbf{t}}$, and $\mathbf{n} \sim p_{\mathbf{n}}$. Then given the assumptions,*

1. *the Fourier transform of the noise distribution $p(\mathbf{n})$ has a full support;*
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 $\mathcal{F}\{\mathbf{C}_{\mathbf{c}_1} + \mathbf{C}_{\mathbf{c}_2}\}$ has a full support;*
3. *the volume $\mathbf{x}$ is non-negative everywhere and has a bounded support; and*
4. *the probability distributions $p_{\boldsymbol{\theta}}$, $p_{\mathbf{c}}$, and $p_{\mathbf{t}}$ are bounded;*

the following holds

$$p(\mathbf{y}|\mathbf{x}_1) = p(\mathbf{y}|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_1 = \mathbf{G}(\mathbf{x}_2), \tag{22}$$

where $\mathbf{G}$ is some member of the set of rotation-reflection operations.

Proof. We first comment on the assumptions. Assumption 1) is true for many common noise distributions including the Gaussian distribution. Assumption 2) is generally true as well. Infact it is used to Wiener filter the clustered projections in classical Cryo-EM reconstruction pipeline. Assumption 3) is true since the volume represents the coulomb potential which is non-negative. Also, the biological structures considered in cryo-EM have finite size.

We denote $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$ with distribution $p_{\text{noiseless}}(\cdot|\mathbf{x})$. We will prove the following in sequence

1. $p(\cdot|\mathbf{x}_1) = p(\cdot|\mathbf{x}_2) \Leftrightarrow p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2)$,
2. $p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_2 = \mathbf{G}(\mathbf{x}_1)$,

For the first part we progress by noting that $\mathbf{y} = \mathbf{y}_{\text{noiseless}} + \mathbf{n}$. Recall that the distribution of the addition of two random variables is equal to the convolution of the distributions of the two random variables. This implies that

$$p(\mathbf{y}|\mathbf{x}) = p_{\text{noiseless}}(\mathbf{y}|\mathbf{x}) * p(\mathbf{n}), \tag{23}$$

$$\mathcal{F}\{p(\cdot|\mathbf{x})\} = \mathcal{F}\{p_{\text{noiseless}}(\cdot|\mathbf{x})\}\mathcal{F}\{p_{\mathbf{n}}(\cdot)\}. \tag{24}$$

By assumption (i), we can now write

$$p_{\text{noiseless}}(\cdot|\mathbf{x}) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{p(\cdot|\mathbf{x})\}}{\mathcal{F}\{p_{\mathbf{n}}(\cdot)\}}\right\}. \tag{25}$$

From this it is easy to see that $p(\cdot|\mathbf{x}_1) = p(\cdot|\mathbf{x}_2) \Leftrightarrow p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2)$. This concludes our first part.

For the second part we will use the result from Theorem 4 (please see "Theoretical Recovery Guarantee in Continuous-Domain"). However, it is based on a continuous-domain volume. But note that $\mathbf{x}$ actually represent a continuous domain volume. Given assumption (iv), the continuous domain representation of $\mathbf{x}$ is

$$f_{\mathbf{x}}(\cdot) = \sum_{i=1}^{V_1}\sum_{j=1}^{V_2}\sum_{k=1}^{V_3} c_{i,j,k}\beta(\cdot - s[i,j,k]^T) \tag{26}$$

where β is a compactly supported basis function, s is the size of a pixel in unit length, and $c_{i,j,k}$ are such that $f_{\mathbf{x}}(s[i,j,k]^T) = \mathbf{x}[i,j,k]$. We assume a β such that $f_{\mathbf{x}}$ is non-negative. The operator $\mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$ is equivalent to applying the continuous-domain forward operator in (14) to $f_{\mathbf{x}}$ and then discretizing and vectorizing the obtained measurement. We can now invoke the result from Theorem 4 which claims that given $f_{\mathbf{x}_1}$ and $f_{\mathbf{x}_2}$, the support of their corresponding $p_{\text{noiseless}}$ is identical if and only if $f_{\mathbf{x}_1}$ and $f_{\mathbf{x}_2}$ are identical upto a rotation-reflection operation. The latter is equivalent to $\mathbf{x}_1 = \mathbf{G}(\mathbf{x}_2)$ where $\mathbf{G}$ is some rotation-reflection operation. This concludes that $p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_1 = \mathbf{G}(\mathbf{x}_2)$.

Theoretical Recovery Guarantee in Continuous-Domain

In the absence of CTF and shifts the recovery guarantee of a function $f : \mathbb{R}^3 \to \mathbb{R}$ from its 2D projections obtained at unknown random poses is given in [51, Theorem 3.1]. We will first go through the notations described in [51] in order to state the required foundational result. We will then extend this theorem for the case when the CTF and shifts are present.

Notations and Preliminaries

Let $SO(3)$ be the space of the special orthogonal matrices and $\mathcal{D}$ be the Borel $\sigma-$algebra induced using the standard Reimannian metric on $SO(3)$. Then $(SO(3), \mathcal{D})$ describes the measurable space of orthogonal matrices. Let $\Delta_N^W = \{\mathbf{x} \in \mathbb{R}^N : \|x\|_2 \leq W\}$ 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 non-negative and are essentially bounded.

For any $f \in \mathbb{F}$ and $\mathbf{A} \in SO(3)$, we denote $\mathbf{y} = \mathcal{P}_{\mathbf{A}}\{f\} = \int_{-\infty}^{\infty} \mathbf{A}f(x, y, z)\,\mathrm{d}z$ 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 $\mathbf{A}$. Infact $\mathbf{A}$ 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

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \Psi_{\mathbf{A}}\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{A}}\{f\} \in \cdot\}. \tag{27}$$

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

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \mathbb{P}_{\text{proj}}(\cdot|\mathbf{R}f),\ \forall f \in \mathbb{F} \text{ and } \mathbf{R} \in O(3), \tag{28}$$

where O_3 is the space of all orthogonal matrices such that $\det \mathbf{A} \in \{-1, 1\}$. The invariance in (28)

is true since

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \Psi\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{A}}\{f\} = \cdot\} \tag{29}$$

$$= \Psi\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{R}^{-1}\mathbf{A}}\{\mathbf{R}f\} = \cdot\} \tag{30}$$

$$= \Psi\{\mathbf{R}\mathbf{A}' \in SO(3) : \mathcal{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\} \tag{31}$$

$$= \Psi\{\mathbf{A}' \in SO(3) : \mathcal{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\} \tag{32}$$

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

$$(\gamma_A f)(\cdot) = f(\mathbf{A}^{-1}\cdot), \forall \mathbf{A} \in O(3), f \in \mathbb{F}. \tag{33}$$

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

Equipped with the notation we can now restate the Theorem 3.1 in [51]. We discuss here the sketch of the proof given in [51].

Theorem 2 ([51, Theorem 3.1])**.** *Let $p_{\mathbf{A}}$ be any bounded distribution on $SO(3)$ and let the assumptions of Theorem 1 be true, then $\forall f, g \in \mathbb{F}$,*

$$[f] \neq [g] \implies \mathbb{P}_{\text{proj}}(\cdot|f) \perp \mathbb{P}_{\text{proj}}(\cdot|g). \tag{34}$$

Proof Sketch. Without loss of generality we provide the proof sketch for the case when $p_{\mathbf{A}}$ is uniform. For the case when $p_{\mathbf{A}}$ is non-uniform the argument remains the same provided that $\Psi_{\mathbf{A}}$ associated with the non-uniform distribution $p_{\mathbf{A}}$ is absolutely continuous w.r.t. Ψ ($\Psi_{\mathbf{A}} \ll \Psi$). This has been stated in [51]. 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 Proposition 7.8 in [52] which states that the following

Proposition 3 ([52, Proposition 7.8])**.** *Let $f \in \mathbb{F}$ and $S_{\mathbf{A}}$ 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.

Coming back to the proof 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 following equality

$$\mathrm{TV}(\mathbb{P}_1, \mathbb{P}_2) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma}[\mathbb{1}_{\mathbf{y}_1 \neq \mathbf{y}_2}] \tag{35}$$

where TV is the total varitaion distance and $\Pi(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all the joint distributions $\gamma(\mathbf{y}_1, \mathbf{y}_2)$ whose marginals are $\mathbb{P}_1$ and $\mathbb{P}_2$, respectively [26]. Infact $\mathbb{E}[\mathbb{1}_{\mathbf{y}_1 \neq \mathbf{y}_2}]$ is equal to the probability of the event $\mathbf{y}_1 \neq \mathbf{y}_2$. In our context this translates into

$$\mathrm{TV}(\mathbb{P}_{\mathrm{proj}}(.|f), \mathbb{P}_{\mathrm{proj}}(.|g)) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_{\mathrm{proj}}(.|f), \mathbb{P}_{\mathrm{proj}}(.|g))} \mathrm{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2) \text{ where } (\mathbf{y}_1, \mathbf{y}_2) \sim \gamma. \tag{36}$$

The optimum is achieved at the extremas which are sparse joint distributions and are such that the variable $\mathbf{y}_2$ is a function of $\mathbf{y}_1$. 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 $(\mathbf{y}_1, \mathbf{y}_2) = (\mathcal{P}_{\mathbf{A}}\{f\}, \mathcal{P}_{h(\mathbf{A})}\{g\})$ for $\mathbf{A} \sim p_{\mathbf{A}}$.
We can then write

$$\Psi[\mathbf{A} \in SO(3) : \mathcal{P}_{h(\mathbf{A})}\{g\} \in \cdot] = \mathbb{P}_{\mathrm{proj}}(\cdot|g). \tag{37}$$

The task now is to estimate $\mathrm{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2)$ where $(\mathbf{y}_1, \mathbf{y}_2) = (\mathcal{P}_{\mathbf{A}}\{f\}, \mathcal{P}_{h(\mathbf{A})}\{g\})$ for $\mathbf{A} \sim p_{\mathbf{A}}$.
(Continuous h). When h is continuous, then with the help of Proposition 3, it is shown that if $[f] \neq [g]$ then,

$$\Psi[\mathbf{A} \in SO(3) : \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{h(\mathbf{A})}\{g\}\|_2 > 0] = 1. \tag{38}$$

(General h). When the function h is not continuous, 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 is then followed by showing that

$$\Psi[\mathbf{A} \in SO(3) : \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{h(\mathbf{A})}\{g\}\|_2 > 0] \geq \Psi(H_\delta), \tag{39}$$

$$\geq 1 - \delta. \tag{40}$$

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 proof shows that 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 $\mathbb{P}_{\mathrm{proj}}(\cdot|f)$ and $\mathbb{P}_{\mathrm{proj}}(\cdot|g)$ is 2. This ensures that the two Probability measures are mutually singular (the intersection of their support has zero measure). This concludes the proof.

Noiseless CTF-modulated Projections

We now extend the previous result for the case when the CTF is present. 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 $\mathcal{C}$, and by $\Psi_{\mathbf{A},\mathbf{c}}$, the measure on this product space. We then define

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|f) = \Psi_{\mathbf{A},\mathbf{c}}[(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathcal{C}) : \mathbf{C}_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} \in \cdot] \tag{41}$$

where $\mathbf{C}_{\mathbf{c}}$ is the space-domain CTF given in (14).

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

$$[f] \neq [g] \implies \mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g). \tag{42}$$

Proof. Similar 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 $p_{\mathbf{A}}$ is uniform. When this is not the case the proof essentially remains the same. We need to show that $\text{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2) = 1$ where $(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma$ for any arbitrary coupling γ of $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$. For arbitrary coupling γ such that $\text{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2)$ is minimum, we again assign $h : (SO(3) \times \mathcal{C}) \to (SO(3) \times \mathcal{C})$ such that

$$(\mathbf{y}_1, \mathbf{y}_2) = (\mathbf{C}_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\}, \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\}) \text{ where } \mathbf{A} \sim p_{\mathbf{A}}, \mathbf{c} \sim p_{\mathbf{c}}, \tag{43}$$

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

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|g) = \Psi_{\mathbf{A},\mathbf{c}}[(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathcal{C}) : \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\} \in \cdot], \tag{44}$$

We will now show that for any h, the event $\{\mathbf{y}_1 \neq \mathbf{y}_2\}$ has probability 1.

(Continuous h). We first assume that h is continuous and use the same kind of technique as in the proof of [51, Theorem 3.1].

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

$$h(\mathbf{A},\mathbf{c}) = (\mathbf{A}\Gamma_{\mathbf{A},\mathbf{c}}, h_1(\mathbf{A},\mathbf{c})). \tag{45}$$

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

$$h_n(\mathbf{A},\mathbf{c}) = (\mathbf{A}\Gamma_n^i, h_{n,1}^i(\mathbf{A},\mathbf{c}))\ \forall\ (\mathbf{A},\mathbf{c}) \in \{\mathcal{A}_n^i \times \mathcal{C}_n^i\}, \tag{46}$$

such that

$$\Gamma_n^i = \underset{\Gamma \in \{\Gamma_{\mathbf{A},\mathbf{c}} : (\mathbf{A},\mathbf{c}) \in \{\bar{\mathcal{A}}_n^i \times \bar{\mathcal{C}}_n^i\}\}}{\arg\min} \ \min_{(\mathbf{A},\mathbf{c}) \in \{\bar{\mathcal{A}}_n^i \times \bar{\mathcal{C}}_n^i\}} \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{\mathbf{A}\Gamma}\{g\}\| \tag{47}$$

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

$$\mathcal{K} = \{(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathcal{C}) : \|\mathbf{C}_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}\Gamma_{\mathbf{A}}}\{g\}\| > 0\}, \tag{48}$$

$$\mathcal{K}_n = \{(\mathbf{A},\mathbf{c}) \in (\mathcal{A}_n^i \times \mathcal{C}_n^i) : \|\mathbf{C}_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma_n^i g\}\| > 0\}. \tag{49}$$

Similar to [51, Theorem 3.1], we can then show that

$$\Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}] = \lim_{n\to\infty} \sum_{i=1}^{i=n} \Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}_n]. \tag{50}$$

We invoke Proposition 5 which gives $\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}_n] = \Psi_{\mathbf{A},\mathbf{c}}[(\mathcal{A}_n^i \times \mathcal{C}_n^i)]$. 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 $\{\mathbf{y}_1 \neq \mathbf{y}_2\}$ has probability 1 if $[f] \neq [g]$.

(General h). When h is not continuous, we can invoke the Lusin's theorem to claim the same (similar to Theorem 2). This means that for any h if $[f] \neq [g]$, the probability of the event $\{\mathbf{y}_1 \neq \mathbf{y}_2\}$ is 1. Therefore the TV distance between $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$ is 2 *i.e.* $\mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g)$. This concludes the proof.

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

$$\mathcal{K}' = \{(\mathbf{A},\mathbf{c}) \in (\mathcal{A}' \times \mathcal{C}') : \|\mathbf{C}_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma g\}\| > 0\}. \tag{51}$$

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

$$\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}'] = \Psi_{\mathbf{A},\mathbf{c}}[(\mathcal{A}' \times \mathcal{C}')]. \tag{52}$$

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 $S_{\mathbf{A}} = [\mathbf{c} \in \mathcal{C}' :$

$\|\mathbf{C_c} * \mathcal{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma g\}\| = 0]$. We define $S_{\mathcal{A}''} = \cup_{\mathbf{A}\in\mathcal{A}''} S_{\mathbf{A}}$ for any $\mathcal{A}'' \subseteq \mathcal{A}'$. We define

$$\mathcal{A}'_1 = [\mathbf{A} \in \mathcal{A}' : S_{\mathbf{A}} \text{ is an uncountable set}], \tag{53}$$

$$\mathcal{A}'_2 = [\mathbf{A} \in \mathcal{A}' : S_{\mathbf{A}} \text{ is a countable non-empty set}]. \tag{54}$$

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

$$\Psi_{\mathbf{A},\mathbf{c}}[\mathcal{K}'^c] = \sum_{i=1}^{2} \Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathcal{A}'_i}(\mathbf{A} \times S_{\mathbf{A}})] \tag{55}$$

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. Now note that, for any $\mathbf{A} \in \mathcal{A}'_1$, the following is true

$$\mathbf{C_c} * I_f = \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * I_g, \ \forall \mathbf{c} \in S_{\mathbf{A}} \tag{56}$$

$$\widehat{\mathbf{C}}_{\mathbf{c}} \cdot \hat{I}_f = \widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})} \cdot \hat{I}_g, \ \forall \mathbf{c} \in S_{\mathbf{A}} \tag{57}$$

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

  We denote by $ze(\hat{I}) = [\omega \in \mathbb{R}^2 : \widehat{I}(\omega) = 0]$, by $\omega_\alpha = \{[r\cos\alpha, r\sin\alpha] : r > 0\}$, and by $ze_\alpha(\hat{I}) = ze(\hat{I}) \cap \omega_\alpha$. From (57) we can write

$$ze(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup ze(\hat{I}_f) = ze(\widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})}) \cup ze(\hat{I}_g), \ \forall \mathbf{c} \in S_{\mathbf{A}}. \tag{58}$$

  Two remarks are in order. Firstly, by assumption (ii) of Theorem 1, $ze(\widehat{\mathbf{C}}_{\mathbf{c}_1}) \cap ze(\widehat{\mathbf{C}}_{\mathbf{c}_2}) = \emptyset$ for $\mathbf{c}_1 \neq \mathbf{c}_2$. Note that $ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}})$ for any $\alpha \in [0, \pi]$ is non-empty (see "Image Formation Theory). Secondly, by assumption (iii) of Theorem 1, support of f and g are compact and nontrivial, so is the support 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

$$ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup ze_\alpha(\hat{I}_f)) = ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap (ze_\alpha(\widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})}) \cup ze_\alpha(\hat{I}_g)), \tag{59}$$

$$ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})})) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g)), \tag{60}$$

$$ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g) \tag{61}$$

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

We can now write

$$\cup_{\mathbf{c} \in S_{\mathbf{A}}} ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = \cup_{\mathbf{c} \in S_{\mathbf{A}}} ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g). \tag{62}$$

for any $\alpha \in S_\alpha$. The set on L.H.S. in (62) has uncountably infinite cardinality since there are uncountably many $\mathbf{c} \in \mathcal{S}_{\mathbf{A}}$ and for each $\mathbf{c}$ there are distinct $ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}})$. Contrary to that the set in the R.H.S. is countable for a given $\alpha \in S_\alpha$. Therefore, for any $\alpha \in S_\alpha$ the two sets have different cardinality. This raises contradiction. The only possible scenario in which (58) is true is when $h_1(\mathbf{A}, \mathbf{c}) = \mathbf{c}$. Using (57) 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 by Proposition 3, $[f] = [g]$.

Now note that,

$$\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A} \in \mathcal{A}_1'} (\mathbf{A} \times S_{\mathbf{A}})] \leq \underbrace{\Psi[\mathcal{A}_1']}_{0} \underbrace{\Psi_{\mathbf{c}}[\cup_{\mathbf{A} \in \mathcal{A}_1'} S_{\mathbf{A}}]}_{\text{finite}} \tag{63}$$

$$= 0. \tag{64}$$

- *(When $S_{\mathbf{A}}$ is countable and non-empty).* Since $S_{\mathbf{A}}$ for this case is a countable set, 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 function $b(z, \cdot)$. Since it is a graph, $\Psi_{\mathbf{A},\mathbf{c}}[q(z)] = 0$.

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

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

Information on Synthetic Data Experiment

Experimental dataset: We construct a synthetic cryo-EM dataset that mimics the real β-galactosidase dataset (EMPIAR-10061) from [16]. We generate 41,000 synthetic β-galactosidase particles using our cryo-EM image-formation model (see *Online Methods*). The ground-truth volume is generated by fitting a 5Å density map on the PDB-*5a1a* atomic model in Chimera [53]. This gives a volume of size $(302 \times 233 \times 163)$ with pixel size of 0.637Å, that is then padded, averaged, and downsampled to size $(180 \times 180 \times 180)$ with pixel size of 1.274Å. This corresponds to a Nyquist resolution of 2.548 Å for the reconstructed volume.

The projections poses are sampled from a uniform distribution over $SO(3)$, where $SO(3)$ is the group of 3D rotations around the origin of $\mathbb{R}^3$. For the CTF, a micrograph from the EMPIAR-10061 dataset is randomly selected and its CTF parameters are extracted using Relion [9]. We then apply the CTF with these parameters to the clean projections. The parameter B of the envelope function of the CTF (see Equation (18)) is chosen such that it decays to a value of 0.2 at the Nyquist frequency. Noisy projections are obtained by adding a randomly-selected background patch from the same micrograph to each noiseless projection. The noise patch is first normalized to zero mean and scaled. The scaling is such that the ratio of the signal energy to the noise energy (SNR) is kept at 0.55, which is equivalent to -2.6 dB.

The dataset is randomly divided into two halves, and the algorithm is applied separately on both halves to generate the half-maps.

Generator settings: We reconstruct a volume of size $180 \times 180 \times 180$ pixels for each half-dataset. The pixel size is 1.274 Å. The volumes are initialized with zeros, and the D2 symmetry of β-galactosidase is enforced during reconstruction.

We use our image-formation model to generate 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 keep the same approach than used to generate the dataset. However, we assume that the final SNR of each projection is unknown, *i.e.*, we learn the scaling parameter that controls the ratio between the projections and the noise patches.

We apply a binary spherical mask of size $(171 \times 171 \times 171)$ on the learned volume. To handle the sharp transition at the mask borders, we enforce some clipping constraints on the masked volume. The clipping value linearly increases with the distance from the center of the projection to the border of the mask, while its minimum value at the center linearly increases from 0 to 10% of the maximum protein value with the number of epochs (*i.e.*, a full pass through each half-dataset). This enforces positivity during the initial phases of reconstruction, which increases the stability of the algorithm.

Discriminator architecture: The architecture of the discriminator network is detailed in the *Online Methods*. The discriminator is initialized identically for both half-datasets. All projections (*i.e.*, the picked particles and the ones generated by the simulator) are normalized to zero-mean and with standard-deviation of 1 before being given to the discriminator.

General settings: The adversarial learning scheme is implemented in Pytorch [17]. For the optimization, we use [29] ($\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$) with a learning rate of 10^{-3} and a batch size of 8. The learning rate decreases by 8% at every epoch. The parameter for the gradient penalty term is kept to $\lambda = 0.001$ (*cf* Equation (10)). The discriminator is trained 4 times for every training of the generator (*i.e.*, $n_{\text{discr}} = 4$ in Algorithm 1.

For the back-propagations, the norm of the gradients for the discriminator are clipped to a maximal value of 10^6. For the generator, the gradients for each pixel are clipped to a maximal value of 10^3. The clipping values linearly increase from zero to those maximas in the first two epochs. Doing so increases the stability of the adversarial learning scheme in the starting, in particular 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 P100 GPU with 18GB memory. Each epoch lasts 10 minutes. The algorithm is run for 16 epochs which, in the current implementation, takes 160 minutes.

Information on Real Data Experiment

Experimental dataset: The dataset consists of 41,123 β-galactosidase (EMPIAR-10061) particle images extracted from 1539 micrographs [16]. Particle images of size 384×384 are downsampled to 192×192, with pixel size of 1.274 Å. This corresponds to a Nyquist resolution of 2.548 Åfor a reconstructed volume of size $180 \times 180 \times 180$. The dataset is randomly divided in two, and the algorithm is applied separately on both halves to generate half-maps. The defocuses and astigmatism parameters of the CTF are estimated from each micrograph using Relion.

Generator settings: For each half-dataset we reconstruct a volume of size $180 \times 180 \times 180$ pixels. Each pixel is of size 1.274 Å. The volumes are initialized with zeros, and the D2 symmetry of β-galactosidase is enforced during reconstruction. A uniform distribution is assumed for the poses.The CTF parameters estimated in Relion are used in the forward model of the CryoEM Physics Simulator. We set the parameter B of the envelope function of the CTF (see (18)) such that it decays to a value of 0.4 at the Nyquist frequency. The translations (vertical and horizontal) are sampled independently from triangle-shaped distributions.

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 normalised. 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 that was used to estimate the CTF parameters previously applied to that specific projection. The scaling operation weights the contribution of the noise w.r.t. the projection signal. This is handled by multiplying the pixel values of the noise images and the projection images by two scalars that are learnt throughout the procedure. These two scalar values are the same for every pair of noise/projection images, *i.e.*, 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 transition at the mask borders, we enforce the same clipping constraints on the masked volume as in the synthetic experiment.

Discriminator architecture: The architecture of the discriminator network is detailed in the *Online Methods*. The discriminator is initialized identically for both half-datasets. The projection images (real and fake) are smoothed with a Gaussian kernel before being given to the discriminator. The width of the kernel is initially set at 2 and decreases by 2% at every epoch.

General settings: The adversarial learning scheme is implemented in Pytorch [17]. For the opti-

mization, we use [29] ($\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$) with a learning rate of 10^{-3} and a batch size of 8. The learning rate decreases by 8% at every epoch. The parameter for the gradient penalty term is kept to $\lambda = 1$ (*cf* Equation (10)). The discriminator is trained 4 times for every training of the generator (*i.e.*, $n_{\text{discr}} = 4$ in Algorithm 1.

For this dataset, the algorithm is first run for 8 epochs (with translation search switch off) to produce a stable low-resolution reconstruction (15Å). The process is then restarted using this volume, and run for 12 epochs to obtained a high-resolution volume. In this second stage, we limit the refinement to the higher frequencies components above 15Å.

For the back-propagations, the norm of the gradients for the discriminator are clipped to a maximal value of 10^6. For the generator, the gradients for each pixel are clipped to a maximal value of 10^3. The clipping values linearly increase from zero to those maxima in the first two epochs. Doing so increases the stability of the adversarial learning scheme in the starting, in particular that of the discriminator. The gradients corresponding 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 P100 GPU with 18GB memory. Each epoch (*i.e.*, a full pass through each half-dataset) lasts 10 minutes. The algorithm is run for 200 minutes.