

A Stochastic Approximate Expectation-Maximization for Structure Determination Directly from Cryo-EM Micrographs

Shay Kreymer, Amit Singer, and Tamir Bendory

Abstract—A single-particle cryo-electron microscopy (cryo-EM) measurement, called a micrograph, consists of multiple two-dimensional tomographic projections of a three-dimensional molecular structure at unknown locations, taken under unknown viewing directions. All existing cryo-EM algorithmic pipelines first locate and extract the projection images, and then reconstruct the structure from the extracted images. However, if the molecular structure is small, the signal-to-noise ratio (SNR) of the data is very low, and thus accurate detection of projection images within the micrograph is challenging. Consequently, all standard techniques fail in low-SNR regimes. To recover molecular structures from measurements of low SNR, and in particular small molecular structures, we devise a stochastic approximate expectation-maximization algorithm to estimate the three-dimensional structure directly from the micrograph, bypassing locating the projection images. We corroborate our computational scheme with numerical experiments, and present successful structure recoveries from simulated noisy measurements.

Index Terms—expectation-maximization, cryo-electron microscopy, three-dimensional tomographic reconstruction.

I. INTRODUCTION

CRYO-ELECTRON MICROSCOPY (cryo-EM) is an increasingly popular technology in structural biology for elucidating the 3-D structure of biomolecules [1], [2]. In a cryo-EM experiment, individual copies of the target biomolecule are dispersed in a thin layer of vitreous ice. Then, a 2-D tomographic projection image, called a *micrograph*, is produced by an electron microscope [3]. A micrograph contains multiple tomographic projections of individual molecules, taken from unknown viewing directions and placed at unknown locations. Section II introduces the formation model of a micrograph in detail. The goal is to recover a 3-D molecular structure from a set of micrographs [4], [5], [6], [7], [8].

S. Kreymer and T. Bendory are with the School of Electrical Engineering of Tel Aviv University, Tel Aviv, Israel, e-mail: shaykreymer@mail.tau.ac.il, bendory@tauex.tau.ac.il. A. Singer is with the Department of Mathematics and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA (amits@math.princeton.edu). S.K. is supported by the Yitzhak and Chaya Weinstein Research Institute for Signal Processing. A.S. is supported in part by AFOSR under Grant FA9550-20-1-0266, in part by Simons Foundation Math+X Investigator Award, in part by NSF under Grant DMS-2009753, and in part by NIH/NIGMS under Grant R01GM136780-01. T.B. is supported in part by BSF under Grant 2020159, in part by NSF-BSF under Grant 2019752, and in part by ISF under Grant 1924/21.

The prevalent cryo-EM computational paradigm splits the reconstruction process into two main stages. The first stage consists of identifying and extracting the projection images from the micrographs. This stage is called particle picking, see for example [9], [10], [11], [12]. In the second stage, the 3-D structure is reconstructed from the extracted projection images. Clearly, the quality of the reconstruction depends on the quality of the particle picking stage, which in turn depends heavily on the signal-to-noise ratio (SNR) of the micrograph [13]. Therefore, this approach fails when the SNR of the micrograph is very low. In particular, it fails for small molecular structures that induce low SNR because fewer electrons carry information. The detection threshold has been recognized early on as a central limiting factor by the cryo-EM community; it was suggested that particle picking is impossible for molecules with molecular weight below ~ 40 kDa [14], [15]. Indeed, to date, the vast majority of biomolecules whose structures have been determined using cryo-EM have molecular weights not smaller than 100 kDa. Recovering smaller molecular structures is of crucial importance in cryo-EM, and is an active focal point of research endeavors in the field [16], [17], [18], [19], [20], [21], [22], [23], [24], [25].

The failure of the current cryo-EM computational paradigm to recover 3-D structures from low SNR micrographs can be understood through the lenses of classical estimation theory. Assume the 3-D volume is represented by M parameters. Each particle projection is associated with five pose parameters—the 3-D rotation and the 2-D location. Thus, if we wish to jointly estimate the 3-D structure and the pose parameters of the T projection images, like in older cryo-EM algorithms [26], the number of parameters to be estimated is $M + 5T$, namely, grows linearly with the number of particle projections. In this case, it is well-known that the existence of a consistent estimator is not guaranteed; see for example the celebrated “Neyman-Scott paradox” [27] and the multi-image alignment problem [28]. Current approaches in cryo-EM can be thought of as “hybrid” in the sense that they estimate the locations of the particle projections in the particle picking stage (overall $2T$ parameters), and marginalize over the rotations (as well as over small translations relative to the estimated locations), see for example [29]. Thus, the number of parameters is $M + 2T$, which still scales linearly with the number of projections. Indeed, as discussed above, this strategy is not consistent

when the SNR is very low since particle picking fails. In this paper, we follow [30] and aim to marginalize over all nuisance variables—the locations and rotations. In this case, the number of parameters to be estimated is fixed and thus, given enough data, designing a consistent estimator might be feasible. Therefore, from an estimation theory view point, recovery in low SNR environments (and thus of small molecular structures) is potentially within reach.

The authors of [30] proposed to recover the 3-D volume directly from the micrographs using autocorrelation analysis, but their reported reconstructions were limited to low resolution. In this paper, we propose an alternative computational scheme for high resolution structure reconstruction based on the expectation-maximization (EM) algorithm [31]. EM is an algorithm for finding a local maximum of a likelihood function with nuisance variables. It is widely used in many machine learning and statistics tasks, with applications to parameter estimation [32], mixture models [33], deep belief networks [34], and independent component analysis [35], to name but a few. The EM algorithm was introduced to the cryo-EM community in [36], and is by now the most popular method for 3-D recovery from picked particles [29], where the 3-D rotations, but not the 2-D locations, are treated as nuisance variables.

Specifically, in order to recover the molecular structure directly from the micrograph, we aim to develop an EM algorithm that marginalizes over both 2-D translations and 3-D rotations. However, as we show in Section III-A, a direct application of EM is computationally intractable for our model since the number of possible projection locations in the micrograph grows quickly with the micrograph size. Therefore, based on [37], [38], we develop an EM algorithm that maximizes an approximation of the likelihood function. The computational complexity of the algorithm is linear in the micrograph size. To further accelerate the algorithm, we apply a stochastic variant of the approximate EM algorithm that linearly decreases the computational complexity and memory requirement of each iteration (at the potential cost of additional iterations); see Section III-C for further details.

In Section IV, we demonstrate that the proposed approximate EM can accurately estimate molecular structures from simulated data in various levels of noise, outperforming the autocorrelation analysis of [30]. This is inline with previous works on simpler 1-D and 2-D models [38], [37], [39], [40]. The SNR in the results of Section IV is around 1 (namely, the noise level and the power of the projection images are of the same order) due to the computational load of the algorithm (see Section III-E). Section V outlines potential strategies to alleviating the computational complexity of our method so we can apply it to micrographs of extremely low SNR, as expected for reconstruction of small molecular structures. Crucially, our results do not depend, empirically, on the initial point of the algorithm. This suggests that our algorithm is less prone to model bias, where the output of the algorithm is biased by the initial model [6]. Model bias has been recognized as a major pitfall of current cryo-EM algorithms [41].

II. MEASUREMENT FORMATION MODEL

Our micrograph formation model follows the formulation of [30]. Let $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ represent the 3-D electrostatic potential of the molecule to be estimated. We refer to f as the volume. A 2-D tomographic projection of the volume is a line integral, given by

$$I_\omega(x, y) := PR_\omega f = \int_{-\infty}^{\infty} (R_\omega f)(x, y, z) dz, \quad (1)$$

where the operator R_ω rotates the volume by $\omega \in SO(3)$ and P is the tomographic projection operator. The micrograph consists of T tomographic projections, taken from different viewing directions $\{\omega_t\}_{t=1}^T \in SO(3)$, centered at different positions $\{(x_t, y_t)\}_{t=1}^T$,

$$\begin{aligned} \mathcal{I}(x, y) &= \int_{-\infty}^{\infty} \sum_{t=1}^T (R_{\omega_t} f)(x - x_t, y - y_t, z) dz + \varepsilon(x, y) \\ &= \sum_{t=1}^T \int_{-\infty}^{\infty} (R_{\omega_t} f)(x - x_t, y - y_t, z) dz + \varepsilon(x, y) \\ &= \sum_{t=1}^T I_{\omega_t}(x - x_t, y - y_t) + \varepsilon(x, y), \end{aligned} \quad (2)$$

where $\varepsilon(x, y)$ is assumed to be i.i.d. white Gaussian noise with zero mean and variance σ^2 .

We further assume that the micrograph is discretized on a Cartesian grid, the particle projections are centered on the grid, and each projection is of size $L \times L$ pixels; the projection size, L , is assumed to be known. We denote the indices on the grid by $\vec{\ell} = (\ell_x, \ell_y) \in \mathbb{Z}^2$. Thus, our micrograph model $\mathcal{I} \in \mathbb{R}^{N \times N}$ reads

$$\mathcal{I}[\vec{\ell}] = \sum_{t=1}^T I_{\omega_t}[\ell_x - \ell_{x_t}, \ell_y - \ell_{y_t}] + \varepsilon[\vec{\ell}]. \quad (3)$$

The goal is to estimate f from several micrographs while the rotations, translations, and the number of projections are unknown. Importantly, it is possible to reconstruct the target volume only up to a 3-D rotation, a 3-D translation, and a reflection. Similar mathematical models were thoroughly studied in previous works for one- and two-dimensional setups [42], [38], [43], [44], [45], [46], [37]. Fig. 1 presents an example of a noisy micrograph \mathcal{I} at different SNRs, where

$$\text{SNR} := \frac{\mathbb{E} [\|I_{\omega_t}\|_F^2]}{L^2 \sigma^2} \approx \frac{\frac{1}{T} \sum_{t=1}^T \|I_{\omega_t}\|_F^2}{L^2 \sigma^2}, \quad (4)$$

where $\|\cdot\|_F$ is the Frobenius norm. Section V discusses how to include additional aspects of the cryo-EM reconstruction problem in the proposed technique, such as the effect of the contrast transfer function (CTF) [47], colored noise, and non-uniform distribution of the rotations of the particles over $SO(3)$.

Following previous works [43], [44], we also assume that each translation is separated by at least a full projection length, L , from its neighbors, in both the horizontal and vertical axes. Explicitly,

$$|\ell_{x_t} - \ell_{x_s}| \geq 2L - 1 \text{ and } |\ell_{y_t} - \ell_{y_s}| \geq 2L - 1, \text{ for all } t \neq s. \quad (5)$$

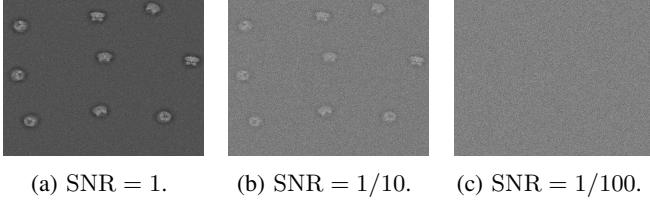


Fig. 1: Three micrographs at different SNRs. Each measurement contains $T = 9$ projections of the target volume. We focus on the low SNR regime, in which the 2-D locations and 3-D rotations of the projection images cannot be estimated reliably.

In Section IV-C, we discuss the implications of mitigating this constraint by allowing the projection images to be arbitrarily close.

A. Volume formation model

Let f be the Coulomb potential representing the molecule, which is smooth and real-valued. Its 3-D Fourier transform, \hat{f} , is finitely expanded by

$$\hat{f}(ck, \theta, \varphi) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} \sum_{s=1}^{S(\ell)} x_{\ell,m,s} Y_{\ell}^m(\theta, \varphi) j_{\ell,s}(k), \quad k \leq 1, \quad (6)$$

where c is the bandlimit, $S(\ell)$ is determined using the Nyquist criterion as described in [48], $j_{\ell,s}$ is the normalized spherical Bessel function, given by

$$j_{\ell,s}(k) = \frac{4}{|j_{\ell+1}(u_{\ell,s})|} j_{\ell}(u_{\ell,s} k), \quad (7)$$

where j_{ℓ} is the spherical Bessel function of order ℓ , $u_{\ell,s}$ is the s -th positive zero of j_{ℓ} , and Y_{ℓ}^m is the complex spherical harmonic, defined by

$$Y_{\ell}^m(\theta, \varphi) := \sqrt{\frac{2\ell+1}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_{\ell}^m(\cos \theta) e^{im\varphi}, \quad (8)$$

where P_{ℓ}^m are the associated Legendre polynomials with the Condon-Shortley phase. We set $c = 1/2$ for sampling at the Nyquist rate [49]. Under this model, we aim to estimate the expansion coefficients $x_{\ell,m,s}$ that describe f . Since f is real-valued, \hat{f} is conjugate-symmetric and thus the expansion coefficients satisfy $x_{\ell,-m,s} = (-1)^{\ell+m} \overline{x_{\ell,m,s}}$.

Let I_{ω} denote the tomographic projection obtained from viewing direction $\omega \in SO(3)$. By the Fourier projection-slice theorem (see, e.g., [50]), its 2-D Fourier transform is given by

$$\widehat{I}_{\omega}(ck, \varphi) = \sum_{\ell, m, m', s} x_{\ell, m, s} D_{m', m}^{\ell}(\omega) Y_{\ell}^{m'}\left(\frac{\pi}{2}, \varphi\right) j_{\ell, s}(k), \quad (9)$$

where $D_{m', m}^{\ell}(\omega)$ is a Wigner-D matrix. Note that the projection images in the micrograph model (3) are expressed in the space domain, whereas (9) is expressed in Fourier space. To bridge this gap, we use the prolate spheroidal wave functions (PSWFs) [51] as explained next.

B. Expressing the projection image in space domain using the prolate spheroidal wave functions

The PSWFs are eigenfunctions of the truncated Fourier transform:

$$\alpha_{N,n} \psi_{N,n}(\mathbf{k}) = \int_{\|\mathbf{r}\|_2 \leq 1} \psi_{N,n}(\mathbf{r}) e^{ic(\mathbf{r}, \mathbf{k})} d\mathbf{r}, \quad (10)$$

where c is the bandlimit of the eigenfunction $\psi_{N,n}$. The eigenfunctions are orthonormal on the unit disk $\mathbb{D} := \{\mathbf{r} \in \mathbb{R}^2, \|\mathbf{r}\|_2 \leq 1\}$, and they are the most energy concentrated among all c -bandlimited functions on \mathbb{D} , i.e., they satisfy

$$\psi_{N,n}(\mathbf{r}) = \arg \max_{\psi} \frac{\|\psi(\mathbf{r})\|_{\mathcal{L}^2(\mathbb{D})}}{\|\psi(\mathbf{r})\|_{\mathcal{L}^2(\mathbb{R}^2)}}. \quad (11)$$

Explicitly, the PSWFs are given in polar coordinates by

$$\psi_{N,n}(r, \varphi) = \begin{cases} \frac{1}{\sqrt{8\pi}} \alpha_{N,n} R_{N,n}(r) e^{iN\varphi}, & r \leq 1, \\ 0, & r > 1, \end{cases} \quad (12)$$

where the range of $N \in \mathbb{Z}$, $n \in \mathbb{N} \cup \{0\}$ is determined by [52, Eq. (8)], the $R_{N,n}$ are a family of real, one-dimensional functions, defined explicitly in [52, Eq. (66)], and $\alpha_{N,n}$ is the eigenvalue corresponding to the (N, n) -th PSWF (10). From (12), we can also see that the PSWFs are steerable [53]—rotating the image is equivalent to multiplying the eigenfunction $\psi_{N,n}$ by a phase dependent only on the rotation and the index N . The indices N and n are referred to, respectively, as the angular index and the radial index.

We may expand the projection (9) in Fourier domain using the PSWFs:

$$\widehat{I}_{\omega}(ck, \theta) = \sum_{N, n} b_{N,n}(\omega) \psi_{N,n}(k, \theta). \quad (13)$$

The coefficients are given by

$$\begin{aligned} b_{N,n}(\omega) &= \frac{4}{\sqrt{2\pi} |\alpha_{N,n}|^2} \int_0^{2\pi} \int_0^1 \widehat{I}_{\omega}(ck, \theta) R_{N,n}(k) e^{-iN\theta} k dk d\theta \\ &= \sum_{l, m', m, s} x_{\ell, m, s} \frac{\sqrt{8\pi}}{\alpha_{N,n}} Y_{\ell}^{m'}\left(\frac{\pi}{2}, 0\right) D_{m', m}^{\ell}(\omega) \\ &\quad \times \frac{1}{2\pi} \int_0^{2\pi} e^{i(m'-N)\theta} d\theta \int_0^1 j_{\ell, s}(k) R_{N,n}(k) k dk \\ &= \sum_{\ell \geq |N|} \sum_{m, s} x_{\ell, m, s} D_{N, m}^{\ell}(\omega) \beta_{\ell, s; N, n}, \end{aligned} \quad (14)$$

where

$$\beta_{\ell, s; N, n} := \frac{\sqrt{8\pi}}{\alpha_{N,n}} Y_{\ell}^N(\pi/2, 0) \int_0^1 j_{\ell, s}(k) R_{N,n}(k) k dk \quad (15)$$

for $\ell \geq |N|$, and 0 otherwise.

Using the property (10), we can express the projection in space domain as (16), where n_{\max} is chosen according to [52, Eq. (8)], and $\widehat{\beta}_{\ell, s; N, n} := \left(\frac{c}{2\pi}\right)^2 \alpha_{N,n} \beta_{\ell, s; N, n}$.

Using (3) and (16), we can now relate the parameters of the volume directly with the micrograph, bypassing the particle picking stage. In the following section, we present our scheme for estimating the coefficients $x_{\ell, m, s}$ from a set of noisy micrographs.

$$\begin{aligned}
I_\omega(r, \varphi) &= \left(\frac{c}{2\pi}\right)^2 \sum_{\ell=0}^{\ell_{\max}} \sum_{N=-\ell}^{\ell} \sum_{m=-\ell}^{\ell} \sum_{n=0}^{n_{\max}(N)} \sum_{s=1}^{S(\ell)} x_{\ell,m,s} \alpha_{N,n} \beta_{\ell,s;N,n} D_{N,m}^\ell(\omega) \psi_{N,n}(r, \varphi) \\
&:= \sum_{\ell=0}^{\ell_{\max}} \sum_{N=-\ell}^{\ell} \sum_{m=-\ell}^{\ell} \sum_{n=0}^{n_{\max}(N)} \sum_{s=1}^{S(\ell)} x_{\ell,m,s} \hat{\beta}_{\ell,s;N,n} D_{N,m}^\ell(\omega) \psi_{N,n}(r, \varphi)
\end{aligned} \tag{16}$$

III. AN APPROXIMATE EXPECTATION-MAXIMIZATION (EM) ALGORITHM FOR CRYO-EM

A. Approximate EM

The EM algorithm estimates the maximum of a likelihood function by iteratively applying the expectation (E) and the maximization (M) steps [31]. For the model (3), given a measurement \mathcal{I} , the maximum likelihood estimator (MLE) is the maximizer of $p(\mathcal{I}; x)$ for the vector of coefficients x (6). The 2-D translations and 3-D rotations associated with the projection images within the micrograph are treated in our analysis as nuisance variables. In the EM terminology, they are dubbed unobserved or latent variables.

In the E-step of the $(k+1)$ -th iteration of the EM algorithm, one computes $Q(x; x_k)$ —the expectation of the complete log-likelihood function, where x_k is the current estimate of the parameters and the expectation is taken over all admissible configurations of translations and rotations. However, for our model, the number of possible translations in the micrograph grows quickly with the micrograph size, N^2 . Consequently, a direct application of EM is computationally intractable. Instead, we follow [38], [37] and partition the micrograph \mathcal{I} into $N_{\text{patches}} = (N/L)^2$ non-overlapping patches $\{\mathcal{I}_p\}_{p=1}^{N_{\text{patches}}}$; each patch is of the size of a projection image $L \times L$. In EM terminology, the patches are dubbed the observed data. The separation condition (5) implies that each patch \mathcal{I}_p can contain either no projection, a full projection, or part of a projection; overall there are $(2L)^2$ possibilities (disregarding rotations). We denote the distribution of translations within a patch by $\rho[\vec{\ell}]$, and require that

$$\sum_{\vec{\ell} \in \mathbb{L}} \rho[\vec{\ell}] = 1, \quad \rho[\vec{\ell}] \geq 0 \text{ for } \vec{\ell} \in \mathbb{L}, \tag{17}$$

where $\mathbb{L} := \{0, \dots, 2L-1\}^2$. Thus, instead of aiming to maximize the likelihood function $p(\mathcal{I}; x)$, we wish to maximize its surrogate $\prod_{p=1}^{N_{\text{patches}}} p(\mathcal{I}_p; x, \rho)$ using EM. Since the number of possible translations in each patch is independent of the micrograph size, applying EM is now tractable.

Specifically, each patch is modeled by

$$\mathcal{I}_p = CT_{\vec{\ell}_p} ZPR_{\omega_p} f + \varepsilon_p, \quad \varepsilon_p \sim \mathcal{N}(0, \sigma^2 I_{L \times L}), \tag{18}$$

where the operator R_{ω_p} rotates the volume f by $\omega_p \in SO(3)$, and the operator P projects the volume into 2-D so that $PR_{\omega_p} f$ is given by I_{ω_p} (16). The operator Z zero-pads L entries to the right and to the bottom of a projection, and $T_{\vec{\ell}_p}$

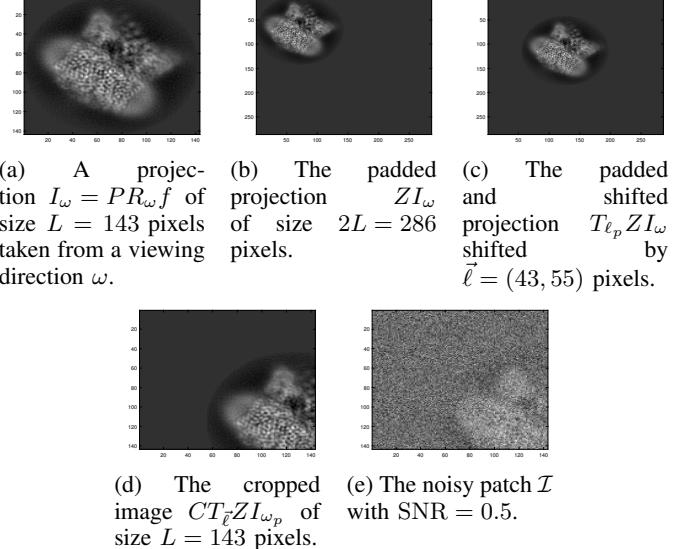


Fig. 2: An illustration of the patch generation model described in (18).

circularly shifts the zero-padded image by $\vec{\ell}_p = (\ell_{x_p}, \ell_{y_p}) \in \mathbb{L}$ positions, that is,

$$\begin{aligned}
&(T_{\vec{\ell}_p} ZPR_{\omega_p} f)[i, j] \\
&= (ZPR_{\omega_p} f)[(i + \ell_{x_p}) \bmod 2L, (j + \ell_{y_p}) \bmod 2L].
\end{aligned} \tag{19}$$

The operator C then crops the first L entries in the vertical and horizontal axes, and the result is further corrupted by additive white Gaussian noise with zero mean and variance σ^2 . The generative model of a patch is illustrated in Fig. 2.

Since in the E-step the algorithm assigns probabilities to rotations, the space of rotations must be discretized. We denote the set of K discrete rotations by $\Omega^K \subseteq SO(3)$, such that $\omega_p \stackrel{\text{i.i.d.}}{\sim} \text{Unif}\{\Omega^K\}$; see Section IV for details. Higher K provides higher accuracy at the cost of running time.

B. EM iterations

1) The E-step: Given a current estimate of the expansion coefficients x_k and the distribution of translations ρ_k , in the E-step, our algorithm calculates the expected log-likelihood

$$Q(x, \rho | x_k, \rho_k) = \mathbb{E} [\log \mathcal{L} | \mathcal{I}_1, \dots, \mathcal{I}_{N_{\text{patches}}}; x_k, \rho_k], \tag{20}$$

where

$$\mathcal{L} = \prod_{p=1}^{N_{\text{patches}}} p(\mathcal{I}_p, \vec{\ell}_p, \omega_p; x, \rho), \tag{21}$$

which is a surrogate of the computationally intractable complete likelihood function. The expectation is taken over the possible translations and rotations, to achieve

$$\begin{aligned} & Q(x, \rho | x_k, \rho_k) \\ &= \sum_{p=1}^{N_{\text{patches}}} \sum_{\vec{\ell} \in \mathbb{L}} \sum_{\omega \in \Omega^K} p(\vec{\ell}_p, \omega_p | \mathcal{I}_p; x_k, \rho_k) \log p(\mathcal{I}_p, \vec{\ell}_p, \omega_p | x, \rho). \end{aligned} \quad (22)$$

Applying Bayes' rule, we have that

$$p(\vec{\ell}_p, \omega_p | \mathcal{I}_p; x_k, \rho_k) = \frac{p(\mathcal{I}_p | \vec{\ell}_p, \omega_p; x_k, \rho_k) p(\vec{\ell}_p, \omega_p | x_k, \rho_k)}{\sum_{\vec{\ell}' \in \mathbb{L}} \sum_{\omega' \in \Omega^K} p(\mathcal{I}_p | \vec{\ell}', \omega'; x_k, \rho_k) p(\vec{\ell}', \omega' | x_k, \rho_k)}, \quad (23)$$

which is just the normalized likelihood function

$$p(\mathcal{I}_p | \vec{\ell}_p, \omega_p; x_k, \rho_k) \propto \exp \left(-\frac{\|\mathcal{I}_p - CT_{\vec{\ell}_p} Z PR_{\omega_p} f\|_F^2}{2\sigma^2} \right), \quad (24)$$

with the normalization $\sum_{\vec{\ell} \in \mathbb{L}} \sum_{\omega \in \Omega^K} p(\mathcal{I}_p | \vec{\ell}, \omega_p; x_k, \rho_k) = 1$, weighted by the prior distribution $p(\vec{\ell}_p, \omega_p | x_k, \rho_k) = \frac{\rho_k[\vec{\ell}_p]}{K}$.

Thus, we can rewrite the expected log-likelihood function (22), up to a constant, as:

$$\begin{aligned} Q(x, \rho | x_k, \rho_k) &= \sum_{p=1}^{N_{\text{patches}}} \sum_{\vec{\ell} \in \mathbb{L}} \sum_{\omega \in \Omega^K} p(\mathcal{I}_p | \vec{\ell}, \omega, x_k) \rho_k[\vec{\ell}] \\ &\times \left(\log p(\mathcal{I}_p | \vec{\ell}, \omega, x) + \log \rho[\vec{\ell}] \right). \end{aligned} \quad (25)$$

2) *The M-step*: The M-step updates x and ρ by maximizing $Q(x, \rho | x_k, \rho_k)$ under the constraint that ρ is a distribution function:

$$\begin{aligned} x_{k+1}, \rho_{k+1} &= \arg \max_{x, \rho} Q(x, \rho | x_k, \rho_k) \\ \text{s.t. } & \sum_{\vec{\ell} \in \mathbb{L}} \rho[\vec{\ell}] = 1, \quad \rho[\vec{\ell}] \geq 0 \text{ for } \vec{\ell} \in \mathbb{L}. \end{aligned} \quad (26)$$

The constrained maximization of (26) can be achieved by maximizing the Lagrangian

$$\mathcal{L}(x, \rho, \lambda) = Q(x, \rho | x_k, \rho_k) + \lambda \left(1 - \sum_{\vec{\ell} \in \mathbb{L}} \rho[\vec{\ell}] \right), \quad (27)$$

where λ is the Lagrange multiplier. As we will see next, the constraint of (17) is automatically satisfied at the maximum of the Lagrangian.

Since $Q(x, \rho | x_k, \rho_k)$ is additively separable for x and ρ , we maximize $\mathcal{L}(x, \rho, \lambda)$ with respect to x and ρ separately. At the maximum of $\mathcal{L}(x, \rho, \lambda)$, we have

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial x_{\ell, m, s}} \\ &= \sum_{p=1}^{N_{\text{patches}}} \sum_{\vec{\ell} \in \mathbb{L}} \sum_{\omega \in \Omega^K} p(\mathcal{I}_p | \vec{\ell}, \omega, x_k) \rho_k[\vec{\ell}] \frac{\partial \log p(\mathcal{I}_p | \vec{\ell}, \omega, x)}{\partial x_{\ell, m, s}}, \end{aligned} \quad (28)$$

Algorithm 1 A stochastic approximate EM for cryo-EM

Input: measurement \mathcal{I} partitioned to N_{patches} patches; patch size L ; parameter K (number of discretized rotations); noise variance σ^2 ; initial guesses x_0 and ρ_0 ; stopping parameter ϵ ; maximal number of iterations N_{it} ; stochastic factor $0 < S \leq 1$.

Output: an estimate of x and ρ .

```

Set  $k \rightarrow 0$ ;
Set  $\mathfrak{I}_{0;S} : \lfloor SN_{\text{patches}} \rfloor$  patches, drawn from a uniform distribution;
Calculate  $p(\mathfrak{I}_{0;S} | \vec{\ell}, \omega; x_0, \rho_0)$  according to (24) for the set  $\mathfrak{I}_{0;S}$ , all  $\omega \in \Omega^K$ , and all  $\vec{\ell} \in \mathbb{L}$ ;
Calculate  $Q_0$  according to (25) for patches in  $\mathfrak{I}_{0;S}$ , and set  $Q_{-1} \rightarrow -\infty$ ;
while  $(Q_k - Q_{k-1}) > \epsilon$  and  $k < N_{\text{it}}$  do
    Update  $x_{k+1}$  according to (29);
    Update  $\rho_{k+1}$  according to (37);
    Set  $\mathfrak{I}_{k+1;S} : \lfloor SN_{\text{patches}} \rfloor$  patches, drawn from a uniform distribution;
    Calculate  $p(\mathfrak{I}_{k+1;S} | \vec{\ell}, \omega; x_{k+1}, \rho_{k+1})$  according to (24) for the set  $\mathfrak{I}_{k+1;S}$ , all  $\omega \in \Omega^K$ , and all  $\vec{\ell} \in \mathbb{L}$ ;
    Calculate  $Q_{k+1}$  according to (25) for patches in  $\mathfrak{I}_{k+1;S}$ ;
    Set  $k \rightarrow k + 1$ ;
end while
return  $x_k, \rho_k$ 
```

resulting in a set of linear equations

$$y_k = A_k x_{k+1}, \quad (29)$$

where y_k is given by (30) and A_k is given by (31), and \odot is the Hadamard product. Notably, we can rewrite (30) as (32), where $q(p, \vec{\ell}, \omega; \ell, m, s)$ is given by (33). Furthermore, we can rewrite A_k from (31) as (34), where $g(\vec{\ell}, \omega; \ell, m, s; \tilde{\ell}, \tilde{m}, \tilde{s})$ is given by (35). The term g is independent of the data $\{\mathcal{I}_p\}_{p=1}^{N_{\text{patches}}}$ and of the current estimate x_k , and thus can be precomputed once before the EM iterations; see Section III-E for a discussion about the computational complexity of the EM algorithm.

In order to update ρ , we maximize $\mathcal{L}(x, \rho, \lambda)$ with respect to ρ :

$$0 = \frac{\partial \mathcal{L}}{\partial \rho[\vec{\ell}]} = \sum_{p=1}^{N_{\text{patches}}} \sum_{\omega \in \Omega^K} p(\mathcal{I}_p | \vec{\ell}, \omega, x_k) \rho_k[\vec{\ell}] \frac{1}{\rho[\vec{\ell}]} - \lambda, \quad (36)$$

for $\vec{\ell} \in \mathbb{L}$. We thus obtain the update rule for ρ as

$$\rho[\vec{\ell}] = \frac{1}{\lambda} \sum_{p=1}^{N_{\text{patches}}} \sum_{\omega \in \Omega^K} p(\mathcal{I}_p | \vec{\ell}, \omega, x_k) \rho_k[\vec{\ell}], \quad (37)$$

and $\lambda = N_{\text{patches}}$ from the normalization (17).

C. Stochastic approximate EM

In order to alleviate the computational burden of the approximate EM scheme (see Section III-E), we apply a stochastic variant. We focus on incremental EM, first introduced in [54]. At each iteration, the incremental EM algorithm applies the E-step to a minibatch of the observed data; the parameters of the problem are updated using the standard M-step. In particular, for the model (3), at each iteration, we

$$\begin{aligned}
y_k[\ell, m, s] &:= \sum_{p=1}^{N_{\text{patches}}} \sum_{\omega \in \Omega^K} \sum_{\tilde{\ell} \in \mathbb{L}} p(\mathcal{I}_p | \tilde{\ell}, \omega, x_k) \rho_k[\tilde{\ell}] \sum_{N=-\ell}^{\ell} \sum_{n=0}^{n_{\max}(N)} \widehat{\beta}_{\ell, s; N, n} D_{N, m}^{\ell}(\omega) \\
&\times \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \mathcal{I}_p[i, j] (Z\psi_{N, n})[(i + \ell_x) \bmod 2L, (j + \ell_y) \bmod 2L]
\end{aligned} \tag{30}$$

$$\begin{aligned}
A_k[\ell, m, s; \tilde{\ell}, \tilde{m}, \tilde{s}] &:= \sum_{p=1}^{N_{\text{patches}}} \sum_{\omega \in \Omega^K} \sum_{\tilde{\ell} \in \mathbb{L}} p(\mathcal{I}_p | \tilde{\ell}, \omega, x_k) \rho_k[\tilde{\ell}] \sum_{\tilde{N}=-\tilde{\ell}}^{\tilde{\ell}} \sum_{\tilde{n}=0}^{n_{\max}(\tilde{N})} \widehat{\beta}_{\tilde{\ell}, \tilde{s}; \tilde{N}, \tilde{n}} D_{\tilde{N}, \tilde{m}}^{\tilde{\ell}}(\omega) \sum_{N=-\ell}^{\ell} \sum_{n=0}^{n_{\max}(N)} \widehat{\beta}_{\ell, s; N, n} D_{N, m}^{\ell}(\omega) \\
&\times \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (Z\psi_{\tilde{N}, \tilde{n}} \odot Z\psi_{N, n})[(i + \ell_x) \bmod 2L, (j + \ell_y) \bmod 2L]
\end{aligned} \tag{31}$$

$$y_k[\ell, m, s] = \sum_{p=1}^{N_{\text{patches}}} \sum_{\omega \in \Omega^K} \sum_{\tilde{\ell} \in \mathbb{L}} p(\mathcal{I}_p | \tilde{\ell}, \omega, x_k) \rho_k[\tilde{\ell}] q(p, \tilde{\ell}, \omega; \ell, m, s) \tag{32}$$

$$q(p, \tilde{\ell}, \omega; \ell, m, s) := \sum_{N=-\ell}^{\ell} \sum_{n=0}^{n_{\max}(N)} \widehat{\beta}_{\ell, s; N, n} D_{N, m}^{\ell}(\omega) \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \mathcal{I}_p[i, j] (Z\psi_{N, n})[(i + \ell_x) \bmod 2L, (j + \ell_y) \bmod 2L] \tag{33}$$

$$A_k[\ell, m, s; \tilde{\ell}, \tilde{m}, \tilde{s}] = \sum_{\omega \in \Omega^K} \sum_{\tilde{\ell} \in \mathbb{L}} \left\{ \sum_{p=1}^{N_{\text{patches}}} p(\mathcal{I}_p | \tilde{\ell}, \omega, x_k) \right\} \rho_k[\tilde{\ell}] g(\tilde{\ell}, \omega; \ell, m, s; \tilde{\ell}, \tilde{m}, \tilde{s}) \tag{34}$$

$$\begin{aligned}
g(\tilde{\ell}, \omega; \ell, m, s; \tilde{\ell}, \tilde{m}, \tilde{s}) &:= \sum_{\tilde{N}=-\tilde{\ell}}^{\tilde{\ell}} \sum_{\tilde{n}=0}^{n_{\max}(\tilde{N})} \widehat{\beta}_{\tilde{\ell}, \tilde{s}; \tilde{N}, \tilde{n}} D_{\tilde{N}, \tilde{m}}^{\tilde{\ell}}(\omega) \sum_{N=-\ell}^{\ell} \sum_{n=0}^{n_{\max}(N)} \widehat{\beta}_{\ell, s; N, n} D_{N, m}^{\ell}(\omega) \\
&\times \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (Z\psi_{\tilde{N}, \tilde{n}} \odot Z\psi_{N, n})[(i + \ell_x) \bmod 2L, (j + \ell_y) \bmod 2L]
\end{aligned} \tag{35}$$

choose $\lfloor SN_{\text{patches}} \rfloor$ patches drawn uniformly from the set of patches, where $0 < S \leq 1$. We denote the set of $\lfloor SN_{\text{patches}} \rfloor$ patches by $\mathcal{J}_{k; S}$, where k is the iteration index. A small S will result in faster and less memory consuming iterations, at the possible cost of additional iterations. In our numerical experiments in Section IV we set $S = 0.05$.

The stochastic approximate EM algorithm is summarized in Algorithm 1.

D. Frequency marching

As another means to reduce the computational complexity of our scheme, we adopt the frequency marching concept, previously applied to cryo-EM tasks [55], [29], [38]. We start our stochastic EM procedure (Algorithm 1) with a low target frequency (small ℓ_{\max}) estimate. When the algorithm is terminated, we use the low frequency estimate as an initial guess for the stochastic EM procedure with a higher target frequency, and continue to gradually increase the frequencies. This way the lion's share of iterations is done over the low frequency estimates. This is crucial since the computational complexity of the algorithm strongly depends on the frequency ℓ_{\max} : see the next section.

E. Complexity analysis

The computational complexity of the approximate EM algorithm depends mainly on the computational complexity of forming and solving the linear system of equations (29) at each iteration.

We start with analyzing the complexity of forming the matrix A_k (31). The computational complexity of computing a single entry of $p(\mathcal{I}_p | \tilde{\ell}, \omega, x_k)$, given by (24), is $O(L^2)$. Recall that we have $N_{\text{patches}} = (N/L)^2$, $2L^2$ possible shifts, and K discrete rotations. In addition, the summations over the indices $\ell, m, s, N, n, \tilde{\ell}, \tilde{m}, \tilde{s}, \tilde{N}$, and \tilde{n} are of $O(\ell_{\max})$ operations each. Consequently, the total number of coefficients $x_{\ell, m, s}$ is $O(\ell_{\max}^3)$, and the total number of matrix entries is $O(\ell_{\max}^6)$. As such, the computational complexity of computing each entry of A_k brute force is $O\left(\frac{N^2}{L^2} KL^2 \ell_{\max}^4 L^2\right) = O(N^2 KL^2 \ell_{\max}^4)$, and the total complexity of computing the matrix is $O(N^2 KL^2 \ell_{\max}^{10})$. However, we recall that $p(\mathcal{I}_p | \tilde{\ell}, \omega, x_k)$ can be computed once at the beginning of each iteration with complexity of $O\left(\frac{N^2}{L^2} KL^2 L^2\right) = O(N^2 KL^2)$. In addition, as was noted in (34), we can massively reduce the computational complexity by precomputing the function $g(\tilde{\ell}, \omega; \ell, m, s; \tilde{\ell}, \tilde{m}, \tilde{s})$

just once at the beginning of the algorithm with complexity of $O(\ell_{\max}^{10} L^4 K)$. That is, the total computational complexity of computing the matrix A_k is

$$\begin{aligned} & O \left(L^4 K \ell_{\max}^{10} + V \left(N^2 K L^2 + \frac{N^2}{L^2} L^2 K \ell_{\max}^6 \right) \right) \\ &= O \left(\ell_{L^4 K \max}^{10} + V (N^2 K L^2 + N^2 K \ell_{\max}^6) \right) \\ &= O \left(L^4 K \ell_{\max}^{10} + V N^2 K (L^2 + \ell_{\max}^6) \right), \end{aligned}$$

where V is the number of iterations. When computing y_k (30), as was the case for computing A_k , $p(\mathcal{I}_p | \vec{\ell}, \omega, x_k)$ is computed once at the beginning of each iteration with complexity of $O(N^2 K L^2)$. Moreover, the multiplication between the patches $\{\mathcal{I}_p\}_{p=1}^{N_{\text{patches}}}$ and the shifted and padded PSWF eigenfunctions $\psi_{N,n}$, denoted as $q(p, \vec{\ell}, \omega; \ell, m, s)$ (33), can be computed just once at the beginning of the algorithm with complexity $O(N^2 L^2 K \ell_{\max}^5)$. All in all, the computational complexity of computing y_k is $O(N^2 L^2 K \ell_{\max}^5 + V N^2 K (L^2 + \ell_{\max}^3))$. The computational complexity of solving the linear system of $O(\ell_{\max}^9)$ equations (29) is of computational complexity of $O(\ell_{\max}^9)$. As such, the total computational complexity of the approximate EM algorithm (Algorithm 1) is given by

$$\begin{aligned} & O \left(L^4 \ell_{\max}^{10} K + N^2 L^2 \ell_{\max}^5 K \right. \\ & \quad \left. + V (N^2 L^2 K + N^2 \ell_{\max}^6 K + \ell_{\max}^9) \right). \end{aligned}$$

The computational complexity consists of precomputations that require $O(L^4 \ell_{\max}^{10} K + N^2 L^2 \ell_{\max}^5 K)$ operations, and from $O(N^2 L^2 K + N^2 \ell_{\max}^6 K + \ell_{\max}^9)$ calculations required per iteration. We note that since we aim to estimate small molecular structures, L^3 (the dimension of the volume), and ℓ_{\max} (the maximum frequency in the expansion (6)) are expected to be rather small. However, N^2 —the total number of pixels in all micrographs—is expected to grow as the SNR decreases. Therefore, we expect that the computational complexity of the approximate EM algorithm will be governed by $O(V N^2 K \ell_{\max}^6)$. Thus, the running time is linear in the number of pixels in the micrograph and the number of rotations considered by the EM framework. The latter presents an accuracy-running time trade-off [37]. Future work includes more sophisticated and hierarchical techniques to control the sampling of the space of rotations, akin to existing methods in cryo-EM software [29], [56]. The benefit of the frequency marching procedure (see Section III-D) is clear from our analysis—the computational complexity depends on ℓ_{\max}^6 .

When using the stochastic variant of our approximate EM (see Algorithm 1), the computational complexity of the algorithm is $O(S \tilde{V} N^2 K \ell_{\max}^6)$, where \tilde{V} is the number of iterations. The computational complexity scales linearly with the stochastic factor $0 < S \leq 1$. However, the number of iterations \tilde{V} is expected to increase as we process less patches per iteration.

IV. NUMERICAL RESULTS¹

In this section, we present numerical results for the stochastic approximate EM recovery algorithm described in Section III. The micrographs for the experiments were generated as follows. We sample rotation matrices from $SO(3)$ uniformly at random as described in [57]. Given a volume and a sampled rotation matrix, we generate projections of the volume corresponding to the rotation matrix using the ASPIRE software package [58]. The projections are then placed in the measurement one by one; for each added projection, it is verified that the separation condition (5) is not violated. The number of projections in the measurement, T , is determined by the required density $\gamma = T \frac{N^2}{\tilde{L}^2}$, where \tilde{L}^2 is the size of a projection within the micrograph. Ultimately, the micrograph is corrupted by i.i.d. Gaussian noise with zero mean and variance corresponding to the desired SNR.

We present reconstructions from micrographs generated in two different ways:

- In the first method, we generate the micrograph from volumes of the original size. Due to computational constraints, we downsample the micrograph by a factor of \tilde{L}/L , where \tilde{L}^3 is the original size of the volume, and L^2 is the required projection size for computations; we aim to estimate a downsampled volume of size L^3 . Importantly, we assume that the downsampled target volumes are bandlimited (as assumed in Section II-A), but do not force it upon them. That is to say, the volume recovery error is bounded by this approximation error. This method imitates the procedure one will perform on experimental data sets.
- In the second method, we generate the micrograph from volumes downsampled to size L^3 , expanded using the expansion (6) up to the maximal frequency. By that, we follow exactly our mathematical model of the micrograph generation (see Section II) and we expect this method to outperform the first (more realistic) method.

We follow the standard convention in the cryo-EM literature, and measure the accuracy of the reconstruction using the Fourier shell correlation (FSC) metric. The FSC is calculated by correlating the 3-D Fourier components of two volumes (the ground truth f_{true} and the estimation f_{est}) and summing over spherical shells in Fourier space:

$$\text{FSC}(r) = \frac{\sum_{r_i \in S_r} \hat{f}_{\text{true}}(r_i) \cdot \hat{f}_{\text{est}}^*(r_i)}{\sqrt{\sum_{r_i \in S_r} |\hat{f}_{\text{true}}(r_i)|^2 \sum_{r_i \in S_r} |\hat{f}_{\text{est}}(r_i)|^2}}, \quad (38)$$

where S_r is the spherical shell of radius r . We use the 0.5 resolution cutoff: the resolution is determined as the frequency where the FSC curve drops below 0.5.

A note about the SNR metric: In the first method mentioned above, the micrographs are downsampled. The downsampling improves the SNR since the spectra of the volumes decay faster than the spectrum of the noise. Therefore, in the following reported numerical results, while we report the SNR for reproducibility purposes, we present a representative image of a micrograph for proper visual evaluation of the SNR.

¹The code to reproduce all numerical experiments is publicly available at <https://github.com/krshay/Stochastic-Approximate-EM-for-cryo-EM>.

A. Volume reconstructions

We present reconstructions of four volumes. All reconstructions were achieved by applying the stochastic approximate EM algorithm (Algorithm 1) with stochastic factor of $S = 0.05$ and $K = 552$ discrete rotations. The algorithm was initialized from an initial guess of the size of the target volume, whose entries were drawn i.i.d. from a Gaussian distribution with mean 0 and variance 1. In all experiments, we used 4 micrographs of size $N^2 = 2992^2$ pixels, with $T = 118374$ total projections. Each experiment consists of $N_{\text{patches}} = 295936$ patches, where in each iteration we use $\lfloor SN_{\text{patches}} \rfloor = 14796$ patches. The experiments were performed on a machine with 96 cores of Intel(R) Xeon(R) Gold 6252 CPU @ 2.10GHz with 1.51 TB of RAM, and took less than approximately 45 minutes per EM iteration with $\ell_{\max} = 6$, approximately 75 minutes per iteration with $\ell_{\max} = 8$, and approximately 5 hours per iteration with $\ell_{\max} = 10$. The molecular visualizations were produced using UCSF Chimera [59].

1) *The 3-D Shepp-Logan phantom [60]:* We consider the 3-D Shepp-Logan phantom of size $\tilde{L}^3 = 11^3$ voxels. Fig. 3b presents the estimation results from micrographs with $\text{SNR} = 3.5$. First, we use the first micrograph generation method, i.e., we simulate the micrographs with the true volume. A representative excerpt of the micrograph is presented in Fig. 3a. We have conducted 7 EM iterations with $\ell_{\max} = 6$, 10 iterations with $\ell_{\max} = 8$, and 8 iterations with $\ell_{\max} = 10$. A visual comparison between the true and estimated volumes is presented in Fig. 3b, and the FSC curve is given in Fig. 3c. Next, we use a 3-D Shepp-Logan phantom of size $\tilde{L}^3 = 11^3$ voxels expanded up to $\ell_{\max} = 10$ as our ground truth. The simulated micrographs are of the same $\text{SNR} = 3.5$. A representative excerpt of the micrograph is presented in Fig. 3d. We have conducted 31 EM iterations with $\ell_{\max} = 6$, 19 iterations with $\ell_{\max} = 8$, and 18 iterations with $\ell_{\max} = 10$. A visual comparison between the true and estimated volumes is presented in Fig. 3e, and the FSC curve is provided in Fig. 3f.

We can see, both visually and in terms of the FSC metric, that the estimation of the Shepp-Logan phantom is quite successful when we follow the mathematical model of Section II. However, when we aim to estimate the true, non-expanded, volume, the estimation is less accurate. This is since the true volume is not bandlimited, as is assumed in the expansion of Section II-A, so our expanded estimate fails to fully describe the true volume.

2) *The TRPV1 structure [61]:* The volume is available at the Electron Microscopy Data Bank (EMDB) as EMD-8117². The true structure is of size $\tilde{L}^3 = 192^3$ voxels. We consider the downsampled version of the volume of size $L^3 = 11^3$ voxels. For this experiment, the micrographs were simulated with $\text{SNR} = 1$. We use the first micrograph generation method, i.e., we generate the micrographs with the true 192^3 voxels volume, and then downsample the micrographs such that each projection is of size 11×11 pixels. A representative excerpt of the downsampled micrograph is presented in Fig. 4a. We have conducted 14 EM iterations

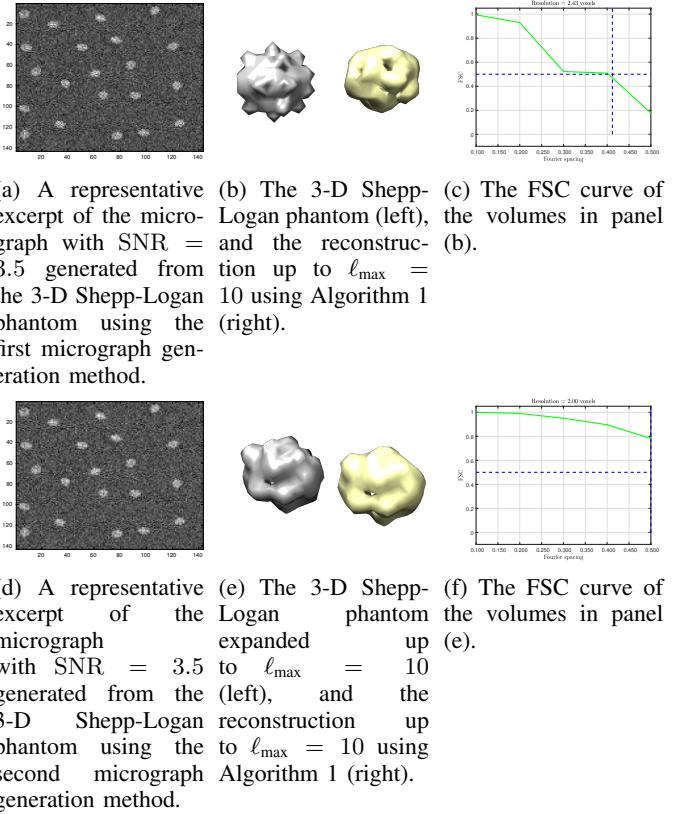


Fig. 3: Results for estimating the 3-D Shepp-Logan phantom directly from a micrograph.

with $\ell_{\max} = 6$, 9 iterations with $\ell_{\max} = 10$, and 12 iterations with $\ell_{\max} = 10$. A visual comparison between the true and the estimated volumes is presented in Fig. 4b, and the FSC curve is given in Fig. 4c. Using the second micrograph generation method, we have generated the micrographs with $\text{SNR} = 3.5$. A representative excerpt of the micrograph is presented in Fig. 4d. We have conducted 25 EM iterations with $\ell_{\max} = 6$, 8 iterations with $\ell_{\max} = 8$, and 4 iterations with $\ell_{\max} = 10$. A visual comparison is presented in Fig. 4e, and the FSC curve is provided in Fig. 4f.

Remarkably, we achieve accurate estimates of the downsampled TRPV1 structure from micrographs generated using both methods, perhaps due to its symmetrical structure. As expected, the result when we aim to estimate an expanded version of the volume is more accurate.

3) *The Plasmodium falciparum 80S ribosome [62]:* The volume is available at the EMDB as EMD-2660. The structure was cropped (to remove zeros at the boundaries) to size of $\tilde{L}^3 = 240^3$ voxels. First, we use the first micrograph generation method. For this experiment, the micrographs were simulated with $\text{SNR} = 1$. A representative excerpt of the downsampled micrograph is presented in Fig. 5a. We have conducted 12 EM iterations with $\ell_{\max} = 6$, 16 iterations with $\ell_{\max} = 10$, and 8 iterations with $\ell_{\max} = 10$. A visual comparison between the true and the estimated volumes is presented in Fig. 5b, and the FSC curve is given in Fig. 5c. Next, we use a downsampled 80S ribosome of size $\tilde{L}^3 = 11^3$ voxels expanded up to $\ell_{\max} = 10$ as our ground truth. The simulated

²<https://www.ebi.ac.uk/emdb/>

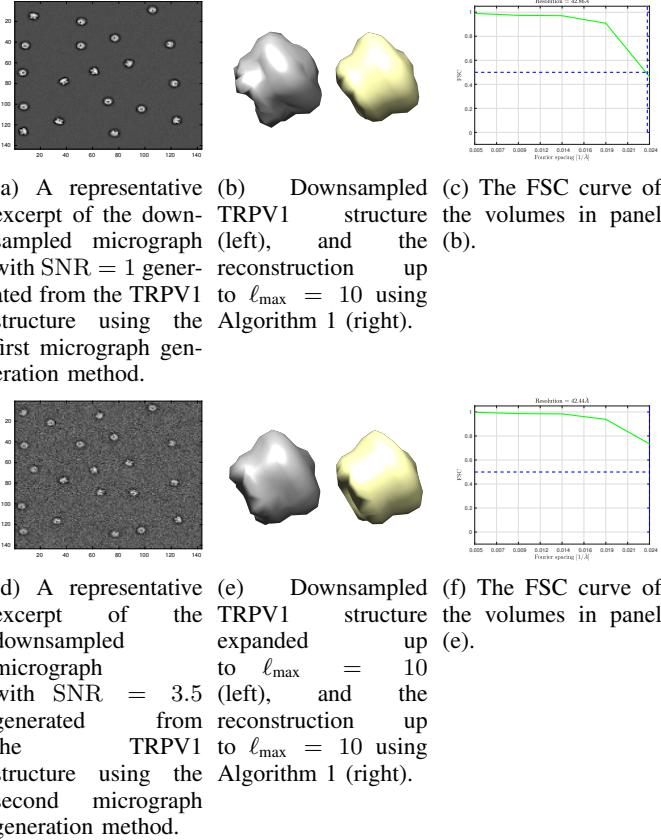


Fig. 4: Results for estimating the TRPV1 structure directly from a micrograph.

micrographs are of SNR = 3.5. A representative excerpt of the micrograph is presented in Fig. 5d. We have conducted 8 EM iterations with $\ell_{\max} = 6$, 8 iterations with $\ell_{\max} = 8$, and 13 iterations with $\ell_{\max} = 10$. A visual comparison between the true and the estimated volumes is provided in Fig. 5e, and the FSC curve is presented in Fig. 5f.

The estimation of the 80S ribosome is successful when we follow the mathematical model of Section II. However, when we aim to estimate the true, non-expanded, volume, the result is less accurate; the estimate follows the shape of the true downsampled volume, but the fine details are blurred. We believe that increasing the parameter K , the number of discrete rotations in the EM scheme, might improve the reconstruction accuracy, at the cost of running time (see Section III-E).

4) The Bovine Pancreatic Trypsin Inhibitor (BPTI) mutant [63]: The volume was generated in [30] from the atomic model from the Protein Data Bank (PDB)³, available as 1QLQ. The structure was cropped to size $\tilde{L}^3 = 31^3$ voxels. We consider the downsampled version of the volume of size $L^3 = 11^3$ voxels. First, we use the first micrograph generation method. For this experiment, the micrographs were simulated with SNR = 0.75. A representative excerpt of the downsampled micrograph is presented in Fig. 6a. We have conducted 15 EM iterations with $\ell_{\max} = 6$, 10 iterations with $\ell_{\max} = 8$, and 9 iterations with $\ell_{\max} = 10$. A visual

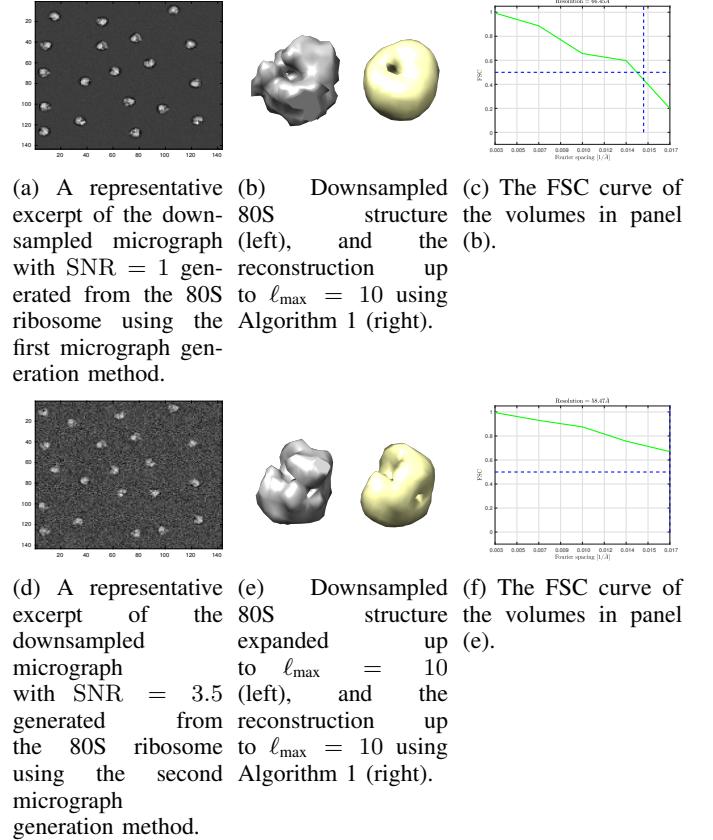


Fig. 5: Results for estimating the Plasmodium falciparum 80S ribosome directly from a micrograph.

comparison between the true and the estimated volumes is presented in Fig. 6b, and the FSC curve is given in Fig. 6c. Next, we use a downsampled BPTI volume of size $\tilde{L}^3 = 11^3$ voxels expanded up to $\ell_{\max} = 10$ as our ground truth. The simulated micrographs are of SNR = 3.5. A representative excerpt of the micrograph is presented in Fig. 6d. We have conducted 22 EM iterations with $\ell_{\max} = 6$, 6 iterations with $\ell_{\max} = 8$, and 15 iterations with $\ell_{\max} = 10$. A visual comparison between the true and the estimated volumes is presented in Fig. 6e, and the FSC curve is provided in Fig. 6f.

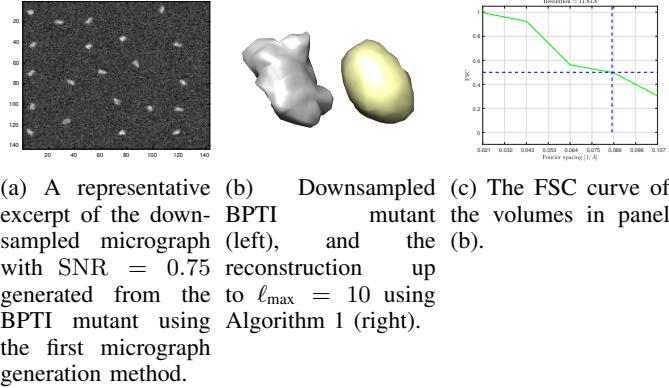
We notice a similar phenomenon to the 80S ribosome estimation—the estimation of the BPTI mutant is successful when we follow the accurate mathematical model of Section II, and less accurate when we aim to estimate the non-expanded volume.

In comparison to the results of [30] using autocorrelation analysis, our results are of much higher frequency— $\ell_{\max} = 10$ versus $\ell_{\max} = 2$. In addition, the recovery in [30] is from clean autocorrelations, corresponding to $N \rightarrow \infty$, while our recoveries were done from noisy micrographs. We note that the recovery in [30] was done for a larger volume, of size $L^3 = 31^3$ voxels. Fig. 7 presents a visual comparison between our estimate and the downsampled estimate from [30].

B. “Particle picking” using the approximate EM algorithm

During our EM procedure, we calculate the likelihood $p(\ell_p, \omega_p | \mathcal{I}_p, x_k)$ (see (23)) for each patch, where x_k is

³<https://www.rcsb.org/>



(d) A representative excerpt of the BPTI mutant downsampled to $\ell_{\max} = 10$ with SNR = 3.5 (left), and the generated from reconstruction up to $\ell_{\max} = 10$ using the second Algorithm 1 (right).

(e) A representative excerpt of the BPTI mutant downsampled to $\ell_{\max} = 10$ with SNR = 0.75 (left), and the generated from reconstruction up to $\ell_{\max} = 10$ using the first micrograph generation method.

(f) The FSC curve of the volumes in panel (e).

Fig. 6: Results for estimating the Bovine Pancreatic Trypsin Inhibitor (BPTI) mutant directly from a micrograph.

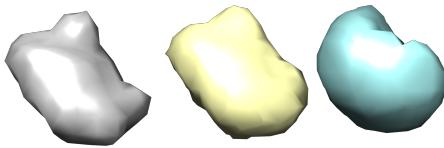


Fig. 7: A visual comparison between the downsampled BPTI mutant (left), the estimate of the volume expanded up to $\ell_{\max} = 10$ using Algorithm 1 (middle), and the downsampled estimate of the volume using autocorrelation analysis, expanded up to $\ell_{\max} = 2$, as was estimated in [30] (right). The estimation in [30] was done from clean autocorrelations, corresponding to $N \rightarrow \infty$.

the current estimate of the vector of coefficients. By averaging over the rotations, we get $p(\vec{\ell}_p | \mathcal{I}_p, x_k)$, which can be interpreted as the probability of each shift in the patch \mathcal{I}_p . Algorithm 2 introduces an algorithm to estimate the most probable shift in each patch. We stress that we do not suggest our algorithm as an alternative to existing particle pickers, but merely want to show that it succeeds to predict the true shifts if the SNR is high enough.

We measure the success of the “particle picking” procedure in two ways. First, we evaluate the performance of the algorithm to detect particles. Accurate detection of

Algorithm 2 “Particle picking” using the approximate EM algorithm

Input: measurement \mathcal{I} partitioned to N_{patches} patches; patch size L ; parameter K ; noise variance σ^2 ; estimate x .

Output: the shifts $\{\vec{\ell}_{\text{picked}_p}\}_{p=1}^{N_{\text{patches}}}$, an estimate of the locations of particles within patches.

Calculate $p(\vec{\ell}_p, \omega_p | \mathcal{I}_p, x)$ according to (23) for $p = 1, \dots, N_{\text{patches}}$;

Calculate $p(\vec{\ell}_p | \mathcal{I}_p, x) := \frac{1}{K} \sum_{\omega_p \in \Omega^K} p(\vec{\ell}_p, \omega_p | \mathcal{I}_p, x)$ for $p = 1, \dots, N_{\text{patches}}$;

Calculate $\vec{\ell}_{\text{picked}_p} := \arg \max_{\vec{\ell}_p \in \mathbb{L}} p(\vec{\ell}_p | \mathcal{I}_p, x)$ for $p = 1, \dots, N_{\text{patches}}$;

return $\{\vec{\ell}_{\text{picked}_p}\}_{p=1}^{N_{\text{patches}}}$

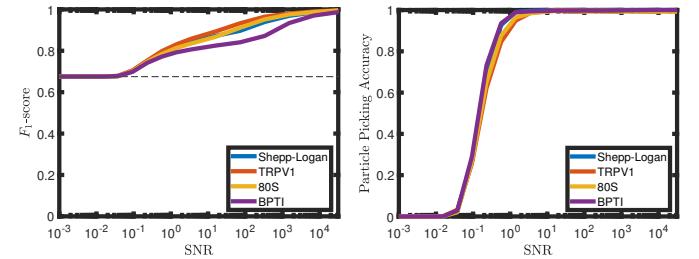


Fig. 8: Results for Algorithm 2 as a function of SNR. The “particle picking” is done with the ground truths. We report results for the F_1 -score for identifying empty patches (left panel), and for the shift estimation accuracy (right panel). The black dashed horizontal line (left panel) indicates the F_1 -score for a classifier which always identifies a patch as empty. For all volumes, the particle picking is successful down to a certain SNR; for highly noisy micrographs, as is expected, the particle picking fails.

empty patches is important since those patches do not add information regarding the target volume. To this end, we define empty patches as “positive,” and calculate the F_1 -score:

$$F_1 = 2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}, \quad (39)$$

where the precision is the ratio of true positives over all positives, and the sensitivity is the ratio of true positives over all patches detected as positive (both true positives and false positives). Second, we evaluate the performance of the algorithm in terms of accurately detecting the locations of the particles within patches. We allow an error of one pixel in each axis. We calculate the share of successfully detected particles from patches that are at least half-occupied by a projection image, i.e., patches which were generated (see (18)) with a shift $\vec{\ell}$ such that $\ell_x < L/2$ and $\ell_y < L/2$, or $\ell_x > 3L/2$ and $\ell_y > 3L/2$.

Fig. 8 presents the results of Algorithm 2 as a function of the SNR, using an accurate estimate of the target volume. We used micrographs of size 990×990 pixels with density $\gamma = 0.4$, and $K = 552$ discrete rotations. All in

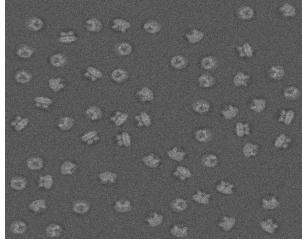


Fig. 9: A micrograph with an arbitrary spacing distribution of projection images, with SNR = 1. In contrast to Fig. 1, here the projection images can be arbitrarily close.

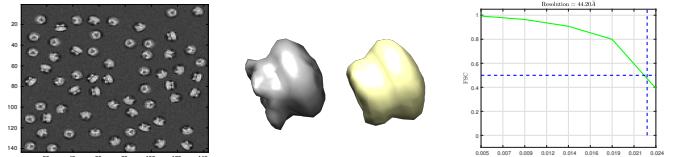
all, we processed $N_{\text{patches}} = 8100$ patches, with $T = 3240$ projection images. The results suggest that, in high SNR, Algorithm 2 succeeds both in detecting empty patches and in picking the correct locations of the projection images within the patches. We also show that the accuracy deteriorates as the noise level increases—up to complete failure in extremely noisy environments, as expected; in low-SNR regimes, the F_1 -score is no better than simply identifying all patches as empty.

C. Arbitrary spacing distribution of projection images within the micrograph

So far, we considered the case of well-separated micrographs, where each projection image is separated by at least a full projection length from its neighbor (5). This allowed us to develop the approximate EM framework (Algorithm 1) under the assumption that each patch contains at most one projection image.

We now discuss the case of an arbitrary spacing distribution of projection images within the micrograph, which better reflects practical cryo-EM micrographs. In this model, we assume only that the projection images do not overlap; see Fig. 9 for an example of such simulated micrograph. Consequently, when partitioning the measurement to N_{patches} of size $L \times L$, each patch can now consist of up to 4 projection images. In order to integrate the arbitrary spacing distribution model into the approximate EM framework, one needs to account for the possibility of multiple projections in a single patch. Namely, the set of possible shifts within a patch of all four potential image projections is \mathbb{L}^4 , and the space of possible rotations is the group $SO(3) \times SO(3) \times SO(3) \times SO(3)$. A similar analysis was conducted for 1-D [38] and 2-D [45] related models. Clearly, this mechanism will greatly inflate the computational complexity of the algorithm.

In the following experiment, we examine the performance of Algorithm 1 (that assumes well-separated micrographs) to micrographs with an arbitrary spacing distribution of projection images. Approximately half of the non-empty patches contain parts of more than one projection image. Nevertheless, as depicted in Fig. 10, we achieve an accurate estimate of the TRPV1 structure from a downsampled micrograph with an arbitrary spacing distribution of projection images, though the estimate is moderately worse than the estimate of the well-separated case (Fig. 4). This indicates that the well-separated model may suffice to achieve recoveries of reasonable resolution. We have conducted 17 EM iterations with $\ell_{\max} = 6, 12$



(a) A representative excerpt of a downsampled TRPV1 structure. (b) The volumes in panel (a) sampled micrograph (left), and the estimated volume (right). (c) The FSC curve of the reconstruction up to an arbitrary spacing distribution of projection images.

Fig. 10: Results for estimating the TRPV1 structure directly from a micrograph with an arbitrary spacing distribution of projection images.

iterations with $\ell_{\max} = 8$, and 7 iterations with $\ell_{\max} = 10$. A visual comparison between the true and estimated volumes is presented in Fig. 10b, and the FSC curve is given in Fig. 10c.

V. CONCLUSIONS

In this paper, we demonstrated successful 3-D volume reconstructions directly from simulated cryo-EM micrographs, using the stochastic approximate EM algorithm. Our approach allows us to estimate the target volume directly from the measurement, without particle picking. Therefore, it might be possible to reconstruct small molecular structures, and in particular structures that are too small to be detected by particle picking methods.

Recovering structures from highly noisy micrographs would require processing more data. The main hurdle to achieve this goal is the computational burden of the algorithm. In this paper, we accelerated the algorithm by using a frequency marching scheme, and a stochastic strategy. Nevertheless, further acceleration measures should be considered. A possible path is to sample the space of rotations more densely as the iterations progress (similarly to the frequency marching scheme), or to design a branch-and-bound algorithm that inexpensively rules out regions of the search space with low probability to achieve the maximum of the likelihood function [56]. Further acceleration can be attained by applying the EM algorithm on a projection of the data onto a lower dimensional subspace [64]. In addition, adding a prior (in Bayesian sense) on the 3-D volume is natural in the EM algorithm, and is expected to improve robustness (at the cost of possible model bias). In addition, the EM iterations could be initialized by efficient computational techniques, such as autocorrelation analysis [30].

An alternative direction is to replace EM by other estimation techniques, with emphasis on generative models. Generative models have been recently applied to various inverse imaging problems, including cryo-EM [65], [66]. Such models include variational auto-encoders (VAEs) [67], [68], [69], generative adversarial networks (GANs) [70], [71], and score-based diffusion models [72].

In addition, our cryo-EM micrograph generation model (see Section II) is simplified and incomplete. As was mentioned in Section IV-C, practical cryo-EM micrographs do not follow the well-separated model (5), but rather have an arbitrary spacing distribution of projection images. An initial experiment (see Fig. 10) suggests that perhaps applying the stochastic approximate EM algorithm also to those micrographs will result in sufficiently accurate estimates. Moreover, the noise in experimental micrographs might be colored, and the viewing angles of the projection images are typically not distributed uniformly. The cryo-EM measurement is also affected by contaminations and by the electron microscope's point spread function [47]. Addressing those modeling issues is essential to apply the proposed algorithm to experimental data sets.

ACKNOWLEDGMENTS

The authors would like to thank Eitan Levin for his kind assistance in the initial stages of the code implementation.

REFERENCES

- [1] E. Nogales, "The development of cryo-EM into a mainstream structural biology technique," *Nature Methods*, vol. 13, no. 1, pp. 24–27, 2016.
- [2] X.-C. Bai, G. McMullan, and S. H. Scheres, "How cryo-EM is revolutionizing structural biology," *Trends in Biochemical Sciences*, vol. 40, no. 1, pp. 49–57, 2015.
- [3] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: Visualization of biological molecules in their native state*. Oxford University Press, 2006.
- [4] D. Elmlund and H. Elmlund, "Cryogenic electron microscopy and single-particle analysis," *Annual Review of Biochemistry*, vol. 84, no. 1, pp. 499–517, 2015.
- [5] Y. Cheng, N. Grigorieff, P. A. Penczek, and T. Walz, "A primer to single-particle cryo-electron microscopy," *Cell*, vol. 161, no. 3, pp. 438–449, 2015.
- [6] F. J. Sigworth, "Principles of cryo-EM single-particle image processing," *Microscopy*, vol. 65, no. 1, pp. 57–67, 2016.
- [7] A. Singer and F. J. Sigworth, "Computational methods for single-particle electron cryomicroscopy," *Annual Review of Biomedical Data Science*, vol. 3, pp. 163–190, 2020.
- [8] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 58–76, 2020.
- [9] F. Wang, H. Gong, G. Liu, M. Li, C. Yan, T. Xia, X. Li, and J. Zeng, "DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM," *Journal of Structural Biology*, vol. 195, no. 3, pp. 325–336, 2016.
- [10] A. Heimowitz, J. Andén, and A. Singer, "APPLE picker: Automatic particle picking, a low-effort cryo-EM framework," *Journal of Structural Biology*, vol. 204, no. 2, pp. 215–227, 2018.
- [11] T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, and B. Berger, "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature Methods*, vol. 16, no. 11, pp. 1153–1160, 2019.
- [12] A. Eldar, B. Landa, and Y. Shkolnisky, "KLT picker: Particle picking using data-driven optimal templates," *Journal of Structural Biology*, vol. 210, no. 2, p. 107473, 2020.
- [13] F. J. Sigworth, "Classical detection theory and the cryo-EM particle selection problem," *Journal of Structural Biology*, vol. 145, no. 1-2, pp. 111–122, 2004.
- [14] R. Henderson, "The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules," *Quarterly Reviews of Biophysics*, vol. 28, no. 2, pp. 171–193, 1995.
- [15] R. M. Glaeser, "Electron crystallography: Present excitement, a nod to the past, anticipating the future," *Journal of Structural Biology*, vol. 128, no. 1, pp. 3–14, 1999.
- [16] S. Wu, A. Avila-Sakar, J. Kim, D. S. Booth, C. H. Greenberg, A. Rossi, M. Liao, X. Li, A. Alian, S. L. Griner, N. Juge, Y. Yu, C. M. Mergel, J. Chaparro-Riggers, P. Strop, R. Tampé, R. H. Edwards, R. M. Stroud, C. S. Craik, and Y. Cheng, "Fabs enable single particle cryoEM studies of small proteins," *Structure*, vol. 20, no. 4, pp. 582–592, 2012.
- [17] R. Danev and W. Baumeister, "Expanding the boundaries of cryo-EM with phase plates," *Current Opinion in Structural Biology*, vol. 46, pp. 87–94, 2017.
- [18] G. Scapin, C. S. Potter, and B. Carragher, "Cryo-EM for small molecules discovery, design, understanding, and application," *Cell Chemical Biology*, vol. 25, no. 11, pp. 1318–1325, 2018.
- [19] Y. Liu, D. T. Huynh, and T. O. Yeates, "A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold," *Nature Communications*, vol. 10, no. 1, pp. 1–7, 2019.
- [20] K. Zhang, S. Li, K. Kappel, G. Pintilie, Z. Su, T.-C. Mou, M. F. Schmid, R. Das, and W. Chiu, "Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution," *Nature Communications*, vol. 10, no. 1, pp. 1–6, 2019.
- [21] M. Wu and G. C. Lander, "How low can we go? Structure determination of small biological complexes using single-particle cryo-EM," *Current Opinion in Structural Biology*, vol. 64, pp. 9–16, 2020.
- [22] T. O. Yeates, M. P. Agdanowski, and Y. Liu, "Development of imaging scaffolds for cryo-electron microscopy," *Current Opinion in Structural Biology*, vol. 60, pp. 142–149, 2020.
- [23] X.-c. Bai, "Seeing atoms by single-particle cryo-EM," *Trends in Biochemical Sciences*, vol. 46, no. 4, pp. 253–254, 2021.
- [24] X. Wu and T. A. Rapoport, "Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies)," *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, p. e2115001118, 2021.
- [25] L. Zheng, N. Liu, X. Gao, W. Zhu, K. Liu, C. Wu, R. Yan, J. Zhang, X. Gao, Y. Yao, B. Deng, J. Xu, Y. Lu, Z. Liu, M. Li, X. Wei, H.-W. Wang, and H. Peng, "Uniform thin ice on ultraflat graphene for high-resolution cryo-EM," *Nature Methods*, pp. 1–8, 2022.
- [26] G. Harauz and F. Ottensmeyer, "Direct three-dimensional reconstruction for macromolecular complexes from electron micrographs," *Ultramicroscopy*, vol. 12, no. 4, pp. 309–319, 1983.
- [27] J. Neyman and E. L. Scott, "Consistent estimates based on partially consistent observations," *Econometrica: Journal of the Econometric Society*, pp. 1–32, 1948.
- [28] C. Aguerreberre, M. Delbracio, A. Bartesaghi, and G. Sapiro, "Fundamental limits in multi-image alignment," *IEEE Transactions on Signal Processing*, vol. 64, no. 21, pp. 5707–5722, 2016.
- [29] S. H. Scheres, "RELION: implementation of a Bayesian approach to cryo-EM structure determination," *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.
- [30] T. Bendory, N. Boumal, W. Leeb, E. Levin, and A. Singer, "Toward single particle reconstruction without particle picking: Breaking the detection limit," *SIAM Journal on Imaging Sciences*, to be published.
- [31] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.
- [32] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 477–489, 1988.
- [33] N. Segol and B. Nadler, "Improved convergence guarantees for learning Gaussian mixture models by EM and gradient EM," *Electronic Journal of Statistics*, vol. 15, no. 2, pp. 4510–4544, 2021.
- [34] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [35] A. Hyvärinen, "Independent component analysis: Recent advances," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, p. 20110534, 2013.
- [36] F. J. Sigworth, "A maximum-likelihood approach to single-particle image refinement," *Journal of structural biology*, vol. 122, no. 3, pp. 328–339, 1998.
- [37] S. Kreymer, A. Singer, and T. Bendory, "An approximate expectation-maximization for two-dimensional multi-target detection," *IEEE Signal Processing Letters*, vol. 29, pp. 1087–1091, 2022.
- [38] T.-Y. Lan, T. Bendory, N. Boumal, and A. Singer, "Multi-target detection with an arbitrary spacing distribution," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1589–1601, 2020.
- [39] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer, "Bispectrum inversion with application to multireference alignment," *IEEE Transactions on Signal Processing*, vol. 66, no. 4, pp. 1037–1050, 2017.

- [40] E. Abbe, T. Bendory, W. Leeb, J. M. Pereira, N. Sharon, and A. Singer, “Multireference alignment is easier with an aperiodic translation distribution,” *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3565–3584, 2018.
- [41] R. Henderson, “Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 45, pp. 18 037–18 041, 2013.
- [42] 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.
- [43] N. F. Marshall, T.-Y. Lan, T. Bendory, and A. Singer, “Image recovery from rotational and translational invariants,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5780–5784.
- [44] T. Bendory, T.-Y. Lan, N. F. Marshall, I. Rukshin, and A. Singer, “Multi-target detection with rotations,” *Inverse Problems and Imaging*, vol. 17, no. 2, pp. 362–380, 2023.
- [45] S. Kreymer and T. Bendory, “Two-dimensional multi-target detection: An autocorrelation analysis approach,” *IEEE Transactions on Signal Processing*, vol. 70, pp. 835–849, 2022.
- [46] Y. Shalit, R. Weber, A. Abas, S. Kreymer, and T. Bendory, “Generalized autocorrelation analysis for multi-target detection,” in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 5907–5911.
- [47] H. Erickson and A. Klug, “Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs,” *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 261, no. 837, pp. 105–118, 1971.
- [48] T. Zhang, A. Singer, and T. Bhamre, “Anisotropic twicing for single particle reconstruction using autocorrelation analysis,” May 13 2021, US Patent App. 16/605,987.
- [49] E. Levin, T. Bendory, N. Boumal, J. Kileel, and A. Singer, “3D ab initio modeling in cryo-EM by autocorrelation analysis,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1569–1573.
- [50] F. Natterer, *The Mathematics of Computerized Tomography*. USA: Society for Industrial and Applied Mathematics, 2001.
- [51] D. Slepian, “Prolate spheroidal wave functions, Fourier analysis and uncertainty—IV: Extensions to many dimensions; generalized prolate spheroidal functions,” *Bell System Technical Journal*, vol. 43, no. 6, pp. 3009–3057, 1964.
- [52] B. Landa and Y. Shkolnisky, “Steerable principal components for space-frequency localized images,” *SIAM Journal on Imaging Sciences*, vol. 10, no. 2, pp. 508–534, 2017.
- [53] W. T. Freeman and E. H. Adelson, “The design and use of steerable filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
- [54] R. M. Neal and G. E. Hinton, “A view of the EM algorithm that justifies incremental, sparse, and other variants,” in *Learning in Graphical Models*. Springer, 1998, pp. 355–368.
- [55] A. Barnett, L. Greengard, A. Pataki, and M. Spivak, “Rapid solution of the cryo-EM reconstruction problem by frequency marching,” *SIAM Journal on Imaging Sciences*, vol. 10, no. 3, pp. 1170–1195, 2017.
- [56] 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, 2017.
- [57] K. Shoemake, “Uniform random rotations,” in *Graphics Gems III (IBM Version)*. Elsevier, 1992, pp. 124–132.
- [58] G. Wright, J. Andén, V. Bansal, J. Xia, C. Langfield, J. Carmichael, R. Brook, Y. Shi, A. Heimowitz, G. Pragier, I. Sason, A. Moscovich, Y. Shkolnisky, and A. Singer, “ComputationalCryoEM/ASPIRE-Python: v0.10.1,” Jan. 2023.
- [59] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF Chimera—a visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [60] L. A. Shepp and B. F. Logan, “The Fourier reconstruction of a head section,” *IEEE Transactions on Nuclear Science*, vol. 21, no. 3, pp. 21–43, 1974.
- [61] Y. Gao, E. Cao, D. Julius, and Y. Cheng, “TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action,” *Nature*, vol. 534, no. 7607, pp. 347–351, 2016.
- [62] W. Wong, X.-c. Bai, A. Brown, I. S. Fernandez, E. Hanssen, M. Condron, Y. H. Tan, J. Baum, and S. H. Scheres, “Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine,” *Elife*, vol. 3, p. e03080, 2014.
- [63] H. Czapinska, J. Otlewski, S. Krzywda, G. M. Sheldrick, and M. Jaskólski, “High-resolution structure of bovine pancreatic trypsin inhibitor with altered binding loop sequence,” *Journal of Molecular Biology*, vol. 295, no. 5, pp. 1237–1249, 2000.
- [64] N. C. Dvornek, F. J. Sigworth, and H. D. Tagare, “SubspaceEM: A fast maximum-a-posteriori algorithm for cryo-EM single particle reconstruction,” *Journal of Structural Biology*, vol. 190, no. 2, pp. 200–214, 2015.
- [65] C. Donnat, A. Levy, F. Poitevin, E. D. Zhong, and N. Miolane, “Deep generative modeling for volume reconstruction in cryo-electron microscopy,” *Journal of Structural Biology*, p. 107920, 2022.
- [66] Z. Zhao, J. C. Ye, and Y. Bresler, “Generative models for inverse imaging problems: From mathematical foundations to physics-driven applications,” *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 148–163, 2023.
- [67] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in *2nd International Conference on Learning Representations, ICLR 2014*, 2014.
- [68] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis, “CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks,” *Nature Methods*, vol. 18, no. 2, pp. 176–185, 2021.
- [69] D. Rosenbaum, M. Garnelo, M. Zielinski, C. Beattie, E. Clancy, A. Huber, P. Kohli, A. W. Senior, J. Jumper, C. Doersch, S. M. A. Eslami, O. Ronneberger, and J. Adler, “Inferring a continuous distribution of atom coordinates from cryo-EM images using VAEs,” *arXiv preprint arXiv:2106.14108*, 2021.
- [70] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [71] H. Gupta, M. T. McCann, L. Donati, and M. Unser, “CryoGAN: A new reconstruction paradigm for single-particle cryo-EM via deep adversarial learning,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 759–774, 2021.
- [72] Y. Song, L. Shen, L. Xing, and S. Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *The Tenth International Conference on Learning Representations, ICLR 2022*, 2022.