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

Harshit Gupta, Michael T. McCann, *Member, IEEE*, Laurène Donati, and Michael Unser, *Fellow, IEEE*

***Abstract*—We present CryoGAN, a new paradigm for single-particle cryo-electron microscopy (cryo-EM) reconstruction based on unsupervised deep adversarial learning. In single-particle cryo-EM, the structure of a biomolecule needs to be reconstructed from a large set of noisy tomographic projections with unknown orientations. Current reconstruction techniques are based on a marginalized maximum-likelihood formulation that requires calculations over the set of all possible poses for each projection image, a computationally demanding procedure. Our approach is to seek a 3D structure that has simulated projections that match the real data in a distributional sense, thereby sidestepping pose estimation or marginalization. We prove that, in an idealized mathematical model of cryo-EM, this approach results in recovery of the correct structure. Motivated by distribution matching, we propose CryoGAN, a specialized GAN that consists of a 3D structure, a cryo-EM physics simulator, and a discriminator neural network. During reconstruction, the 3D structure is optimized so that its projections obtained through the simulator resemble real data (to the discriminator). Simultaneously, the discriminator is trained to distinguish real projections from simulated projections. CryoGAN takes as input only real projection images and the distribution of the cryo-EM imaging parameters. It involves neither prior training nor an initial estimation of the 3D structure. CryoGAN currently achieves a 10.8 Å resolution on a realistic synthetic dataset. Preliminary results on experimental β-galactosidase and 80S ribosome data demonstrate the ability of CryoGAN to exploit data statistics under standard experimental imaging conditions. We believe that this paradigm opens the door to a family of novel likelihood-free algorithms for cryo-EM reconstruction.**

***Index Terms*—Single-particle cryo-electron microscopy (cryo-EM), tomographic reconstruction, deep learning, generative adversarial networks, cryo-EM physics simulator.**

Manuscript received November 9, 2020; revised April 1, 2021 and June 21, 2021; accepted June 21, 2021. Date of publication July 13, 2021; date of current version August 6, 2021. This work was supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, under Grant Agreement 692726 GlobalBioIm: Global integrative framework for computational bio-imaging. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tobias Lasser. *(Corresponding author: Harshit Gupta.)*

Harshit Gupta, Laurène Donati, and Michael Unser are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne, Switzerland (e-mail: harshit.gupta@epfl.ch; laurene.donati@epfl.ch; michael.unser@epfl.ch).

Michael T. McCann is with the Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI USA (e-mail: mccann13@msu.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TCI.2021.3096491, provided by the authors.

Digital Object Identifier 10.1109/TCI.2021.3096491

I. Introduction

SINGLE-PARTICLE cryo-electron microscopy (cryo-EM) is a powerful method to determine the atomic structure of macro-molecules by imaging them with electron rays at cryogenic temperatures [1]–[3]. Its popularity has rocketed in recent years, culminating in 2017 with the Nobel Prizes awarded to Jacques Dubochet, Richard Henderson, and Joachim Frank. In single-particle cryo-EM, one collects 2D noisy projections from numerous copies (dubbed "particles") of a 3D biomolecule of interest, where each copy is randomly oriented. (In practice, these particles may not be exact duplicates due to impurities in sample preparation and different conformations of the target biomolecule; this work assumes that these can be removed during preprocessing.) There exists a multitude of software packages to produce high-resolution 3D structure(s) from these 2D measurements [4]–[12]. These packages are based on sophisticated algorithms, which include projection-matching approaches [13], [14], frequency marching [15], common-lines [16], maximum-likelihood optimization frameworks, and a range of other methods [17]–[22], enabling the determination of structures with unprecedented atomic resolution.

Reconstruction in single-particle cryo-EM is a challenging task. A key reason behind this difficulty is that the imaged particles have unknown poses. Hence, the tomographic reconstruction task involves a high-dimensional, nonconvex optimization problem with numerous local minima. Currently, most software packages rely on a marginalized maximum-likelihood (ML) formulation [23] that is solved through an expectation-maximization algorithm [9], [11]. The latter involves calculations over the discretized space of poses for each projection, a computationally demanding procedure. Moreover, the outcome of the global process is predicated on the quality of the initial reconstruction [24], [25]. Additionally, one often relies on the input of an expert user for appropriate processing decisions and parameter tuning [26]. Even for more automated methods, the risk of outputting incorrect and misleading 3D reconstructions is ever-present.

To bypass these limitations, we introduce CryoGAN, an unsupervised reconstruction algorithm for single-particle cryo-EM that exploits the remarkable ability of generative adversarial networks (GANs) to model data distributions [27]. CryoGAN is driven by the competitive training of two entities: one that tries to capture the distribution of real data, and another that

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

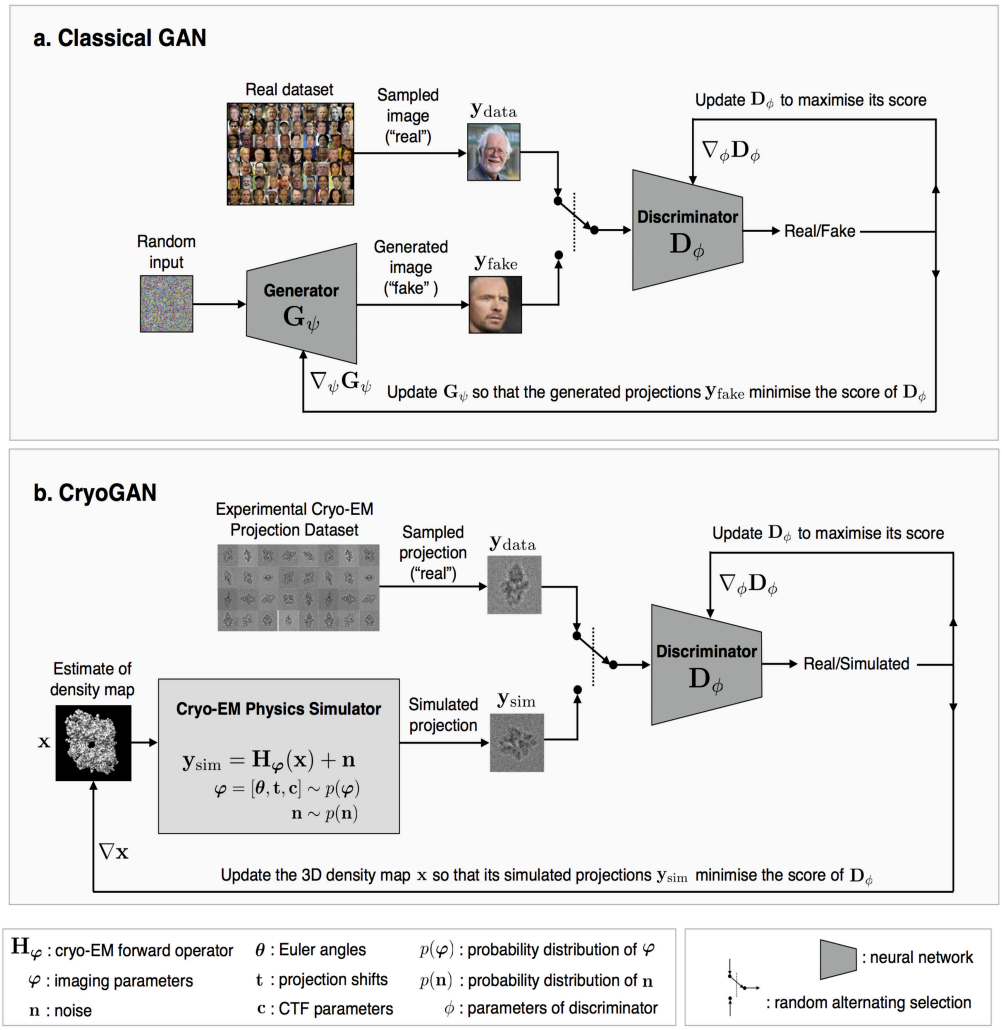

Fig. 1. Schematic comparison between (a) a standard GAN architecture and (b) the CryoGAN architecture. Both frameworks rely on a deep adversarial learning scheme to model the distribution of the real data. CryoGAN looks for the volume whose simulated projections have a distribution that matches the real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

discriminates between generated samples and samples from the real dataset. In a standard GAN, the two entities are each a convolutional neural network (CNN). They are known as the generator and the discriminator and are trained simultaneously using backpropagation. The important twist with CryoGAN is that we replace the generator network by a cryo-EM physics simulator. By doing so, CryoGAN learns the 3D density map whose simulated projections are the most consistent with a given dataset of 2D measurements in a distributional sense. We summarise in Figure 1 the fundamental difference between a typical GAN and CryoGAN.

A. *Contributions and Outline*

The main contributions of this paper are (1) a mathematical framework that deploys distribution matching for cryo-EM reconstruction, including a proof that builds on [28] and that shows that this approach can recover the true structure in an idealized setting; and (2) CryoGAN, an algorithm for cryo-EM reconstruction motivated by this framework, along with experimental validation on simulated and experimental cryo-EM data.

CryoGAN has several attractive qualities: No estimation of the poses is attempted during the learning procedure; rather, the reconstruction is obtained through distributional matching performed in a likelihood-free manner. Hence, CryoGAN sidesteps many of the computational drawbacks associated with likelihood-based methods. The CryoGAN algorithm requires no prior knowledge of the 3D structure; its learning process is purely unsupervised. The user needs only to feed the particle images and estimates of parameters of the contrast transfer function (CTF). No initial volume is needed: the algorithm may start with a volume initialized with zeros. Unlike most deep-learning methods that solve inverse problems in imaging [29], [30], our method requires no training dataset nor prior training.

The rest of the manuscript is organized as follows: We describe related works on cryo-EM reconstruction in Section II. In Section III, we state the cryo-EM reconstruction problem and formulate our forward imaging model. In Section IV, we present our distribution-matching approach and its mathematical justification, including our main theorem. In Section V, we present our proposed algorithm, CryoGAN. In Section VI we present our experiments and results on synthetic and experimental datasets. We end in Section VII with a discussion.

II. Related Works

Related works fall into two main categories: the reconstruction methods used in most standard cryo-EM software packages and recently proposed deep-learning techniques.

Current Cryo-EM Reconstruction. The main challenge in cryo-EM reconstruction is that every particle has an unknown pose when imaged— if the poses were known, ML or maximum *a posteriori* (MAP) estimation of the volume could be performed by solving a standard linear inverse problem, with the large number of measurements counteracting their low SNR.

The classical paradigm to overcome this problem is to refine an initial volume by iteratively estimating the unknown poses. Pose estimation can be achieved with a variety of strategies, including the popular projection-matching approach [13], [14]. Whatever the method used, pose estimation is challenging because the SNR of individual projection images is extremely low. It also requires the estimation of additional parameters and, at every iteration of the reconstruction pipeline, the generation of the projections of the intermediate reconstruction at a large number of poses; this is extremely costly from a computational point of view.

Another approach is to formulate the reconstruction as an ML estimation problem in which the unknown poses are marginalized away [9], [11], [31], [32]. This is attractive in that no extra parameters need to be estimated. The problem can then be solved using the expectation-maximization algorithm (*e.g.*, [9], [11]), where marginalization over poses during the so-called E-step is computationally expensive. This does however still require a first estimate of the 3D volume.

Alternatively, the ML problem can be minimized *ab initio*, without an initial volume, using stochastic gradient descent (SGD) (*e.g.*, during the first phase of [11]). Here, the challenge is that the involved gradients require computations over all poses. For a more in-depth discussion, see [23], [26], [33]. For additional mathematical details on the relationship between likelihood-based methods and CryoGAN, see Supplementary Materials Section B.

Likelihood-free methods for single-particle cryo-EM reconstruction are relatively few. An early approach is [34], which proposes to reconstruct an *ab initio* structure such that the first few moments of the distribution of its simulated cryo-EM measurements match the ones of the real dataset. However, the method assumes that the poses of the particles have a uniform distribution, which can be restrictive. This moment-matching technique has been recently extended in [35] to reconstruct an *ab initio* structure in the case of nonuniform pose distributions. The authors show that, under appropriate conditions, matching the first two analytical moments is sufficient to uniquely determine the 3D structure. Although our proposed method and moment-matching methods are conceptually similar, the assumptions, practical considerations, and actual reconstruction algorithms are very different.

Recently, the work in [36] uses the Wasserstein distance in a fashion similar to ours. However, the distance there is only used to cluster the tomographic projections and not to perform reconstruction.

Deep Learning for Cryo-EM. Deep learning has already had a profound impact on a wide range of image-reconstruction applications [29], [37], [38]; yet, its current utilization in single-particle cryo-EM is mostly restricted to preprocessing steps such as micrograph denoising [39] or particle picking [40]–[44]. Very recently, deep learning has also been used to solve the reconstruction problem in cryo-EM. The work in [45] uses neural networks to model the continuous generative factors of structural heterogeneity. Another work [46] uses a variational autoencoder trained using a discriminator-based objective to find a low-dimensional latent representation of the particles. These representations are then used to estimate the poses.

From a more general standpoint, if deep learning is now extensively used to solve inverse problems in imaging [29], [47]–[49], most methods are based on supervised learning and thus rely on training data. Interestingly, the reconstruction of a 3D structure from its 2D views is a prominent problem in computer vision [50]. Many recent deep-learning algorithms have been used in this regard [51], [52]. While this problem is ostensibly similar to that of single-particle cryo-EM reconstruction, its measurement process is distinct and the level of noise much lower. A recent work [53] includes simple forward models within a GAN; however the aim of the method is image generation (like a standard GAN) and not image reconstruction, which also restricts the comparison.

III. Single-Particle Cryo-EM

We begin with a mathematical statement of the problem that CryoGAN addresses, namely, single-particle cryo-EM image reconstruction. We then give our measurement model for cryo-EM, including a stochastic formulation.

A. Problem Statement

The goal of single-particle cryo-EM reconstruction is to estimate an unknown 3D density map from a large collection of its noisy, 2D projection images. Following a standard cryo-EM model [54], we write the linear relationship between the volume and each projection as

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

where $\mathbf{y}^i \in \mathbb{R}^M$ is a (vectorized) 2D projection, $\mathbf{x} \in \mathbb{R}^V$ is the (vectorized) 3D density, $\mathbf{H}_{\boldsymbol{\varphi}} \in \mathbb{R}^{M\times V}$ is the forward operator with parameters $\varphi^i \in \mathbb{R}^8$, and $\mathbf{n} \in \mathbb{R}^M$ is an additive noise following distribution $p_{\mathbf{n}}$. Details of $\mathbf{H}_{\varphi}$ are given in Section III-B.

The problem of cryo-EM image reconstruction is to estimate $\mathbf{x}$ from a set $\{\mathbf{y}^1, \ldots, \mathbf{y}^{N_{\text{tot}}}\}$ given by (1). The key challenges

that separate cryo-EM from other tomography problems are the fact that the parameters $\boldsymbol{\varphi}^i$ of the forward operator are different for each projection and *unknown* [23], the large number of projections (N_{tot} is typically between 10^4 and 10^7), and the high level of noise in the projections.

B. Image-Formation Model

Our cryo-EM forward operator $\mathbf{H}_\varphi$ is based on [54]–[56], It is given by

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

where $\mathbf{P}_{\boldsymbol{\theta}} : \mathbb{R}^V \to \mathbb{R}^M$ is a projection operator at orientation $\boldsymbol{\theta}$ (mathematically speaking, an X-ray transform [57]), $\mathbf{S}_{\mathbf{t}} : \mathbb{R}^M \to \mathbb{R}^M$ is a shift operator, and $\mathbf{C}_{\mathbf{c}} : \mathbb{R}^M \to \mathbb{R}^M$ is a convolution operator. Thus, the imaging parameters, $\boldsymbol{\varphi}$, comprise the projection (Euler) angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, the projection shifts $\mathbf{t} = (t_1, t_2)$, and the CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\mathrm{ast}})$, where d_1 is the defocus-major, d_2 is the defocus-minor, and α_{ast} is the angle of astigmatism.

The projection operator $\mathbf{P}_{\boldsymbol{\theta}}$ is a discretization of the continuous-domain operator [57]

$$P_{\boldsymbol{\theta}}\{f\}(x_1, x_2) = \int_{-\infty}^{\infty} R_{\boldsymbol{\theta}}\{f\}(x_1, x_2, x_3)\,\mathrm{d}x_3, \tag{3}$$

where $f : \mathbb{R}^3 \to \mathbb{R}$ is a density map and $R_{\boldsymbol{\theta}}\{f\}(\mathbf{x}) = f(\mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{x})$ is a 3D rotation of f, where $\mathbf{R}_{\boldsymbol{\theta}}$ is the rotation matrix associated with the Euler angles $\boldsymbol{\theta}$. A plethora of software packages for computing discretized versions of (3) exist; we use the ASTRA toolbox [58].

The shift operator $\mathbf{S}_{\mathbf{t}}$ models the fact that the projection measurements are cropped subimages of a larger micrograph and can thus be off-center. In the continuous domain,

$$S_{\mathbf{t}}\{y\}(x_1, x_2) = y(x_1 - t_1, x_2 - t_2), \tag{4}$$

where $y : \mathbb{R}^2 \to \mathbb{R}$ is a projection and $\mathbf{t}$ gives the 2D offset. We implement the discretized version of $S_{\mathbf{t}}$ by using whole-pixel shifts.

The convolution operator $\mathbf{C}_{\mathbf{c}}$ models electron optics [55], [56]. It is a discretization of the continuous convolution operator $C_{\mathbf{c}}$, the frequency response of which is given by the CTF. Detailed expressions for the CTF can be found Supplementary Materials Section A.

C. Stochastic Model of Cryo-EM

The noisy projections acquired from a given volume $\mathbf{x}$ depend on the imaging parameters $\boldsymbol{\varphi}$, which can be assumed to be sampled from a probability distribution $p_{\boldsymbol{\varphi}}$. Along with the probability distribution of the noise, it induces a distribution onto the acquired projections. This means that those can be assumed to be sampled according to

$$p(\mathbf{y}|\mathbf{x}) = \int_{\boldsymbol{\varphi}\in\Phi} p_{\mathbf{n}}(\mathbf{y} - \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}) p_{\boldsymbol{\varphi}}(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi}, \tag{5}$$

where Φ is the set of all imaging parameters. We denote a noiseless projection as $\mathbf{y}_{\mathrm{clean}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$.

Correspondingly, we denote by $\mathbf{x}_{\mathrm{true}}$ the 3D structure from which the projection dataset has been acquired. Therefore, the projections in the real dataset can be seen as samples of the distribution $p(\mathbf{y}|\mathbf{x}_{\mathrm{true}})$ which we denote by p_{data}.

IV. The Distribution-Matching Approach

The CryoGAN algorithm is motivated by a distribution-matching approach, the goal being to find a reconstruction $\mathbf{x}_{\mathrm{rec}}$ such that its statistical projection distribution matches the data distribution $p(\mathbf{y}|\mathbf{x}_{\mathrm{true}})$. (In practice, we have access only to a large number of samples from this distribution. Please see Section VII-A.) This means that we reconstruct a 3D structure whose set of projections resemble the acquired projection dataset. Intuitively, this suggests that the 3D structure itself should also resemble the 3D structure. We prove in Theorem 1 that this intuition is indeed correct: in an idealized setting, distribution matching recovers the true structure.

Two remarks are in order. Firstly, distribution matching for cryo-EM reconstruction was first considered in [28]. However, it involved a restrictive setting where the forward model lacked noise and optical effects. While our proof for Theorem 1 is based on the mathematical tools used in [28], our result generalizes it to a more realistic forward model. Secondly, current ML methods [9], [11] can be interpreted as distribution-matching approaches. Thus, Theorem 1 provides a justification for them as well. To our knowledge, it is the first theoretical result that states that the ML formulation is apt to recover the correct structure. For more details on this connection, see Supplementary Materials Section B.

In the continuous domain, we have that $y = H_{\boldsymbol{\varphi}} f + n$ where $y : \mathbb{R}^2 \to \mathbb{R}$ is the 2D measurement obtained from the 3D volume f and n is the noise. Here, $H_{\boldsymbol{\varphi}} = C_{\mathbf{c}} S_{\mathbf{t}} P_{\boldsymbol{\theta}}$ is the continuous-domain forward operator, $P_{\boldsymbol{\theta}}$ is the projection operator, $S_{\mathbf{t}}$ is the shift operator, and $C_{\mathbf{c}}$ is the operator for convolution with the CTF. We assume that $\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}$, $\mathbf{c} \sim p_{\mathbf{c}}$, $\mathbf{t} \sim p_{\mathbf{t}}$, and that $\mathbb{P}_n$ is the probability measure associated with n. We denote the probability measure on the projection space given the 3D structure f as $\mathbb{P}(y|f)$.

We have the following assumptions on the forward model:

1) the characteristic functional [59, Chapter 2] $\hat{\mathbb{P}}_n$ of the noise probability measure $\mathbb{P}_n$ is nonzero everywhere over its domain and n is pointwise-defined everywhere in $\mathbb{R}^2$;
2) the support of $p_{\mathbf{c}}$ is such that, for any $\mathbf{c}_1, \mathbf{c}_2 \in \mathrm{Support}\{p_{\mathbf{c}}\}$ and $\mathbf{c}_1 \neq \mathbf{c}_2$, $|\mathcal{F}\{C_{\mathbf{c}_1}\}| + |\mathcal{F}\{C_{\mathbf{c}_2}\}|$ is positive everywhere;
3) the volume f is nonnegative everywhere and has a bounded support;
4) the probability distributions $p_{\boldsymbol{\theta}}$, $p_{\mathbf{c}}$, and $p_{\mathbf{t}}$ are bounded.

Before stating Theorem 1, we first comment on these assumptions. Assumption 1) holds for many noise distributions used in cryo-EM models, including white Gaussian noise filtered with any kernel having compact support. Assumption 2) is commonly used, for instance, to justify the application of Wiener filter to the clustered projections in classical cryo-EM reconstruction pipelines [3]. Assumption 3) is reasonable in cryo-EM because

the volume represents a 3D electron density map, which is nonnegative and effectively has a finite size.

Theorem 1: Let $\mathbb{P}(y|f_{\rm true})$ be the probability measure associated with the projection dataset obtained from the true structure $f_{\rm true}$ such that

$$y = H_{\boldsymbol{\varphi}} f_{\rm true} + n, \text{ where } \boldsymbol{\varphi} \sim p_{\boldsymbol{\varphi}}^{\rm true} \text{ and } n \sim \mathbb{P}_n^{\rm true}. \quad (6)$$

Let $\mathbb{P}^{\rm rec}(y|f_{\rm rec})$ be the probability measure associated with the projections obtained from the reconstructed structure $f_{\rm rec}$ with imaging distribution $p_{\boldsymbol{\varphi}}^{\rm rec}$ and noise measure $\mathbb{P}_n^{\rm rec}$. We assume that these distributions and structures satisfy the assumptions of the forward model. Then, given $p_{\boldsymbol{\varphi}}^{\rm true} = p_{\boldsymbol{\varphi}}^{\rm rec}$ and $\mathbb{P}_n^{\rm true} = \mathbb{P}_n^{\rm rec}$, it holds that

$$\mathbb{P}(y|f_{\rm true}) = \mathbb{P}(y|f_{\rm rec}) \Leftrightarrow f_{\rm rec} = G(f_{\rm true}), \quad (7)$$

where G is a rotation-reflection operation.

Theorem 1 states that if the reconstructed and true structures have the same projection distributions then the structures are the same except for some rotation-reflection operation. This result supports distribution matching as an approach for cryo-EM reconstruction. Note that this result assumes that the distribution of the imaging parameters and the probability measure of the noise for the reconstructed projections exactly match that of the true projections.

Discrete-Domain Extension. In practice, the cryo-EM measurements are acquired on a detector grid and are therefore discrete. Theorem 1 then holds approximately, up to some error that results from discretization of the measurements, forward model, and 3D density map.

V. CryoGAN

Theorem 1 provides motivation for our algorithm, CryoGAN, which aims to match the statistical distribution of the acquired and simulated data. Based on adversarial learning, CryoGAN, attempts to reconstruct a voxel-domain $\mathbf{x}_{\rm rec}$ such that $p(\mathbf{y}|\mathbf{x}_{\rm rec}) = p(\mathbf{y}|\mathbf{x}_{\rm GT})$, meaning that the corresponding probability measures are equal as well. Barring the approximation error stemming from discretization, Theorem 1 ensures that if CryoGAN successfully finds such a structure, then it is indeed a rotated-reflected version of the true structure.

The main goal now is to tractably find such a volume. In order to do so, we formulate the reconstruction task as the minimization problem

$$\mathbf{x}_{\rm rec} = \arg\min_{\mathbf{x}} D\left(p(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\mathbf{x}_{\rm true})\right), \quad (8)$$

where D is a valid distance between the two distributions. In essence, (8) states that the appropriate reconstruction is the 3D density map whose projection distribution is the most similar to the real dataset in a distributional sense. For the sake of conciseness, we shall henceforth use the notation $p(\mathbf{y}|\mathbf{x}) = p_{\mathbf{x}}(\mathbf{y})$.

As the distance in (8), we choose the Wasserstein distance defined as

$$D(p_1, p_2) = \inf_{\gamma \in \Pi(p_1, p_2)} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma}[\|\mathbf{y}_1 - \mathbf{y}_2\|], \quad (9)$$

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. Our choice is driven by works demonstrating that the Wasserstein distance is more amenable to minimization than other popular distances (*e.g.*, total-variation or Kullback-Leibler divergence) for distribution-matching applications [60]. Using (9), the minimization problem (8) expands as

$$\mathbf{x}_{\rm rec} = \arg\min_{\mathbf{x}} \inf_{\gamma \in \Pi(p_{\mathbf{x}}, p_{\rm data})} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma}[\|\mathbf{y}_1 - \mathbf{y}_2\|]. \quad (10)$$

By using the formalism of [60]–[62], this minimization problem can also be stated in its dual form

$$\mathbf{x}_{\rm rec} = \arg\min_{\mathbf{x}} \max_{f:\|s\|_L \leq 1} \left(\mathbb{E}_{\mathbf{y} \sim p_{\rm data}}[s(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}}[s(\mathbf{y})]\right), \quad (11)$$

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

A. Connection With Wasserstein GANs

The scheme in (10) falls under the framework of Wasserstein GANs (WGANs) [60]. In the standard WGAN implementation, the function s is parameterized by a neural network $\mathbf{D}_\phi$, called the discriminator, with parameters ϕ. The role of the discriminator is to learn to differentiate between real samples (typically coming from an experimental dataset) and fake samples. The latter are produced by another neural network, called the generator, which aims at producing samples that are realistic enough to fool the discriminator. This adversarial-learning scheme progressively drives the WGAN to capture the distribution of the experimental data.

In CryoGAN, we adapt this adversarial scheme to learn the volume $\mathbf{x}$ whose simulated projections follow the distribution of the data. To do so, we rely on a cryo-EM physics simulator whose role is to produce projections of a volume estimate $\mathbf{x}$ using (1). These simulated projections then follow a distribution $\mathbf{y} \sim p_{\mathbf{x}}$. Hence, (11) translates into

$$\mathbf{x}_{\rm rec} = \arg\min_{\mathbf{x}} \max_{\|\mathbf{D}_\phi\|_L \leq 1} \left(\mathbb{E}_{\mathbf{y} \sim p_{\rm data}}[\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}}[\mathbf{D}_\phi(\mathbf{y})]\right). \quad (12)$$

As proposed in [63], the Lipschitz constraint $\|\mathbf{D}_\phi\|_L \leq 1$ can be enforced by penalizing the norm of the gradient of $\mathbf{D}_\phi$ with respect to its input. This gives the final formulation of our reconstruction problem as

$$\begin{aligned}\mathbf{x}_{\rm rec} = \arg\min_{\mathbf{x}} \max_{\mathbf{D}_\phi} \big(&\mathbb{E}_{\mathbf{y} \sim p_{\rm data}}[\mathbf{D}_\phi(\mathbf{y})] - \mathbb{E}_{\mathbf{y} \sim p_{\mathbf{x}}}[\mathbf{D}_\phi(\mathbf{y})] \\ &- \lambda \cdot \mathbb{E}_{\mathbf{y} \sim p_{\rm int}}[(\|\nabla_{\mathbf{y}} \mathbf{D}_\phi(\mathbf{y})\| - 1)^2]\big).\end{aligned} \quad (13)$$

Here, $p_{\rm int}$ denotes the uniform distribution along the straight line between points sampled from $p_{\rm data}$ and $p_{\mathbf{x}}$, while $\lambda \in \mathbb{R}_+$ is an appropriate penalty coefficient (see [63], Section 4).

B. The CryoGAN Algorithm

Equation (13) is a min-max optimization problem. By replacing the expected values with their empirical counterparts

Algorithm 1: CryoGAN.

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

for n_{train} **do**
  for n_{discr} **do**
    sample $\{\mathbf{y}^1_{\text{batch}}, \ldots, \mathbf{y}^M_{\text{batch}}\}$ from real data
    sample $\{\mathbf{y}^1_{\text{sim}}, \ldots, \mathbf{y}^M_{\text{sim}}\}$ from current $p_{\mathbf{x}}$ ▷ (Algo. 2)
    sample $\{\alpha_1, \ldots, \alpha_M\} \sim U[0,1]$
    compute $\mathbf{y}^m_{\text{int}} = \alpha_m \mathbf{y}^m_{\text{batch}} + (1-\alpha_m)\mathbf{y}^m_{\text{sim}}$
    update discriminator parameters ϕ using $\nabla_\phi L_S$ (14)
  sample $\{\mathbf{y}^1_{\text{sim}}, \ldots, \mathbf{y}_{\text{sim}}{}^M\}$ from current $p_{\mathbf{x}}$ ▷ (Algo. 2)
  update volume $\mathbf{x}$ using $\nabla_{\mathbf{x}} L_S$ (14)

(sums) [63], we reformulate it as the minimization of

$$L_S(\mathbf{x}, \mathbf{D}_\phi) = \sum_{m=1}^{M} \mathbf{D}_\phi(\mathbf{y}^{n_m}_{\text{data}}) - \sum_{m=1}^{M} \mathbf{D}_\phi(\mathbf{y}^m_{\text{sim}}) - \lambda \sum_{m=1}^{M} (\|\nabla_{\mathbf{y}} \mathbf{D}_\phi(\mathbf{y}^m_{\text{int}})\| - 1)^2), \tag{14}$$

where M is the number of samples in the empirical estimates, $\{n_m\}_{m=1}^M$ is a set of random indices into the dataset of measured projection images, $\{\mathbf{y}^m_{\text{sim}}\}_{m=1}^M$ is a set of projections from the current estimate $\mathbf{x}$ generated by the cryo-EM physics simulator, and $\mathbf{y}^m_{\text{int}} = \alpha_m \mathbf{y}^{n_m} + (1-\alpha_m)\mathbf{y}^m_{\text{sim}}$, where α_m is sampled from a uniform distribution between 0 and 1.

The empirical estimate (14) is identical to (13) only in the limit when the number of measurements grows to infinite. Therefore, the success of the algorithm in real-world scenarios hinges on the fidelity of this empirical estimate, which presupposes that the number of particles is very large. More details on the difference between the empirical estimate (14), Wasserstein-distance based formulation (13), and the original distribution-matching formulation (8) are given in Section VII-A.

In practice, we minimize (14) through SGD using batches; specifically, we use the Adam optimizer [64]. We alternatively update the discriminator $\mathbf{D}_\phi$ (for n_{discr} iterations) and the volume $\mathbf{x}$ (for one iteration). The pseudocode and a schematic view of the CryoGAN algorithm are given in Algorithm 1 and Figure 1 b, respectively. The code for our implementation of CryoGAN is written in Python using the PyTorch [65] package.[1] We now provide more detail on the CryoGAN physics simulator and discriminator network.

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

[1] Code for the CryoGAN algorithm is [Online]. Available: https://github.com/harshit-gupta-cor/CryoGAN

Algorithm 2: Pseudocode for Cryo-EM Physics Simulator.

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

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 $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}}) \sim p_{\mathbf{c}}$
generate a synthetic noiseless projection based on (2), with $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$
sample the noise $\mathbf{n} \sim p_{\mathbf{n}}$. Add to the projection as $\mathbf{y}_{\text{sim}} = \mathbf{y}_{\text{noiseless}} + \mathbf{n}$.

We now detail the steps of the generator. Recall that the set of imaging parameters is given by $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$. We first sample the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ from a distribution $p_{\boldsymbol{\theta}}$ decided *a priori* based on the acquired dataset. Similarly, the projection shifts $\mathbf{t} = (t_1, t_2)$ are sampled from the prior distribution $p_{\mathbf{t}}$. The CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$ are sampled from the prior distribution $p_{\mathbf{c}}$. In practice, we exploit the fact that the CTF parameters can often be efficiently estimated for all micrographs. We then uniformly sample from the whole set of extracted CTF parameters.

We generate noiseless projections $\mathbf{y}_{\text{noiseless}}$ by applying $\mathbf{H}_{\boldsymbol{\varphi}}$ to the current volume estimate $\mathbf{x}$. The projection operator $\mathbf{P}_{\boldsymbol{\theta}}$ in (2) is implemented using the ASTRA toolbox [58]. We propagate gradients back through the projection operation to the volume $\mathbf{x}$ by using the ASTRA backprojection algorithm.

The simulator must then add realistic noise to the clean projections so that the distribution of the simulated projections may closely match that of the real data, which is highly noisy (Figure 5). To produce realistic noise realizations, we extract random background patches directly from the micrographs themselves, at locations where particles do not appear, and add them pixel-wise to the clean projection. For consistency, the noise patch added to a given noiseless projection is taken from the same micrograph that was used to estimate the CTF parameters previously applied to that specific projection. We detail this approach when describing our experiments in Section VI.

In summary, the physics simulator in CryoGAN differs for the generator network in a standard GAN in several ways: the learnable parameters are the voxel values of the reconstruction rather than weights and biases; the architecture comes from a physics model rather than the typical convolutional layers, and the randomness comes from the random selection of the imaging parameters and noise patch, rather than from a multivariate Gaussian input.

Discriminator Network. The cryo-EM physics simulator is paired with a discriminator network whose architecture is similar to those used in standard GANs. Its role is to differentiate between projections from the experimental dataset and projections

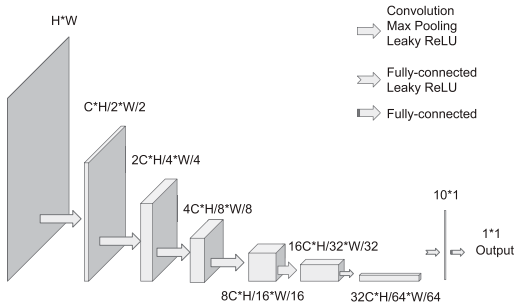

Fig. 2. Architecture of the discriminator. The parameter for the channel size is $C = 96$ in every experiment. The input image with size $(H \times W)$ is successively processed and downsampled to output a scalar.

generated by the cryo-EM physics simulator. The discriminator is not pretrained; its weights are optimized via the adversarial learning scheme described in Section V during reconstruction. The gradients from the discriminator carry information on the difference between the real and simulated projections at a given runtime. Those gradients are used by the cryo-EM physics simulator to update itself, thus improving on the realism of the simulated projections.

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

VI. Experiments and Results

We now describe our experiments and results on synthetic and experimental datasets.

For each experiment, we randomly divided the dataset into halves and performed separate reconstructions. We then aligned the two reconstructed structures and calculated the FSC between them using the FOCUS software [66], [67]. We report the half-half resolution where this FSC crosses the 0.143 cutoff. Additionally, we aligned and averaged these half-maps and report the resolution (with FSC cutoff 0.5 and 0.143) with respect to the true structure. For the experiments associated with β-galactosidase (synthetic and experimental data), we considered the density map obtained from the PDB-*5a1a* atomic model as the true structure (see Section VI-A). Similarly, for the 80S Ribosome case, we used the preprocessed publicly available EMDB-2660 as the true structure (see Section VI-D). These true structures were used to create the datasets in every synthetic experiment.

TABLE I
Reconstruction Resolution (Å) for Synthetic β-Galactosidase

	SNR (dB), translation (%)			
metric	-20, 0	-5.2, 0	-20, 3	-20, 20
half-map FSC = 0.143	8.6	7.5	10.8	14.3
truth FSC = 0.5	15.3	8.3	14.7	23.2
truth FSC = 0.143	11.7	6.5	11.5	19.8

A. Synthetic Experiment

We first assess the performance of CryoGAN on a synthetic dataset that mimics the β-galactosidase dataset (EMPIAR-10061) from [68] in terms of noise level and CTF parameters. In the context of this experiment, we refer to these synthetic data as "real," in contrast to the projections coming from the CryoGAN physics simulator, which we term "simulated."

Experimental Setup. For the ground-truth volume, we generated a 2.5 Å density map from PDB-*5a1a* atomic model using the Chimera software [69]. This gave us a volume of size $(302 \times 233 \times 163)$ with voxel size 0.637 Å. The volume was then padded, averaged, and downsampled to a size $(180 \times 180 \times 180)$ with voxel size 1.274 Å. This corresponds to a Nyquist resolution of 2.548 Å for the reconstructed volume.

We generated 41000 projections from this volume according to the forward model (2). Shifts were not used. The projection directions were sampled from a uniform distribution over 3D rotations. To set the CTF parameters for each projection, we randomly picked a micrograph in the EMPIAR- 10061 dataset, extracted its CTF parameters using CTFFIND4 [70], and applied them to the projection. The parameter B of the envelope function of the CTF (see Supplemenary Material (42)) was chosen such that it decays to 0.2 at the Nyquist frequency. To simulate a realistic noise for each projection, we randomly selected a background patch from the same micrograph used for the CTF parameters. This noise patch was downsampled to size (192×192), normalized to zero-mean, and scaled and added to the projection. The scaling was such that the signal-to-noise ratio was -20 dB.

Generator Settings. For each half-map reconstruction, the generator was initialized with a volume of size $184 \times 184 \times 184$ voxels; the pixel size was 1.274 Å. Because β-galactosidase has D2 symmetry, this volume was parameterized by a quarter-volume (the first $92 \times 92 \times 184$ voxels) that was copied and flipped to create the full volume. We applied a simple, *ad hoc* regularization scheme intended to promote nonnegativity during the initial phases of reconstruction and to increase the stability of the algorithm, details of which being given in the Supplementary Materials Section D.

The cryo-EM simulator in the generator matched the process used to generate the data, except that we assumed that the final SNR of each projection was unknown, leading us to learn the scaling parameter that controls the ratio between the projections and the noise patches. Details on the scaling operation can be found in the Supplementary Materials Section D.

Training Settings. For the optimization, we used Adam [64] ($\beta_1 = 0.5, \beta_2 = 0.9, \epsilon = 10^{-8}$) with a learning rate of 10^{-3} and

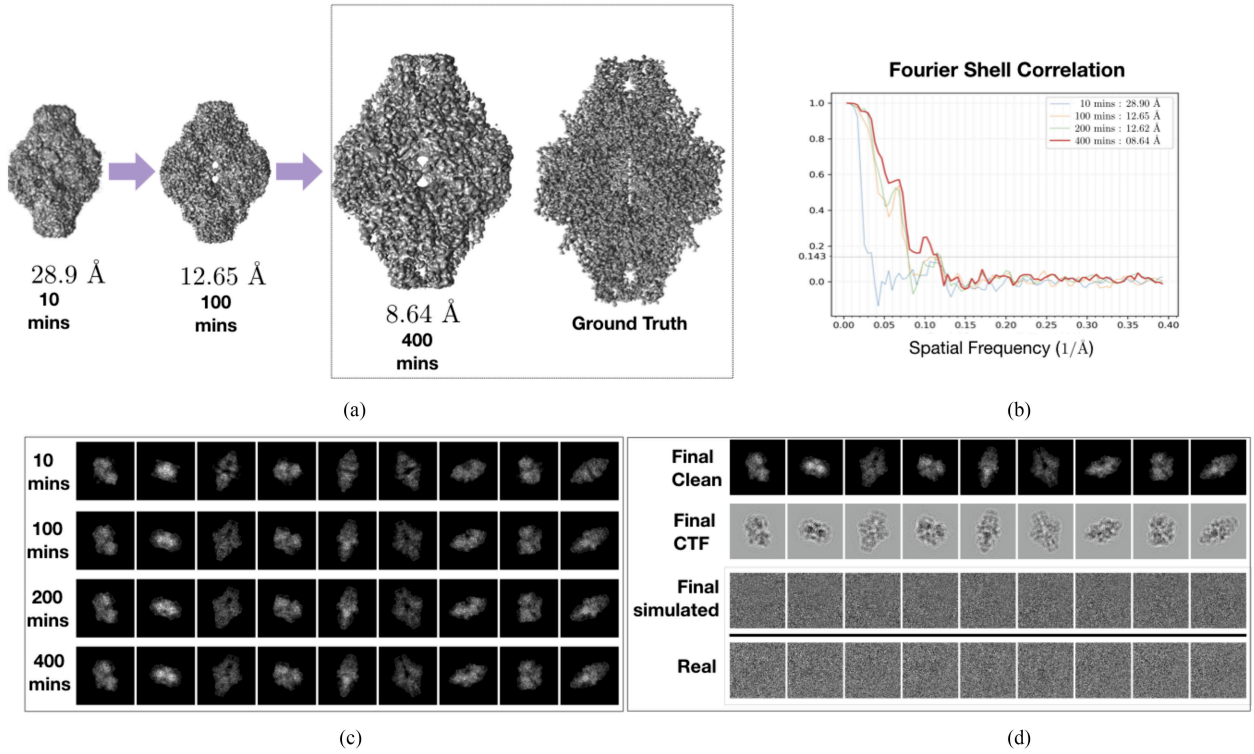

Fig. 3. CryoGAN is applied on a synthetic projection dataset generated from a 2.5Å β-galactosidase volume. We refer to these synthetic projections as "real," in contrast to the projections coming from the CryoGAN physics simulator, which we term "simulated." (a) The volume is initialized with zeros and is progressively updated to produce projections whose distribution matches that of the real projections. (b) Corresponding evolution of half-half FSC through time. (c) Evolution during training of some clean projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. (d) *Row 1:* Clean, simulated projections (before CTF and noise) generated at the final stage of training. *Row 2:* CTF-modulated simulated projections (before noise) generated at the final stage of training. *Row 3:* Simulated projections (with CTF and noise) generated at the final stage of training. *Row 4:* Real projections, for comparison.

a batch size of 8. The algorithm was run for 40 epochs and the learning rate decreased by 1% at every epoch. The parameter for the gradient-penalty term (13) was set to $\lambda = 0.001$. The discriminator was updated 4 times for every update ($n_{\text{discr}} = 4$ in Algorithm 1).

Discriminator Settings. The discriminator weights were initialized to random values using PyTorch default distributions. All projections were normalized to zero mean and unit standard deviation before entering the discriminator.

Results. We ran the CryoGAN algorithm for 400 minutes on an NVIDIA V100 GPU and obtained a reconstruction with a half-map resolution of 8.64 Å (additional metrics given in Table I). This confirms that our novel scheme produces reconstruction with an acceptable quality, although the resolution is still significantly below the limit that can be expected from the data. We show in Figure 3 the progression of the CryoGAN reconstruction. It demonstrates that CryoGAN quickly finds a reasonable density map, then fills in details.

B. Additional Imaging Conditions

To understand the effect of different imaging conditions on the quality of the CryoGAN reconstruction, we created additional synthetic datasets: a dataset with a low noise level (-5.2 dB SNR); a dataset with a realistic noise level (-20 dB SNR) and moderate translations (3% of the image size); and a dataset with a realistic noise level (-20 dB SNR) and large translations (20% of the image size). For the last two cases, the translation (both horizontal and vertical) for each projection is first sampled from a zero-mean symmetric triangular distribution whose total width is 6% and 40% of the image size, quantized to the nearest integer. All other conditions remain identical to the main synthetic experiment, except that the number of epochs for reconstruction was 100 and the volume clipping value in the generator was -5%.

Results. The resolutions of the reconstructions are given in Table I. As expected, the decrease of the noise level from -20 dB to -5.2 dB improves the reconstruction resolution (from 8.64 Å to 7.53 Å) and results in a sharper looking reconstruction (Figure 4). For the second and third case, the presence of translations results in slightly lower resolutions. We suspect this is because the discriminator must learn to ignore the translations, which, intuitively, makes its task more challenging.

C. Effect of a Mismatch in the Pose Distribution

We performed another series of synthetic experiments to understand the robustness of CryoGAN with respect to the mismatch between the pose distributions associated with the real projections $p_{\boldsymbol{\theta}}^{\text{true}}$ and the simulated projections $p_{\boldsymbol{\theta}}^{\text{rec}}$. We

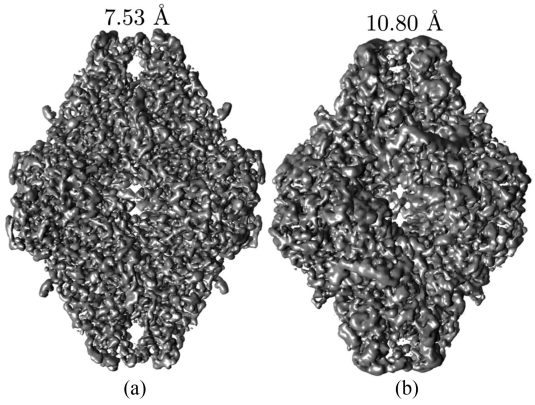

Fig. 4. Additional CryoGAN reconstructions for synthetic datasets with different imaging conditions. (a) Reconstruction for a low noise case (-5.2 dB SNR). (b) Reconstruction for a realistic noise level (-20 dB SNR) and with translations (3% of the image size) in the data. The reported resolutions correspond to half-map FSC at 0.143 cutoff.

TABLE II
RECONSTRUCTION RESOLUTION (Å) FOR β-GALACTOSIDASE

	reconstruction distribution		
true distribution	uniform	$\sigma = 2$	$\sigma = 3$
half-map FSC = 0.143			
uniform	8.6	13.2	17.2
$\sigma = 2$	9.2	9.4	14.7
$\sigma = 3$	8.6	10.6	12.7
ground truth FSC = 0.5			
uniform	15.3	16.1	18.6
$\sigma = 2$	16.4	15.0	15.5
$\sigma = 3$	16.4	16.9	16.1
ground truth FSC = 0.143			
uniform	11.7	12.6	14.5
$\sigma = 2$	11.6	10.4	10.7
$\sigma = 3$	11.5	14.0	12.3

used three synthetic datasets, each one created using different pose distributions $p_{\boldsymbol{\theta}}^{\text{true}}$. The first dataset was the same as in the main synthetic experiment (uniform distribution). The second and the third datasets used nonuniform distributions of the poses, which were sampled by generating random samples from a 3D multivariate Gaussian and computing the corresponding Euler angles. The Gaussians had zero mean and diagonal covariance $\boldsymbol{\Sigma} = \text{diag}(\sigma, 2, 1)$ with $\sigma = 2$ and 3. On each dataset, we performed three CryoGAN reconstructions characterized using the same three pose distributions $p_{\boldsymbol{\theta}}^{\text{rec}}$ in the physics simulator.

Results. The reconstruction resolutions are shown in Table II. We observe that the best resolution is usually achieved when the assumed pose distribution matches the real one (diagonal values tend to be the lowest in each row and column, especially when the FSC is computed with respect to the ground truth). However, assuming a uniform distribution during reconstruction provides reasonable reconstructions in all cases (never more than 1.4 Å worse than the best choice of distribution), which justifies its use when the true distribution is not known.

TABLE III
RECONSTRUCTION RESOLUTION (Å) FOR SYNTHETIC RIBOSOME

	noise level (dB)				
metric	20	0.0	-5.2	-14	-20
half-map FSC = 0.143	15.5	18.8	19.5	63.0	57.1
truth FSC = 0.5	20.9	21.9	24.1	81.0	104.3
truth FSC = 0.143	15.3	19.6	20.5	45.1	57.2

D. 80S Ribosome Datasets With Various Noise Levels

We reconstructed synthetic datasets that mimic the 80S Ribosome dataset (EMPIAR-10028) from [71]. The ground-truth structure with dimensions $(180 \times 180 \times 180)$ and voxel size 2.68 Å was obtained by processing the EMDB-2660 reconstruction from EMPIAR-10028. We created five datasets with identical imaging conditions (same as EMPIAR-10028), except for the noise levels. Their SNR were kept at 20 dB, 0 dB, -5.2 dB, -14 dB, and -20 dB. The noise patches were extracted from the micrograph background in EMPIAR-10028. In all datasets, translations were kept at 5% of the image size. We then performed reconstructions using CryoGAN for 20 epochs. For more details, see Supplementary Materials Section D.

Results. The reconstruction quality is reported in Table III. As expected, the resolution decreases with the increase in the noise level. When the noise level is low (-5.7 dB), we achieve 19.5 Å, with an isosurface rendering provided in Figure 6. For similar imaging conditions, the resolution achieved for Ribosome is worse than that of β-galactosidase. We believe this is because of the lack of symmetry in Ribosome, the fact that the larger pixel size for its projections (2.68 Å), and because of its large blob-like structure. We show in Table III and Figure 6 that the main culprit behind the loss of resolution is the noise level.

E. Experimental Data—βbeta-Galactosidase (EMPIAR-10061)

To assess the capacity of CryoGAN to reconstruct real, experimental data, we deploy it on the β-galactosidase projection data provided in EMPIAR-10061 [68]. These data consist of 1539 micrographs and 41123 particle locations. These were obtained via automated particle picking followed by 3D classification, as described in [68].

The details of this experiment matched those of the synthetic experiment (Section VI-A) with the following changes:

Dataset. The projections, originally of size (384×384) were downsampled to (192×192), with a pixel size of 1.274 Å. This corresponds to a Nyquist resolution of 2.548 Å for a reconstructed volume of size $(180 \times 180 \times 180)$. The defocus and astigmatism parameters of the CTF were estimated for each micrograph using CTFFIND4 [70].

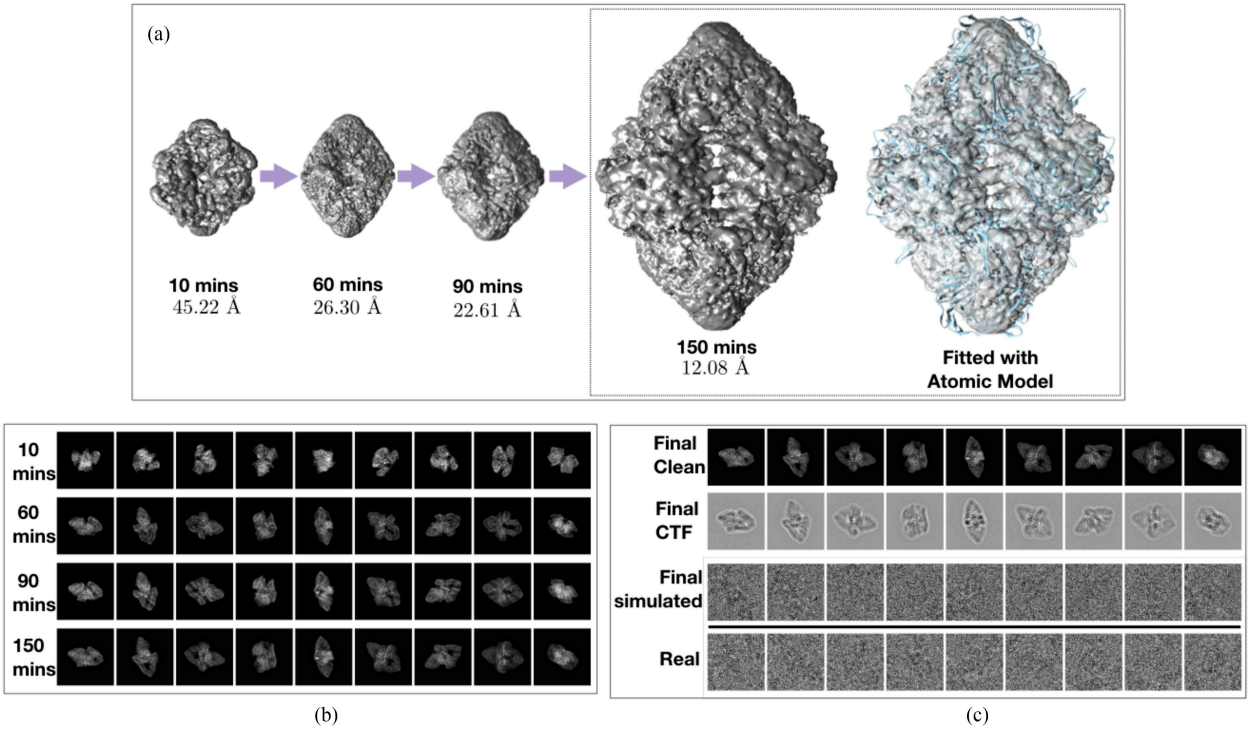

Fig. 5. Evolution of CryoGAN while reconstructing the experimental β-galactosidase dataset (EMPIAR-10061) from [68]. (a) The volume is initialized with zeros and is progressively updated to produce projections whose distribution matches that of the experimental dataset. (b) Evolution during the training of the clean projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. (c) *Row 1:* Clean, simulated projections generated at the final stage of training. *Row 2:* CTF-modulated, simulated projections (before noise) generated at the final stage of training. *Row 3:* Simulated projections (with CTF and noise) generated at the final stage of training. *Row 4:* Real projections.

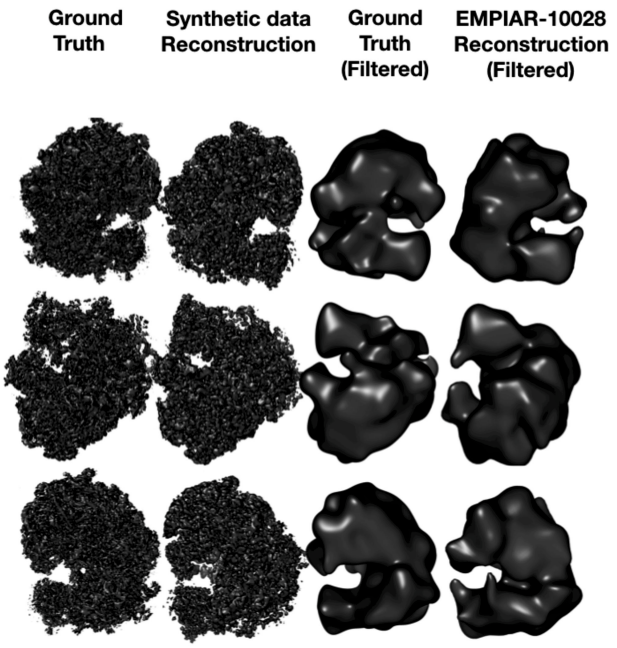

Fig. 6. Different views of the 80S Ribosome and reconstruction using CryoGAN. (First column) 80S Ribosome (EMD-2660) from EMPIAR-10028. (Second-column) CryoGAN reconstruction from synthetic data with -5.2 dB noise and 5% translations. (Third column) 80S Ribosome (EMD-2660) filtered with a Gaussian with standard deviation of 5 voxels. (Fourth column) CryoGAN reconstruction filtered with a Gaussian with standard deviation of 5 voxels.

Generator Settings. We reconstruct a volume of size $(180 \times 180 \times 180)$ voxels for each half dataset. The CTF parameters estimated with CTFFIND4 [70] are used in the forward model of the cryo-EM physics simulator. The parameter B of the envelope function of the CTF (see (42)) decays to 0.4 at the Nyquist frequency. The translations are set to zero.

To handle the noise, we randomly extract (prior to the learning procedure) 41123 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 from that micrograph. Each noise patch is then downsampled to size (192×192) and normalized. Then, during run-time, the noise patches are sampled from this collection, scaled, and added to the simulated projections. For consistency, the noise patch added to a given simulated projection is taken from the same micrograph as the one that was used to estimate the CTF parameters previously applied to that specific projection. The scaling operation weighs the contribution of the noise with respect to the projection signal. More details on the addition of noise and the regularization constraints enforced on the volume during the reconstruction are given in Supplementary Materials Section D.

Discriminator Architecture. The discriminator is initialized identically for the two half datasets. The projection images (real and simulated) are smoothed with a Gaussian kernel before being fed to the discriminator. The standard deviation of the kernel is

initially set at 2 and changes in every iteration so that it decreases by a total of 2% in each epoch.

Training Settings. The algorithm is run for 15 epochs and the learning rate decreases by 8% at every epoch. The parameter for the gradient-penalty term is set to $\lambda = 1$ (see (13)) and $n_{\text{discr}} = 4$.

The norm of the gradients for the discriminator is clipped to a maximal value of 10^7. For the generator, the gradients for each pixel are clipped to a maximal value of 10^3. The gradients that correspond to the learning of the scaling ratios between the noise and projection images are clipped to a value of 10.

Results. We let CryoGAN reconstruct the real EMPIAR-10061 dataset for 150 minutes on an NVIDIA V100 GPU. For reproducibility, we ran the reconstruction five times and report the averaged metrics. We achieved a mean resolution of 11.2 Å with a standard deviation of 1.14 Å. We obtained the mean resolutions of 26.6 Å and 19.9 Å with respect to the true structure for 0.5 and 0.143 FSC cutoffs, respectively. The standard deviation for these metrics were 0.45 Å and 0.49 Å, respectively. The results are displayed in Figure 5. We note that higher-resolution details are progressively introduced in the estimated volume throughout the run, as illustrated by the evolution of the FSC curves between the reconstructed half-maps (Figure 5 d). This resulted in a 12.08 Å β-galactosidase structure whose simulated projections closely resemble the real ones (Figure 5 c).

F. Reconstruction of βbeta-Galactosidase With Pose Distribution From Relion

We performed another CryoGAN reconstruction on β-galactosidase (EMPIAR-10061) where, instead of a uniform distribution, we set $p_{\boldsymbol{\theta}}^{\text{rec}}$ to be the distribution associated with the poses obtained from Relion [9]. This distribution acts as a proxy for the true pose distribution $p_{\boldsymbol{\theta}}^{\text{true}}$. Hence, this experiment considers the case when the mismatch between $p_{\boldsymbol{\theta}}^{\text{true}}$ and $p_{\boldsymbol{\theta}}^{\text{rec}}$ is minimal. All the other reconstruction parameters were kept the same as in the main experiment. More details are available in the Supplementary Materials.

Results. We achieved a half-half resolution of 11.47 Å. The resolution with respect to the true structure was 32.8 Å and 20.9 Å for 0.5 and 0.143 FSC cutoffs, respectively. The reconstruction quality is similar to the case when the uniform distribution is assumed. This, we believe, is because of two main reasons: firstly, the distribution obtained from the Relion poses is very similar to a uniform distribution. This can be seen in Figure 7.d in the Supplementary Materials and has also been reported for EMPIAR-10061 in [72]. Secondly, as shown on the synthetic dataset in Table II, reconstructions using a uniform distribution give fairly good and consistent results and are similar to the one that can be achieved with using a distribution with less mismatch.

G. Experimental Data—80S Ribosome (EMPIAR-10028)

We also deployed CryoGAN on the experimental projection data obtained from 80S Ribosome [71] provided in EMPIAR-10028. This dataset contains 102547 projections. For more details about the dataset and the reconstruction procedure, see the Supplementary Materials.

Results. We achieved a half-half resolution of 55.5 Å resolution. The resolution with respect to the true structure was 100.7 Å and 54.2 Å for 0.5 and 0.143 FSC cutoffs, respectively. We provide the reconstruction in Figure 6. It matches reasonably well a smoothed version of the ground truth.

Like was the case with the synthetic data, the results obtained with the experimental 80S Ribosome dataset are not as good as those for β-galactosidase. This is probably because the former does not have any symmetry, while its structure is larger and more blobby. These issues decrease the effective SNR of the Ribosome data (more number of variables to predict) and, hence, result in lower resolutions. As shown in Table III, an increase in SNR improves the resolution for Ribosome in the synthetic experiments. We believe that, for experimental data, methods that help mitigate the effect of noise, for instance multiresolution approaches for GANs [73], [74], would improve the resolution for Ribosome-like structures. See Section VII-B for more discussions. Moreover, the pixel size in the Ribosome data (2.68 Å) is twice coarser than in β-galactosidase data. This limits access to the information needed to reconstruct at higher resolutions.

VII. Discussion

We have presented CryoGAN, an algorithm for single-particle cryo-EM reconstruction motivated by distribution matching. Our experiments showed that CryoGAN is able to reconstruct reasonable structures, even from highly noisy experimental data. We now discuss the connections between the CryoGAN algorithm and the theory of distribution matching that we developed in Section IV, as well as future steps to improve the CryoGAN algorithm.

A. CryoGAN and Theorem 1

It is important to understand that, while the CryoGAN algorithm attempts to solve a distribution-matching problem, there are several reasons why Theorem 1 does not guarantee that CryoGAN will provide a perfect reconstruction. Firstly, several details complicate the successful minimization of the distribution distance in (11): the empirical estimates in (11) require infinite data to be perfectly accurate, the loss function in (11) is non-convex, and the discriminator D_ϕ in (13) may not span all the valid functions s with Lipschitz less than or equal to 1 in (11). Secondly, the algorithm warrants discretization of projections, structures, and the forward model in (8) which is a slight (but controllable) deviation from continuous-domain formulation of Theorem 1. Additional work is thus required to understand the effect of these nonidealities on the quality of reconstruction. One step in this direction would be to obtain a bound on the reconstruction error as a function of the distributional distance (8).

Another point to note is that there are many algorithms other than CryoGAN that exploit the distribution-matching approach. This includes current ML techniques [9], [11] that minimize the KL-divergence (using it as D in (8)) to achieve distribution matching. However, because of the structure of this divergence, ML methods need to marginalize away the imaging parameters. Unlike them, CryoGAN relies on Wasserstein distance and,

hence, does not need marginalization but only a method to sample from the two distributions (11). For more details, see the Supplementary Materials.

B. Future Improvements

While the spatial resolution of the CryoGAN reconstructions from real data is not yet competitive with the state-of-the-art (which typically is around 3-4 Å), the algorithm is already able to steadily perform the hardest part of the job, which is to obtain a reasonable structure by using nothing but the particle dataset and CTF estimations. We believe that some key developments will help to bring the CryoGAN algorithm to the stage where it becomes a relevant contributor for high-resolution reconstruction in single-particle cryo-EM. We have laid out a roadmap for future improvements that should get us to this stage, and may eventually help us reconstruct dynamic structures.

Beyond simple engineering tweaks (*e.g.*, a tuning of the number of layers in the discriminator, a change of the optimization strategies, or the use of Fourier methods to accelerate the projection), we expect several directions of research to be fruitful.

Multiresolution. A promising line of research is the use of a coarse-to-fine strategy in CryoGAN. Several GAN architectures rely on such approaches, such as the progressive GANs [74] and the styleGANs [73]. In the context of CyroGAN, this could involve the reconstruction of a low-resolution structure from lowpass-filtered measurements followed by a progressive increase in the resolution of the estimated structure, the resolution of the measurements, and the complexity of the discriminator. The motivation is twofold. First, similar to [73], [74], an increased robustness during the low-resolution regime tends to have a positive impact on the convergence of the higher-resolution steps. Second, unique to CryoGAN, the approach could help with highly noisy data, because a lowpass-filtering of the measurements tends to increase their SNR. We expect a multiresolution approach to be especially beneficial in the reconstruction of non-symmetric structures from noisy data (*e.g.*, the 80S ribosome Section VI-G), as we have shown in Section VI-D that noise is a major factor that limits the resolution of the CryoGAN reconstruction.

The impact of multiresolution techniques on CryoGAN could be as important as the one they had on GANs, where they allowed the field to progress in just a few years from the generation of blurry facial images [27] to simulated images indistinguishable from real facial images [73], [74], More generally, the upcoming tools and extensions in GAN architectures could bring significant gains in resolution to the CryoGAN implementation.

The benefits of multiscale reconstructions could be considerable for CryoGAN, given the extremely difficult imaging conditions that prevail in single-particle cryo-EM and that make the convergence of optimization algorithms to good solutions particularly challenging.

Volume Regularization. Many cryo-EM reconstruction packages include some regularization of the volume itself, for instance, the smoothness prior used in Relion [75]. The same could be easily included in CryoGAN, provided that the regularization function is amenable to gradient descent.

Improved Forward Model. The performance of the cryo-EM physics simulator should also improve hand-in-hand with our computational ability to precisely model the physics that govern single-particle cryo-EM. At the moment, CryoGAN assumes that the noise is additive in its image-formation model. One could alternatively consider a Poisson-noise-based forward model [55], [56]. This would however require backpropagation through a Poisson distribution, a nontrivial operation.

Bypassing Particle Picking. Another interesting extension of the simulator would be to directly simulate the patches of nonaligned micrographs or frames (rather than the individual projections) and match their distribution to that of the raw dataset. Doing so would allow CryoGAN to bypass particle picking. Some reconstruction approaches [76], [77] have already attempted to bypass the particle-picking procedure, but with limited success so far.

Learning the Pose Distribution. Similar to likelihood-based methods, the CryoGAN algorithm requires the specification of the distribution of poses. In the case of CryoGAN, one could also parameterize this distribution and learn its parameters during the reconstruction procedure, along the lines of [35]. The same approach could be used to calibrate the distribution of the translations of the projections.

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

Removal of Corrupt Particles. Like all reconstruction algorithms, CryoGAN can fail if the dataset contains too many corrupted particle images, typically those with broken structures or strong optical aberrations. Several solutions could be deployed to handle excessive outliers in the data distribution. One approach would be to include a step that automatically spots and discards corrupted data so that the discriminator never gets to see them. Recent deep-learning-based approaches able to track outliers in data could prove useful in this regard [46].

APPENDIX
A PROOF OF THEOREM 1

We prove the Theorem in three steps. First, we consider the case when the forward model consists only of random projections but lacks corruption by noise and CTF. The recovery guarantee in this case has been given in [28, Theorem 3.1]; its main concept will be used in the second part. In the second step, we prove the recovery when the forward model consists of random projections and CTF but lacks corruption by noise. In the third step, we prove recovery for the full forward model that includes noise.

A. Notations and Preliminaries

We begin by introducing notations, for which we follow [28].

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

For brevity, we use in the next two steps of the proof general $f \in \mathbb{F}$ and $g \in \mathbb{F}$ instead of f_{true} and f_{rec}, $p_{\boldsymbol{\theta}}$ instead of $p_{\boldsymbol{\theta}}^{\text{true}}$ and $p_{\boldsymbol{\theta}}^{\text{rec}}$, and $p_{\mathbf{c}}$ instead of $p_{\mathbf{c}}^{\text{true}}$ and $p_{\boldsymbol{\theta}}^{\text{rec}}$. Note that, by the assumptions of Theorem 1, $p_{\boldsymbol{\theta}}^{\text{true}} = p_{\boldsymbol{\theta}}^{\text{rec}}$ and $p_{\mathbf{c}}^{\text{true}} = p_{\mathbf{c}}^{\text{rec}}$.

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

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

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \Psi_{\mathbf{A}}[\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{A}}\{f\} \in \cdot\}]. \tag{15}$$

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

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \mathbb{P}_{\text{proj}}(\cdot|\mathbf{R}f), \ \forall f \in \mathbb{F} \text{ and } \mathbf{R} \in O(3), \tag{16}$$

where O_3 is the space of all orthogonal matrices such that $\det \mathbf{A} \in \{-1, 1\}$. The invariance in (16) is true since

$$\begin{aligned}
\mathbb{P}_{\text{proj}}(\cdot|f) &= \Psi[\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{A}}\{f\} = \cdot\}] \\
&= \Psi[\{\mathbf{A} \in SO(3) : \mathcal{P}_{\mathbf{R}^{-1}\mathbf{A}}\{\mathbf{R}f\} = \cdot\}] \\
&= \Psi[\{\mathbf{R}\mathbf{A}' \in SO(3) : \mathcal{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\}] \\
&= \Psi[\{\mathbf{A}' \in SO(3) : \mathcal{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\}], \qquad (17)
\end{aligned}$$

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

$$(\gamma_A f)(\cdot) = f(\mathbf{A}^{-1}\cdot), \forall \mathbf{A} \in O(3), f \in \mathbb{F}. \tag{18}$$

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

B. Recovery in the Absence of CTF and Noise

We now consider a simple forward model consisting only of a projection operation. The guarantee of recovery for this case was given in [28, Theorem 3.1], which we now restate and discuss briefly.

Theorem 2 ([28, Theorem 3.1]): Let $p_{\mathbf{A}}$ be any bounded distribution on $SO(3)$ and let the assumptions of Theorem 1 be true; then, $\forall f, g \in \mathbb{F}$,

$$[f] \neq [g] \Rightarrow \mathbb{P}_{\text{proj}}(\cdot|f) \perp \mathbb{P}_{\text{proj}}(\cdot|g). \tag{19}$$

Sketch of the Proof: Without loss of generality, we provide the sketch of the proof for the case when $p_{\mathbf{A}}$ is uniform. For the case when $p_{\mathbf{A}}$ is nonuniform the argument remains the same, provided that $\Psi_{\mathbf{A}}$ associated with the nonuniform distribution $p_{\mathbf{A}}$ is absolutely continuous with respect to Ψ ($\Psi_{\mathbf{A}} \ll \Psi$). This has been stated in [28]. 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 [78, Proposition 7.8] which we restate here as Proposition 1.

Proposition 1 ([78, Proposition 7.8]): Let $f \in \mathbb{F}$ and let $S_{\mathbf{A}}$ be an uncountably infinite subset of $SO(3)$. Then, f is determined by the collection $\{\mathcal{P}_A\{f\}\}_{\mathbf{A} \in S_{\mathbf{A}}}$ ordered with respect to $\mathbf{A} \in S_{\mathbf{A}}$.

Note that this proposition assumes that the angles of the projections are known. Even though the angles are unknown in our case, we shall see that this proposition remains useful.

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

$$\text{TV}(\mathbb{P}_1, \mathbb{P}_2) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{(y_1, y_2) \sim \gamma}[\mathbb{1}_{y_1 \neq y_2}], \tag{20}$$

where TV is the total-variation distance and $\Pi(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all the joint distributions $\gamma(y_1, y_2)$ whose marginals are $\mathbb{P}_1$ and $\mathbb{P}_2$ [61]. In fact, $\mathbb{E}[\mathbb{1}_{y_1 \neq y_2}]$ is equal to the probability of the event $y_1 \neq y_2$. In our context, this translates into

$$\text{TV}(\mathbb{P}_{\text{proj}}(.|f), \mathbb{P}_{\text{proj}}(.|g)) = 2 \inf_{\gamma \in \Xi} \text{Prob}(y_1 \neq y_2), \tag{21}$$

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

We can then write that

$$\Psi[\{\mathbf{A} \in SO(3) : \mathcal{P}_{h(\mathbf{A})}\{g\} \in \cdot\}] = \mathbb{P}_{\text{proj}}(\cdot|g). \tag{22}$$

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

When h is continuous, Proposition 1 implies that, if $[f] \neq [g]$, then

$$\Psi[\{\mathbf{A} \in SO(3) : \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{h(\mathbf{A})}\{g\}\|_2 > 0\}] = 1. \tag{23}$$

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

leads to

$$\Psi[\{\mathbf{A} \in SO(3) : \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{h(\mathbf{A})}\{g\}\|_2 > 0\}] \geq \Psi(H_\delta) \geq 1 - \delta. \tag{24}$$

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

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

C. Recovery in the Presence of CTF and Absence of Noise

We now extend the previous result to the case when the CTF is also present in the forward model. For the sake of simplicity, we do not take into account shifts in the forward model. Note that it is trivial to generalize the results to shifts since they do not affect the information content in the projections, only their location.

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

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|f) = \Psi_{\mathbf{A},\mathbf{c}}[\{(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathcal{C}) : C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} \in \cdot\}], \tag{25}$$

where $C_{\mathbf{c}}$ is the space-domain CTF given in the Supplementary Materials.

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

$$[f] \neq [g] \Rightarrow \mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g). \tag{26}$$

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

$$(y_1, y_2) = (C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\}, C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\}), \tag{27}$$

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

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|g) = \Psi_{\mathbf{A},\mathbf{c}}[\{(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathcal{C}) : C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\} \in \cdot\}]. \tag{28}$$

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

We first assume that h is continuous and follow the strategy of [28, Theorem 3.1]. Since $SO(3)$ is transitive, we can write that $h(\mathbf{A},\mathbf{c}) = (\mathbf{A}\Gamma_{\mathbf{A},\mathbf{c}}, h_1(\mathbf{A},\mathbf{c}))$.

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

$$h_n(\mathbf{A},\mathbf{c}) = (\mathbf{A}\Gamma_n^m, h_{n,1}^m(\mathbf{A},\mathbf{c})) \; \forall \; (\mathbf{A},\mathbf{c}) \in \{\mathcal{A}_n^m \times \mathcal{C}_n^m\}, \tag{29}$$

such that

$$\Gamma_n^m = \arg\min_{\Gamma \in \eta} \min_{(\mathbf{A},\mathbf{c}) \in \{\bar{\mathcal{A}}_n^m \times \bar{\mathcal{C}}_n^m\}} \|\mathcal{P}_{\mathbf{A}}\{f\} - \mathcal{P}_{\mathbf{A}\Gamma}\{g\}\|, \tag{30}$$

where $\eta = \{\Gamma_{\mathbf{A},\mathbf{c}} : (\mathbf{A},\mathbf{c}) \in \{\bar{\mathcal{A}}_n^m \times \bar{\mathcal{C}}_n^m\}\}$, $\bar{\mathcal{A}}_n^m$, and $\bar{\mathcal{C}}_n^m$ are the closures of $\mathcal{A}_n^m$ and $\mathcal{C}_n^m$, respectively. The sequence h_n converges to h as $n \to \infty$. We denote

$$\mathcal{K} = \{(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathcal{C}) : \|C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}\Gamma_{\mathbf{A}}}\{g\}\| > 0\}, \tag{31}$$

$$\mathcal{K}_n = \{(\mathbf{A},\mathbf{c}) \in (\mathcal{A}_n^m \times \mathcal{C}_n^m) : \|C_{\mathbf{c}} * \mathcal{P}_{\mathbf{A}}\{f\} - C_{h_1(\mathbf{A},\mathbf{c})} * \mathcal{P}_{\mathbf{A}}\{\Gamma_n^m g\}\| > 0\}. \tag{32}$$

Similarly to [28, Theorem 3.1], we can then show that

$$\Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}] = \lim_{n \to \infty} \sum_{m=1}^{m=n} \Psi_{\mathbf{A},\mathbf{d}}[\mathcal{K}_m]. \tag{33}$$

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

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

D. Recovery for the Full Forward Model

We now consider the case with the full forward model. The proof for this case hinges on the result from Appendix A–C. For clarity, we again bring back the notation used in Theorem 1. We denote the probability measure of $y_{\text{clean}} = H_{\boldsymbol{\varphi}} f_{\text{true}}$ with $\mathbb{P}_{\text{clean}}^{\text{true}}(\cdot|f_{\text{true}})$. Similarly, $\mathbb{P}_{\text{clean}}^{\text{rec}}(\cdot|f_{\text{rec}})$ is obtained from f_{rec}. (Recall that $\boldsymbol{\varphi}$ is a random vector that represents imaging parameters; therefore, y_{clean} is also a random entity.) We shall prove the following in sequence:

$$\mathbb{P}(\cdot|f_{\text{true}}) = \mathbb{P}(\cdot|f_{\text{rec}}) \Leftrightarrow \mathbb{P}_{\text{clean}}(\cdot|f_{\text{true}}) = \mathbb{P}_{\text{clean}}(\cdot|f_{\text{rec}}), \tag{34}$$

and

$$\mathbb{P}_{\text{clean}}(\cdot|f_{\text{true}}) = \mathbb{P}_{\text{clean}}(\cdot|f_{\text{rec}}) \Leftrightarrow f_{\text{rec}} = G(f_{\text{true}}). \quad (35)$$

For the first part, we progress by noting that $y = y_{\text{clean}} + n$. Recall that the characteristic function of the probability measure associated to the sum of two random variables is the product of their characteristic functions. Mathematically, $\hat{\mathbb{P}}(\cdot|f_{\text{true}}) = \hat{\mathbb{P}}_{\text{clean}}(\cdot|f_{\text{true}})\hat{\mathbb{P}}_n^{\text{true}}$ and $\hat{\mathbb{P}}(\cdot|f_{\text{rec}}) = \hat{\mathbb{P}}_{\text{clean}}(\cdot|f_{\text{rec}})\hat{\mathbb{P}}_n^{\text{rec}}$.

By Assumption 1), we can now write that

$\hat{\mathbb{P}}_{\text{clean}}(\cdot|f_{\text{true}}) = \frac{\hat{\mathbb{P}}(\cdot|f_{\text{true}})}{\hat{\mathbb{P}}_n^{\text{true}}}$ and $\hat{\mathbb{P}}_{\text{clean}}(\cdot|f_{\text{rec}}) = \frac{\hat{\mathbb{P}}(\cdot|f_{\text{rec}})}{\hat{\mathbb{P}}_n^{\text{rec}}}$.

Then, given that $\mathbb{P}(\cdot|f_{\text{true}}) = \mathbb{P}(\cdot|f_{\text{rec}})$ and $\mathbb{P}_n^{\text{true}} = \mathbb{P}_n^{\text{rec}}$, it is easy to conclude that $\mathbb{P}_{\text{clean}}(\cdot|f_{\text{true}}) = \mathbb{P}_{\text{clean}}(\cdot|f_{\text{rec}})$.

For the second part, we use Theorem 3 which is applicable to the case when only noise is absent from the forward model. We have that $f = f_{\text{true}}$, $g = f_{\text{rec}}$, and $\mathbb{P}_{\text{proj,CTF}} = \mathbb{P}_{\text{clean}}$. Then, given that $p_{\mathbf{c}}^{\text{true}} = p_{\mathbf{c}}^{\text{rec}} = p_{\mathbf{c}}$ and $p_{\boldsymbol{\theta}}^{\text{true}} = p_{\boldsymbol{\theta}}^{\text{rec}} = p_{\boldsymbol{\theta}}$, Theorem 3 implies that $\mathbb{P}_{\text{clean}}(\cdot|f_{\text{true}}) \not\perp \mathbb{P}_{\text{clean}}(\cdot|f_{\text{rec}}) \Rightarrow [f_{\text{rec}}] = [f_{\text{true}}]$. Since $\mathbb{P}_{\text{clean}}(\cdot|f_{\text{true}}) = \mathbb{P}_{\text{clean}}(\cdot|f_{\text{rec}})$ it follows that $f_{\text{rec}} = G(f_{\text{true}})$. This concludes the proof.

ACKNOWLEDGMENT

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 and to Dr. Jonathan Dong (EPFL) and Dr. Jasenko Zivanov (EPFL) for their help in obtaining the pose data for EMPIAR-10061.

REFERENCES

[1] J. Dubochet *et al.*, "Cryo-electron microscopy of vitrified specimens," *Quart. Rev. Biophys.*, vol. 21, no. 2, pp. 129–228, 1988.
[2] R. Henderson, J. M. Baldwin, T. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, "Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy," *J. Mol. Biol.*, 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*. London, U.K.: Oxford Univ. 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 *et al.*, "Xmipp: A new generation of an open-source image processing package for electron microscopy," *J. Struct. Biol.*, vol. 148, no. 2, pp. 194–204, 2004.
[6] G. Tang *et al.*, "Eman2: An extensible image processing suite for electron microscopy," *J. Struct. Biol.*, vol. 157, no. 1, pp. 38–46, 2007.
[7] N. Grigorieff, "Frealign: High-resolution refinement of single particle structures," *J. Struct. Biol.*, vol. 157, no. 1, pp. 117–125, 2007.
[8] M. Hohn *et al.*, "SPARX, a new environment for Cryo-EM image processing," *J. Struct. Biol.*, vol. 157, no. 1, pp. 47–55, 2007.
[9] S. H. Scheres, "RELION: Implementation of a bayesian approach to cryo-EM structure determination," *J. Struct. Biol.*, vol. 180, no. 3, pp. 519–530, 2012.
[10] J. de la Rosa-Trevín *et al.*, "Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy," *J. Struct. Biol.*, 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. F. Reboul, M. Eager, D. Elmlund, and H. Elmlund, "Single-particle cryo-EM-Improved ab initio 3 d reconstruction with SIMPLE/PRIME," *Protein Sci.*, vol. 27, no. 1, pp. 51–61, 2018.
[13] 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.
[14] T. Baker and R. Cheng, "A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy," *J. Struct. Biol.*, vol. 116, no. 1, pp. 120–130, 1996.
[15] A. Barnett, L. Greengard, A. Pataki, and M. Spivak, "Rapid solution of the cryo-EM reconstruction problem by frequency marching," *SIAM J. Imag. Sci.*, vol. 10, no. 3, pp. 1170–1195, 2017.
[16] B. Vainshtein and A. Goncharov, "Determination of the spatial orientation of arbitrarily arranged identical particles of an unknown structure from their projections," *Sov. Phys. Doklady*, vol. 31, pp. 278–280, Apr. 1986.
[17] S. P. Mallick, S. Agarwal, D. J. Kriegman, S. J. Belongie, B. Carragher, and C. S. Potter, "Structure and view estimation for tomographic reconstruction: A bayesian approach," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2006, pp. 2253–2260.
[18] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky, "Detecting consistent common lines in cryo-EM by voting," *J. Struct. Biol.*, vol. 169, no. 3, pp. 312–322, 2010.
[19] L. Wang, A. Singer, and Z. Wen, "Orientation determination of cryo-EM images using least unsquared deviations," *SIAM J. Imag. Sci.*, vol. 6, no. 4, pp. 2450–2483, 2013.
[20] I. Greenberg and Y. Shkolnisky, "Common lines modeling for reference free ab-initio reconstruction in cryo-EM," *J. Struct. Biol.*, vol. 200, no. 2, pp. 106–117, 2017.
[21] G. Pragier and Y. Shkolnisky, "A common lines approach for ab initio modeling of cyclically symmetric molecules," *Inverse Problems*, vol. 35, no. 12, p. 124005, 2019.
[22] M. Zehni, L. Donati, E. Soubies, Z. Zhao, and M. Unser, "Joint angular refinement and reconstruction for single-particle Cryo-EM," *IEEE Trans. Image Process.*, vol. 29, pp. 6151–6163, 2020.
[23] A. Singer and F. J. Sigworth, "Computational methods for single-particle electron cryomicroscopy," *Annu. Rev. Biomed. Data Sci.*, vol. 3, 2020.
[24] C. O. S. Sorzano, R. Marabini, A. Pascual-Montano, S. H. Scheres, and J. M. Carazo, "Optimization problems in electron microscopy of single particles," *Ann. Operations Res.*, vol. 148, no. 1, pp. 133–165, 2006.
[25] R. Henderson *et al.*, "Outcome of the first electron microscopy validation task force meeting," *Structure*, vol. 20, no. 2, pp. 205–214, 2012.
[26] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE Signal Process. Mag.*, vol. 37, no. 2, pp. 58–76, Mar. 2020.
[27] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[28] V. M. Panaretos *et al.* "On random tomography with unobservable projection angles," *Ann. Statist.*, vol. 37, no. 6 A, pp. 3272–3306, 2009.
[29] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, Nov. 2017.
[30] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.
[31] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.: Ser. B. (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
[32] F. J. Sigworth, "A maximum-likelihood approach to single-particle image refinement," *J. Struct. Biol.*, vol. 122, no. 3, pp. 328–339, 1998.
[33] F. J. Sigworth, P. C. Doerschuk, J.-M. Carazo, and S. H. Scheres, *An Introduction to Maximum-Likelihood Methods in Cryo-EM*. Elsevier, vol. 482, ch. 10, 2010, pp. 263–294.
[34] Z. Kam, "The reconstruction of structure from electron micrographs of randomly oriented particles," *J. Theor. Biol.*, vol. 82, pp. 15–39, Jan. 1980.
[35] 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*, vol. 36, no. 4, 2019.
[36] R. Rao, A. Moscovich, and A. Singer, "Wasserstein k-means for clustering tomographic projections," 2020, *arXiv:2010.09989*.
[37] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
[38] 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.

[39] T. Bepler, K. Kelley, A. J. Noble, and B. Berger, "Topaz-denoise: General deep denoising models for cryoEM and cryoET," *Nature Commun.*, vol. 11, no. 1, Oct. 2020.

[40] F. Wang *et al.*, "DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM," *J. Struct. Biol.*, vol. 195, no. 3, pp. 325–336, 2016.

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

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

[43] T. Wagner *et al.* "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM," *Commun. Biol.*, vol. 2, no. 1, p. 218, 2019.

[44] T. Bepler *et al.*, "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature Methods*, vol. 16, pp. 1153–1160, 2019.

[45] E. D. Zhong, T. Bepler, J. H. Davis, and B. Berger, "Reconstructing continuous distributions of 3D protein structure from cryo-EM images," in *Int. Conf. Learn. Representations*, 2020.

[46] 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," 2019, *arXiv:1911.08121*.

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

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

[49] 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 Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, Jun. 2018.

[50] A. Tewari *et al.*, "State of the art on neural rendering," 2020, *arXiv:2004.03805*.

[51] S. Tulsiani, A. A. Efros, and J. Malik, "Multi-view consistency as supervisory signal for learning shape and pose prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 2897–2905.

[52] M. Gadelha, S. Maji, and R. Wang, "3D shape induction from 2D views of multiple objects," in *Proc. Int. Conf. 3D Vis.*, Qingdao, China, 2017, pp. 402–411.

[53] A. Bora, E. Price, and A. G. Dimakis, "AmbientGAN: Generative models from lossy measurements." *Int. Conf. Learn. Representations*, p. 5, 2018.

[54] J. Frank, *Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*. Springer Science & Business Media, 2008.

[55] M. Vulović *et al.*, "Image formation modeling in cryo-electron microscopy," *J. Struct. Biol.*, vol. 183, no. 1, pp. 19–32, Jul. 2013.

[56] H. Rullga°rd, L.-G. öfverstedt, S. Masich, B. Daneholt, and O. öktem, "Simulation of transmission electron microscope images of biological specimens," *J. Microsc.*, vol. 243, no. 3, pp. 234–256, 2011.

[57] F. Natterer, *Math. Computerized Tomogr.. Soc. Ind. Appl. Math.*, Jan. 2001.

[58] W. van Aarle *et al.*, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[59] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[60] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 214–223.

[61] C. Villani, *Optimal Transport: Old and New*. Springer Science & Business Media, 2008.

[62] G. Peyré and M. Cuturi, "Computational optimal transport," *Found. Trends Mach. Learn.*, vol. 11, no. 5/6, pp. 355–607, 2019.

[63] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5767–5777.

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

[65] A. Paszke *et al.* "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.

[66] N. Biyani *et al.*, "Focus: The interface between data collection and data processing in cryo-em," *J. Struct. Biol.*, vol. 198, no. 2, pp. 124–133, 2017.

[67] R. D. Righetto, N. Biyani, J. Kowal, M. Chami, and H. Stahlberg, "Retrieving high-resolution information from disordered 2D crystals by single-particle cryo-EM," *Nature Commun.*, vol. 10, no. 1, pp. 1–10, 2019.

[68] A. Bartesaghi *et al.*, "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.

[69] E. F. Pettersen *et al.*, "UCSF chimera-a visualization system for exploratory research and analysis," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.

[70] A. Rohou and N. Grigorieff, "CTFFIND4: Fast and accurate defocus estimation from electron micrographs," *J. Struct. Biol.*, vol. 192, no. 2, pp. 216–221, 2015.

[71] A. Bartesaghi *et al.*, "Atomic resolution Cryo-EM structure of β-Galactosidase," *Structure*, vol. 26, no. 6, pp. 848–856, 2018.

[72] N. P. Nguyen, I. Ersoy, J. Gotberg, F. Bunyak, and T. A. White, "DRP-net: Automated particle picking in cryo-electron micrographs using deep regression," *BMC Bioinf.*, vol. 22, no. 1, Feb. 2021.

[73] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, 2019, pp. 4401–4410.

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

[75] S. H. Scheres, "A bayesian view on Cryo-EM structure determination," *J. Mol. Biol.*, vol. 415, no. 2, pp. 406–418, Jan. 2012.

[76] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Toward single particle reconstruction without particle picking: Breaking the detection limit," 2018, *arXiv:1810.00226*.

[77] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Multi-target detection with application to cryo-electron microscopy," *Inverse Problems*, vol. 35, no. 10, p. 104003, 2019.

[78] S. Helgason, *The Radon Transform*. Springer, 1980.