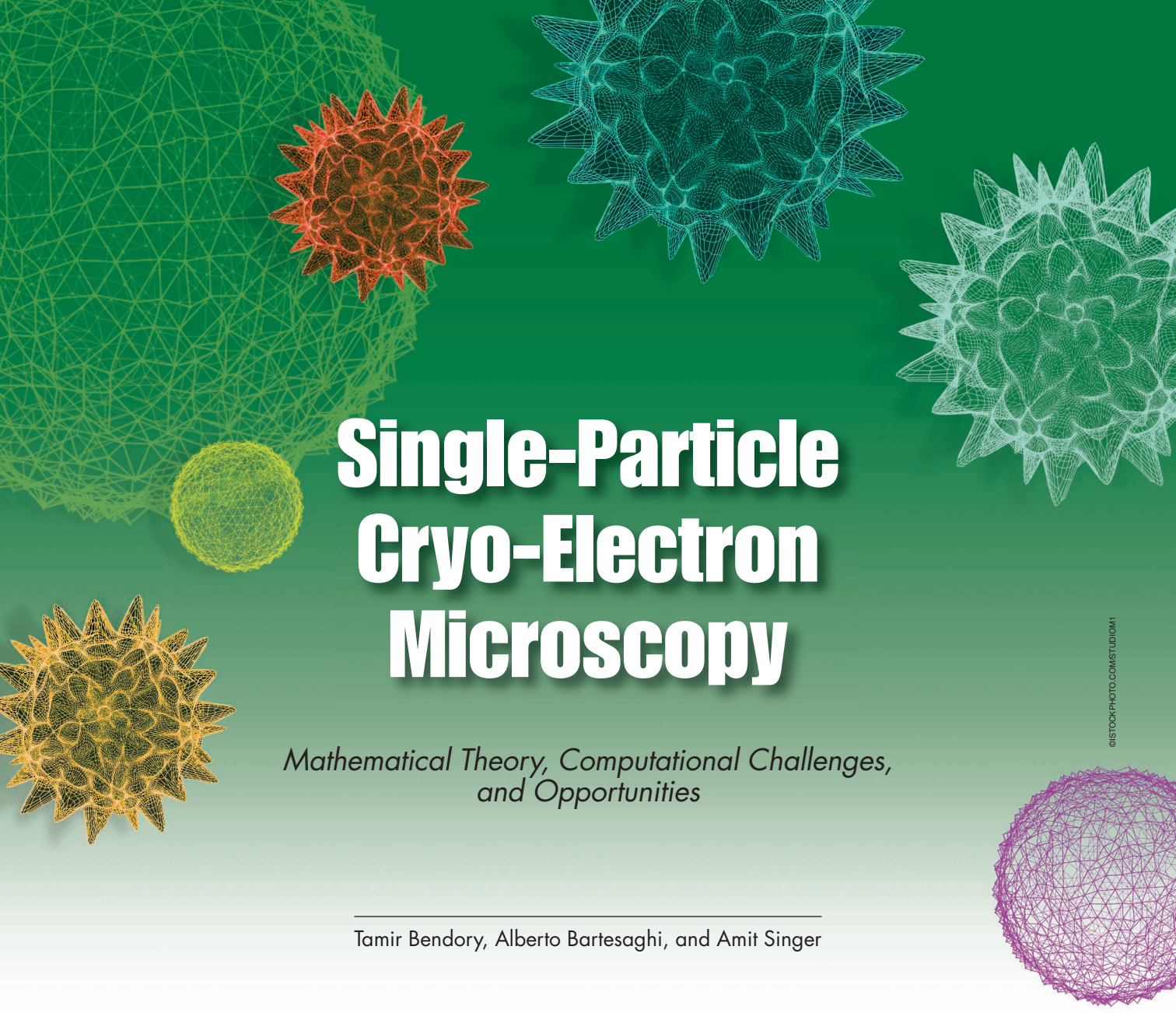

Single-Particle Cryo-Electron Microscopy

Mathematical Theory, Computational Challenges, and Opportunities

Tamir Bendory, Alberto Bartesaghi, and Amit Singer

In recent years, an abundance of new molecular structures have been elucidated using cryo-electron microscopy (cryo-EM), largely due to advances in hardware technology and data processing techniques. Owing to these exciting new developments, cryo-EM was selected by *Nature Methods* as the "Method of the Year 2015," and the Nobel Prize in Chemistry 2017 was awarded to three pioneers in the cryo-EM field: Jacques Dubochet, Joachim Frank, and Richard Henderson "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution" [93].

The main goal of this article is to introduce the challenging and exciting computational tasks involved in reconstructing 3D molecular structures by cryo-EM. Determining molecular structures requires a wide range of computational tools in a variety of fields, including signal processing, estimation and detection theory, high-dimensional statistics, convex and nonconvex optimization, spectral algorithms, dimensionality reduction, and machine learning. The tools from these fields must be adapted to work under exceptionally challenging conditions, including extreme noise levels, the presence of missing data, and massive data sets as large as several terabytes.

In addition, we present two statistical models, multireference alignment (MRA) and multitarget detection (MTD), that abstract away much of the intricacy of cryo-EM while retaining some of its essential features. Based on these abstractions, we discuss some recent intriguing results in the mathematical theory of cryo-EM and delineate relations with group, invariant, and information theories.

Digital Object Identifier 10.1109/MSP.2019.2957822
Date of current version: 26 February 2020

Introduction

Structural biology studies the structure and dynamics of macromolecules to broaden our knowledge about the mechanisms of life and impact the drug-discovery process. Owing to recent

groundbreaking developments, chiefly in hardware technologies and data processing techniques, many new molecular structures have been elucidated to near-atomic resolutions using cryo-EM [8], [44], [47], [52], [81].

In a cryo-EM experiment, biological macromolecules suspended in a liquid solution are rapidly frozen into a thin ice layer. The 3D location and orientation of particles within the ice are random and unknown. An electron beam then passes through the sample, and a 2D tomographic projection, called a *micrograph*, is recorded. The goal is to reconstruct a high-resolution estimate of the 3D electrostatic potential of the molecule (particularly its atomic structure) from a set of micrographs.

The resolution measures the smallest detail that is distinguishable in a recovered 3D structure; structures with better resolutions resolve finer features. For example, at resolutions of 9 Å, α-helices are resolved, at resolutions of 4.8 Å, individual β-strands are resolved, and, at resolutions of 3.5 Å, many amino acid side chains are resolved [8]. Figure 1 shows a gallery of important biomedical structures solved by cryo-EM at increasingly higher resolutions. Figure 2 presents an example of a micrograph of the enzyme β-galactosidase and the corresponding high-resolution 3D reconstruction [15].

The signal-to-noise ratio (SNR) of cryo-EM data is very low due to two compounding reasons. On the one hand, the micrograph's contrast is low due to the absence of contrast enhancement agents, such as heavy-metal stains. On the other hand, the noise level is high because the electron doses must be kept low to prevent damage to the radiation-sensitive biological molecules. The difficulty of estimating the 3D structure in this low-SNR regime, when the orientation and location of the particles are unknown, is the crux of the cryo-EM problem.

Forty years ago, Dubochet and colleagues [27] devised a new technique to preserve biological samples within a thin layer of an amorphous solid form of water called *vitreous ice*. In contrast to "regular" ice, vitreous ice lacks a crystalline molecular arrangement and, therefore, allows preservation of biological structures. As the vitreous ice is maintained at liquid-nitrogen or liquid-helium temperatures, the technique was named *cryo-EM*. In the following decades, the successful application of cryo-EM was limited to the study of large and highly symmetric structures, such as ribosomes and different types of viruses. Before 2013, only a few structures were resolved at resolutions better than 7 Å, and the field was dubbed *blob-ology* due to the blobby appearance of the structures at these resolutions. We refer the reader to [8], [30], [52], and [81] for a more detailed historical account of the development of the technology.

Since 2013, single-particle reconstruction using cryo-EM has been undergoing fast transformations, leading to an abundance of new high-resolution structures and reaching close to

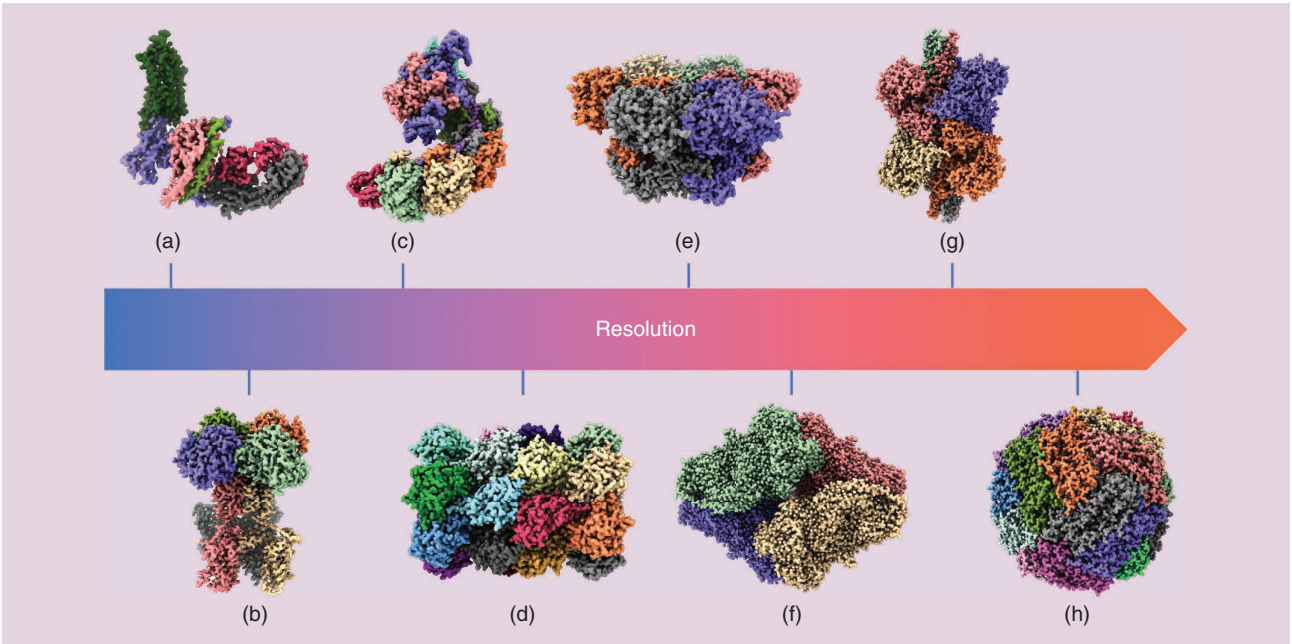

FIGURE 1. A gallery of important biomedical structures solved by single-particle cryo-EM at increasing resolutions. (a) The 4.5-Å structure of the human rhodopsin receptor bound to an inhibitory G protein [43], a member of the family of G-protein-coupled receptors, which are the target of approximately 35% of drugs approved by the U.S. Food and Drug Administration. (b) A 3.3-Å map of a voltage-activated potassium channel, an integral membrane protein responsible for potassium ion transport [49]. (c) The 2.9-Å cryo-EM structure of a clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein implicated in gene editing [36]. (d) A 2.4-Å map of the T20s proteasome, a complex that degrades unnecessary or damaged proteins by proteolysis [obtained using data available from Electron Microscopy Public Image Archive (EMPIAR)-10025] [1]. (e) The 2.3-Å structure of human p97 adenosine triphosphate (ATP)/ATPase associated with diverse cellular activities, a key mediator of several protein homeostasis processes and a target for cancer [12]. (f) The 1.9-Å structure of the β-galactosidase enzyme in complex with a cell-permeant inhibitor [15]. (g) The 1.8-Å structure of the conformationally dynamic enzyme glutamate dehydrogenase [50]. (h) A 1.6-Å map of human apoferritin, a critical intracellular iron-storage protein (obtained using data available from EMPIAR-10200).

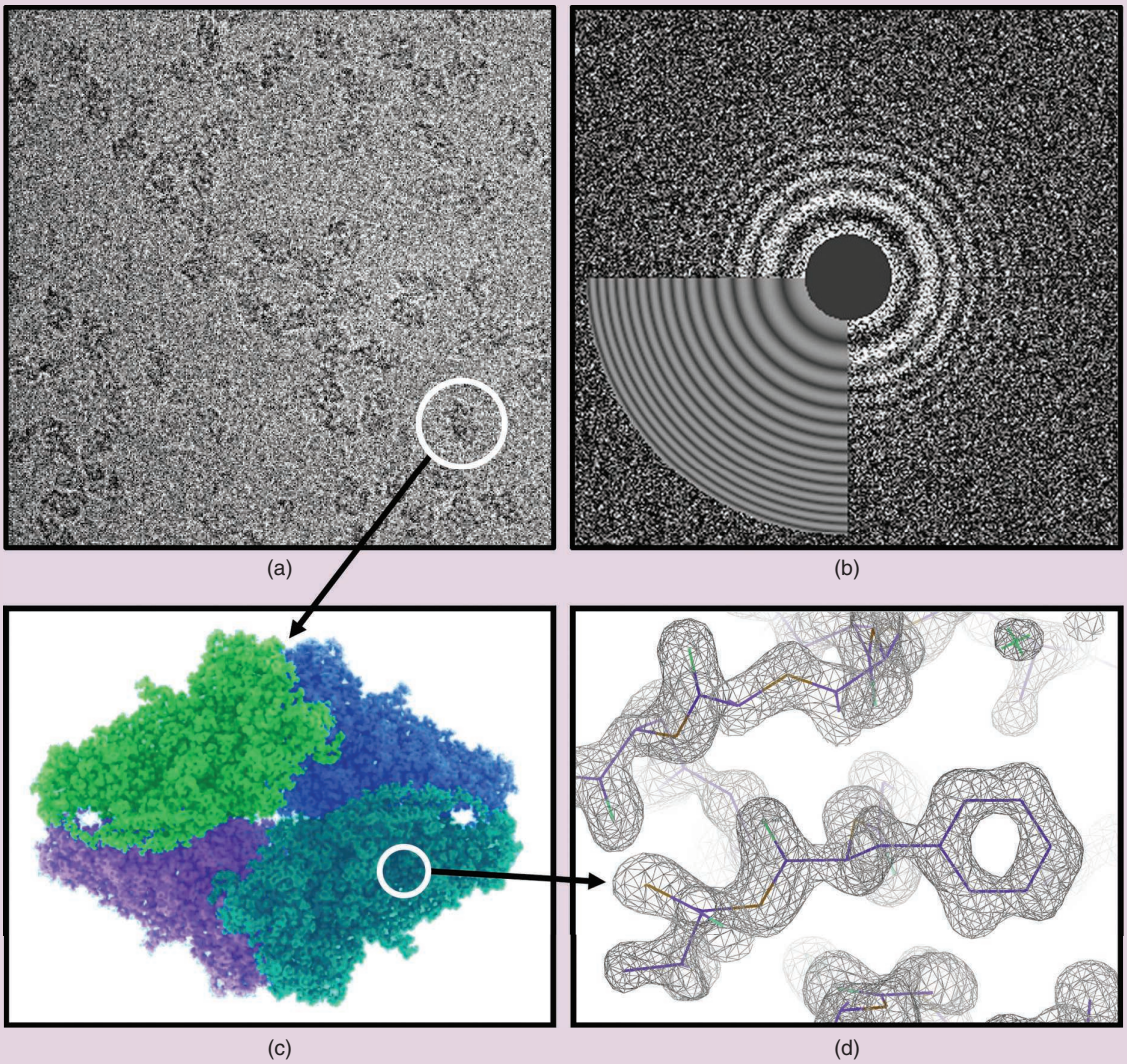

FIGURE 2. Examples of high-resolution cryo-EM imaging of the β-galactosidase enzyme in complex with a cell-permeant inhibitor. (a) A micrograph of β-galactosidase showing individual particle projections (indicated with white circle). (b) The power spectra of the image shown in (a) and estimated CTF matching the characteristic Thon ring oscillations (see the "CTF Estimation and Correction" section). (c) A 1.9-Å resolution map obtained from approximately 150,000 individual particle projections extracted from the publicly available data set EMPIAR-10061 [15], [16]. (d) A close-up view of reconstruction shown in (c), highlighting high-resolution features of the map at the individual amino acid level.

atomic resolution [15]. Figure 3 presents the historical growth of the number of high-resolution structures produced by cryo-EM. Advances in camera technology and data processing contributed to the "resolution revolution" in cryo-EM [44].

First and foremost, a new generation of detectors was developed, called *direct electron detectors* (*DEDs*), that—in contrast to charge-coupled device cameras—do not convert electrons to photons. DEDs dramatically improved image quality and SNR, thereby increasing the attainable resolution of cryo-EM. These detectors have high-output frame rates that allow the recording of multiple frames per micrograph ("movies") rather than the integration of individual exposures [58]. These movies can be used to compensate for motion induced by the electron beam to the sample; see the "Motion Correction" section. In addition, recent hardware developments have enabled the acquisition and storage of huge amounts of data, which, combined with ready access to CPU and GPU resources, have helped propel the field forward.

The prevalent technique in structural biology in the last half century was X-ray crystallography. This technique suffers from three intrinsic weaknesses that can be mitigated by using cryo-EM imaging. First, many molecules, among them different types of membrane proteins, were notoriously difficult to crystallize. In contrast, the sample preparation procedure for a cryo-EM experiment is significantly simpler, does not require crystallization, and needs smaller amounts of sample. Second, crystal contacts may alter the structure of proteins, making it difficult to recover their physiologically relevant conformation. Cryo-EM samples, instead, are rapidly frozen into vitreous ice, which preserves the molecules in a near-physiological

state. Third, the X-ray beam aggregates the information from all molecules simultaneously and, thus, hinders visualization of structural variability.

In stark contrast, the projection of each particle in a cryo-EM experiment is recorded independently, and, thus, multiple structures—associated with different functional states—can be estimated [74]. A third technique, nuclear magnetic resonance (NMR), can be used to elucidate molecular structures in physiological conditions (water solution at room temperature) but is restricted to small structures of up to approximately 50 kDa.

The main goal of this article is to introduce the unique and exciting algorithmic challenges and cutting-edge mathematical problems arising from cryo-EM research. The estimation of molecular structures involves developing and adopting computational tools in signal processing, estimation and detection theory, high-dimensional statistics, convex and nonconvex optimization, spectral algorithms, dimensionality reduction, and machine learning as well as knowledge in group theory, invariant theory, and information theory.

All tools from the aforementioned fields should be adapted to exceptional conditions: an extremely low-SNR environment and the presence of missing data (for instance, 2D location and 3D orientation of samples in the micrograph). In addition, the devised algorithms should be efficient when run on massively large data sets on the order of several terabytes. This article provides an account of the leading software packages in the field and discusses their underlying mathematical, statistical, and algorithmic principles [33], [57], [63], [76].

Before moving on, we want to mention that topics of great importance to practitioners, such as the physics and optics of an electron microscope, sample preparation, and data acquisition, are not discussed in this article. These topics are thoroughly covered by biologically oriented surveys, such as [30], [52], and [81] and the references therein.

The cryo-EM reconstruction problem

Modern electron microscopes produce multiple micrographs, each composed of a series of frames (a "movie"). The first stage of any contemporary algorithmic pipeline is to align and average the frames to mitigate the effects of movement induced by the electron beam and, thus, to improve the SNR. This process is called *motion correction* or *movie frame alignment*. The next step, termed *particle picking*, consists of detecting and extracting the particles' tomographic projections from the micrographs.

Perfect detection requires finding the particles' center of mass, which is difficult to estimate due to the characteristics of the Fourier transform of the microscope's point spread function (PSF), called the *contrast transfer function* (*CTF*); see the "CTF Estimation and Correction" section. The output of the particle-picking stage is a series of images $I_1, \ldots, I_N$ from which we wish to estimate the 3D structure; this section focuses on the problem of 3D reconstruction from picked particles, which is the heart of the computational pipeline of single-particle reconstruction using cryo-EM. Motion correction and particle picking are discussed in more detail in the "Building Blocks in the Computational Pipeline" section.

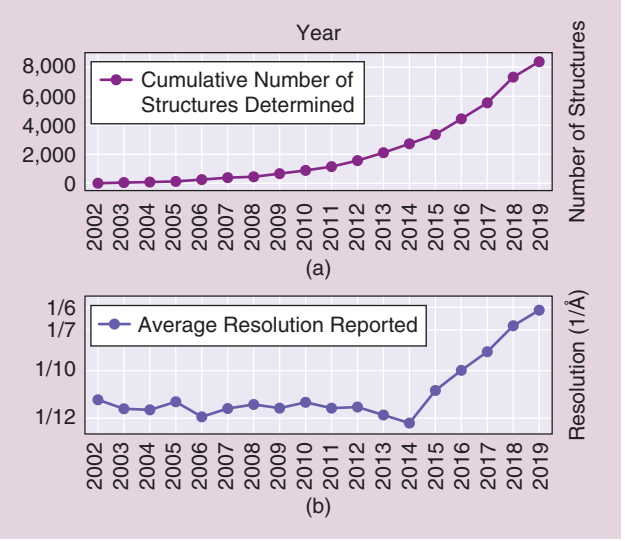

FIGURE 3. The recent growth in the number of high-resolution structures produced by single-particle cryo-EM. (a) The cumulative number of structures solved by single-particle cryo-EM in the last 17 years, as recorded in the Electron Microscopy Data Bank public repository [2]. (b) The corresponding values for the average resolution of maps deposited in the database showing an inflection point after the year 2013, coinciding with the introduction of DED technology in the cryo-EM field.

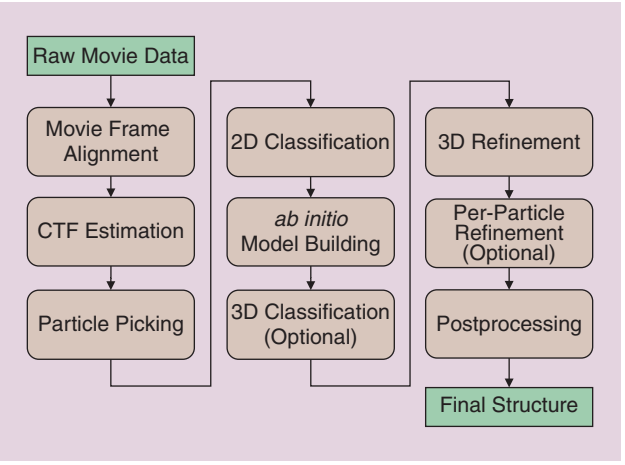

FIGURE 4. A flowchart diagram showing the computational pipeline required to convert raw movie data into high-resolution structures by single-particle cryo-EM. (Adapted from [90].)

Figure 4 shows the complete cryo-EM imaging pipeline. Briefly, raw data are first preprocessed at the movie frame alignment ("Motion Correction" section) and CTF estimation ("CTF Estimation and Correction" section) steps, followed by particle picking to detect and extract the individual projections from micrographs. Occasionally, particle picking is followed by a particle-pruning stage to remove noninformative picked particles ("Particle Picking" section). The output of this stage is a set of 2D images; each (ideally) contains a noisy tomographic projection taken from an unknown viewing direction.

Particle images are then classified ("2D Classification" section): the 2D classes may be used to construct *ab initio* models, as templates for particle picking, to provide a quick assessment of the particles and for symmetry detection. Then, an *ab initio* model is built using the 2D classes or by alternative techniques ("Ab Initio Modeling" section). If the data contain structural variability or a mix of structures (see the "The Cryo-EM Reconstruction Problem" and "Conformational Heterogeneity: Modeling and Recovery" sections), then a 3D classification step is applied to cluster the projection images into the different structural conformations.

Initial models are subjected to 3D high-resolution refinement ("High-Resolution Refinement" section), and an additional per-particle refinement may be applied. Finally, a postprocessing stage is employed to facilitate interpretation of structures in terms of atomic models. Different software packages may use slightly different workflows and, occasionally, some of the steps are applied iteratively. For instance, one can use the 2D classes to repeat particle picking with more reliable templates.

Let $\phi\colon \mathbb{R}^3 \to \mathbb{R}$ represent the 3D molecular structure to be estimated. Under the assumption that the particle picking is executed well (i.e., each particle is detected up to a small translation; we assume for simplicity no false detection), each image $I_1, \ldots, I_N$ is formed by rotating ϕ by a 3D rotation R_ω, integrating along the z-axis (tomographic projection), shifting by a 2D shift T_t, convolving with the PSF of the microscope h_i, and adding noise:

$$I_i = h_i * T_{t_i}\left(\int_{-\infty}^{\infty} \phi(R_{\omega_i}^T r)\,dz\right) + \text{"noise,"} \quad r = (x, y, z), \quad i = 1, \ldots, N, \tag{1}$$

where $*$ denotes convolution. The model can be written more concisely as

$$I_i = h_i * T_{t_i} P R_{\omega_i} \phi + \text{"noise,"} \quad i = 1, \ldots, N, \tag{2}$$

where the tomographic projection is denoted by P, and $(R\phi)(r) = \phi(R^T r)$. The image is then sampled on a Cartesian grid. The integration along the z-axis is called the *X-ray transform* (not to be confused with the radon transform in which the integration is over hyperplanes). The rotations R_ω describe the unknown 3D orientation of the particles embedded in the ice, and they can be thought of as random elements of the special orthogonal group $\omega \in SO(3)$. These rotations can be represented as 3×3 orthogonal matrices with determinant one or by quaternions. The 2D shifts $t \in \mathbb{R}^2$ result from detection inaccuracies, which are usually small. The PSFs h_i are assumed to be known, and their Fourier transforms suffer from many zero crossings, making deconvolution challenging; see the "CTF Estimation and Correction" section for further discussion.

The cryo-EM inverse problem of reconstruction consists of estimating the 3D structure ϕ from the 2D images $I_1, \ldots, I_N$. Importantly, the 3D rotations $\omega_1, \ldots, \omega_N$ and the 2D shifts $t_1, \ldots, t_N$ are called *nuisance variables*; although the rotations and shifts are unknown a priori, their estimation is not an aim by itself. Figure 5 shows an example of the cryo-EM problem in its most simplified version, without noise, CTF, and shifts.

The reconstruction of ϕ is possible with up to three intrinsic ambiguities: a global 3D rotation, the position of the center of the molecule (3D location), and handedness. This last symmetry, also called *chirality*, means that it is impossible to distinguish whether the molecule was reflected about a 2D plane through the origin. The handedness of the structure cannot be determined from cryo-EM images alone because the original 3D object and its reflection give rise to identical sets of projections related by the following conjugation: $\tilde{R}_{\omega_i} = J R_{\omega_i} J^{-1}$, where $J = \mathrm{diag}(1, 1, -1)$.

In the presence of structural variability, ϕ may be thought of as a random signal with an unknown distribution defined over a space of possible structures (which might be unknown as well). In this case, the task is more ambitious and involves estimating the whole distribution of conformations. Usually, the distribution is assumed to be discrete (i.e., in each measurement, we observe one of a few possible structures or conformational states) or to lie in a low-dimensional subspace or manifold. This subject is further discussed in the "Conformational Heterogeneity: Modeling and Recovery" section and in a recent survey [74].

The chief noise source in cryo-EM at the frame level (before motion correction) is shot noise, which follows a Poisson distribution. After movie frames are averaged to produce micrographs, it is customary to assume that the noise is characterized by a Gaussian distribution. Indeed, all current algorithms build on—implicitly or explicitly—the speculated Gaussianity of the noise. While the spectrum of the noise is not white—mainly due to inelastic scattering, variations in the thickness of ice, and the PSF of the microscope—it is assumed to be a 1D radial function (i.e., constant along the angular direction). The common practice is to estimate the parameters of the noise power spectrum during the 3D reconstruction process or from regions in the micrographs that, presumably, contain no signal.

Main computational challenges

The difficulty in determining high-resolution molecular structures using cryo-EM hinges on three characteristic features of the cryo-EM data: high noise level, missing data, and its massive size. This section elaborates on these unique features, while the next sections dive into the different tasks and algorithms involved in cryo-EM data processing.

High noise level

In a cryo-EM experiment, the electron doses must be kept low to mitigate radiation damage due to electron illumination. The low doses induce high noise levels on the acquired raw data frames. Together with the image's low contrast, this results in SNR levels that are usually well below 0 dB and might be as low as −20 dB (i.e., the power of the noise is 100 times greater than the power of the signal). Under such low-SNR conditions, standard tasks, such as aligning, detecting, or clustering signals, become very challenging.

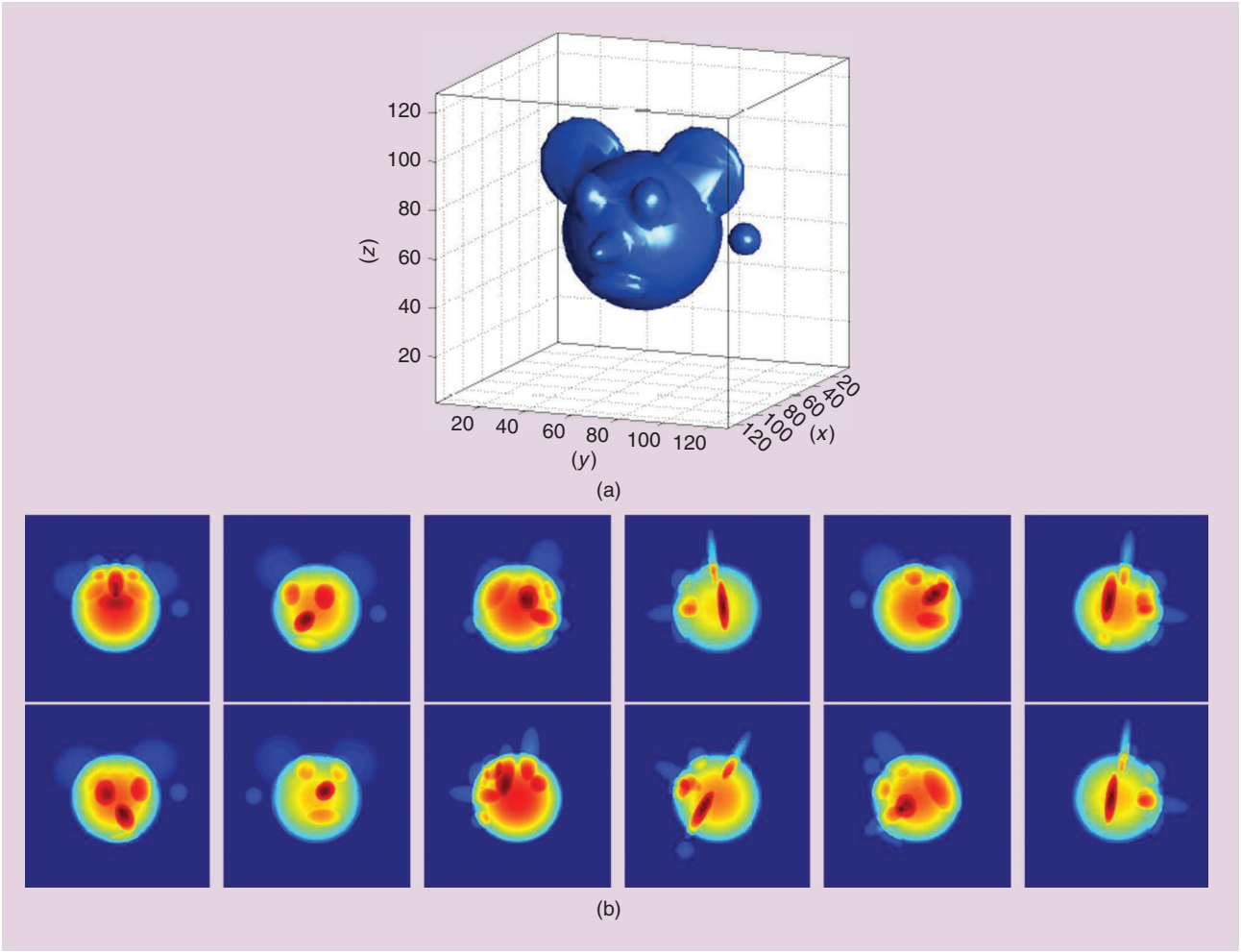

FIGURE 5. (a) A simulated 3D structure and (b) a dozen of its noise-free tomographic projections from different viewing directions. The most simplified version of the cryo-EM problem is estimating the 3D structure from the 2D projection images when the viewing directions are unknown. In practice, the projections are highly noisy, slightly shifted, and convolved with the microscope's PSF. (Used with permission from [71].)

To comprehend the difficulty of performing estimation tasks in a low-SNR environment, let us consider a simple rotation-estimation problem. Let us denote by $x \in \mathbb{R}^L$ the samples of a periodic signal on the unit circle; x is assumed to be known. We wish to estimate a rotation $\theta \in [0, 2\pi)$ from a noisy measurement

$$y = R_\theta x + \varepsilon, \tag{3}$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$, and R_θ denotes the rotation operator, that is, $(R_\theta x)(t) = x(t - \theta)$. To estimate θ, we correlate the signal x with rotated versions of y and choose the inverse of the rotation that maximizes the correlation as the estimator; this technique is called *template matching*. In the absence of noise, template matching simply correlates the signal with its rotated versions; the maximum is attained when y is rotated by $-\theta$.

However, in the presence of noise, we get an additional stochastic term due to the correlation of the noise with the signal. Consequently, if σ^2 is large, it is likely that the peak of the correlation will not be close to $-\theta$. In particular, when $\sigma \to \infty$, the location of the peak is distributed uniformly on the circle. This result can be derived formally using the Neyman–Pearson lemma and holds true for any estimation technique, not necessarily template matching [18]. Even if we collect N measurements (each with a different rotation), it is impossible to estimate the N rotations accurately. In [6], it was shown that the Cramér–Rao bound of this problem is proportional to σ^2 and independent of N. Therefore, if the noise level is high, the variance of any estimator will be high as well.

The same conclusions that were derived for the simple rotation estimation problem (3) remain true for cryo-EM. For example, even if the 3D structure is known, aligning a noisy raw image against multiple noiseless projection templates that correspond to rotations sampled from $SO(3)$ will produce no salient peak in the correlation if the noise level is high. In particular, the higher the noise level, the flatter the distribution over $SO(3)$.

Missing data—Unknown viewing directions and locations

The viewing direction and location associated with each particle in a micrograph are unknown a priori. If they were known, estimating the structure ϕ would be a linear inverse problem, similar to the reconstruction problem in computerized tomography

(CT). The recovery in this case is based on the Fourier slice theorem, which states that the 2D Fourier transform of a tomographic projection is the restriction of the 3D Fourier transform of ϕ to a 2D plane. Mathematically, it can be written succinctly as

$$SRF_3\phi = F_2PR\phi, \tag{4}$$

where F_2, F_3, P, and S are, respectively, the 2D Fourier, 3D Fourier, tomographic projection, and restriction operators; this theorem is a direct corollary of the fact that the Fourier operator F and the rotation operator R commute, that is, $RF = FR$.

The Fourier slice theorem implies that acquiring tomographic projections from known viewing directions is equivalent to sampling the 3D Fourier space. Hence, given enough projections, one can estimate ϕ to a certain resolution. This is the underlying principle behind many CT imaging algorithms, such as the classical filtered-back projection (FBP). However, FBP is not optimal for cryo-EM (even if viewing directions are known) since not all viewing directions are necessarily represented in the data, and the sampling in the Fourier domain is nonuniform; the coverage of viewing directions affects the quality of the solution.

An alternative is the algebraic reconstruction technique (ART), which solves the linear inverse problem by iterative projections; this algorithm is called the *Kaczmarz method* in numerical linear algebra. However, ART is rarely used in cryo-EM because it is slow: it does not exploit the Fourier slice theorem and, thus, cannot be accelerated using fast Fourier transforms (FFTs). Modern approaches were developed to exploit the Fourier slice theorem and account for nonuniform sampling; popular algorithms are based on efficient gridding methods (that compute a uniformly sampled version of a function from a nonuniformly sampled version by choosing proper weights) [55] and using nonuniform FFT packages to harness the structure of the linear system [14], [35].

While the viewing directions in cryo-EM are unknown, there is a rigorous technique to estimate them based on common lines; see the "Estimating Viewing Directions Using Common Lines" section as well as [72] and [79]. Unfortunately, any method for estimating the viewing directions is destined to fail when the SNR is low, for the same reasons that estimating θ in (3) would fail. Bearing in mind that the ultimate goal is to estimate the 3D structure—not the viewing directions—it is essential to consider statistical methods that circumvent rotation estimation, such as the maximum likelihood (ML) and the method of moments.

Massive data sets and high dimensionality

A single session of data collection in a typical cryo-EM experiment produces a few thousand micrographs, each containing several hundred individual particle projections. Depending on the type of detector used during acquisition, micrographs can range in size from a few tens of megapixels up to 100 megapixels for the newest detectors. Moreover, the new generation of detectors can record each micrograph as a rapid burst of frames, producing large movie files that result in data sets of several terabytes in size. The sheer volume of data must be taken into account early on in the algorithmic design process, and in addition to storage considerations, steps must be taken to ensure the efficient use of computational resources to keep up with the ever-increasing throughput of data produced by modern cameras.

Another issue is the dimensionality of the reconstruction problem. The number of voxels of a typical $200 \times 200 \times 200$ volume is 8 million. Estimating so many parameters poses a challenge, both from the computational complexity side (for example, how to find the maximum of the likelihood function of 8 million variables) and also on the statistical-estimation front (see the "Denoising and Dimensionality Reduction Techniques" section). The problem is even more severe when multiple structures, or even a continuum of structures, need to be estimated (see the "Remaining Computational and Theoretical Challenges" section). Getting to high resolution is a major bottleneck due to a compounding computational burden effect: as more parameters need to be estimated, more data are required for their estimation, and the computation becomes ever more expensive and challenging.

3D reconstruction from projections

High-resolution refinement

The reconstruction procedure of high-resolution 3D structures is usually split into two stages: constructing an initial low-resolution model, which is later refined by applying an iterative algorithm ("refinement algorithm"); see Figure 4. This section is devoted to refinement techniques, while different approaches to constitute low-resolution estimates using *ab initio* modeling—initialization-free models—are discussed in the next section.

Refinement techniques for cryo-EM can be broadly classified into two categories: hard and soft angular-assignment methods. The hard-assignment approach is based on template matching. At each iteration, multiple projections are generated from the current estimate of the structure; the projections should, ideally, densely cover $SO(3)$. Then, a single viewing direction is assigned to each experimental image based on the projection with which it correlates best. The angular assignment can be performed either in real or Fourier space.

The advantage of working in Fourier space is that it is not necessary to rotate the molecule: projections can be computed fast using off-the-shelf nonuniform FFT packages (e.g., [14]and [35]), as implied by the Fourier slice theorem. Moreover, with this representation, the CTF is simply a diagonal operator. Once the viewing directions of all experimental images are assigned, the 3D structure is constructed using standard linear-inversion techniques; see the discussion in the "Main Computational Challenges" section. The algorithm iterates between hard angular assignment and structure construction until convergence. Although the quality of hard angular assignments may be influenced by the high noise levels, there are several examples of packages that follow this type of approach, including EMAN2 [76] and cisTEM [33], among others.

A second class of strategies comprises the soft-assignment methods. Similar to hard-assignment methods, in each iteration, the experimental images are correlated with multiple template projections. However, instead the best match being chosen, a similarity score is given to each pair of experimental image and projection. These scores, also called *weights* or *responsibilities*, are computed according to the generative statistical model of the experimental images and used for reconstruction, as explained later in this section.

The soft assignment methods are instances or variants of the well-known expectation-maximization (ML-EM) algorithm [25]. Notably, this relation classifies the 3D reconstruction problem as a problem of ML estimation or, more generally, a problem in Bayesian statistics and, thus, provides a solid theoretical and algorithmic framework. In particular, it enables the incorporation of priors that, essentially, act as regularizers in the reconstruction process. In the context of cryo-EM, ML-EM was first applied to 2D cryo-EM images by Sigworth [68] and was later implemented for 3D reconstruction by the software packages RELION [63] and cryoSPARC [57].

In what follows, we describe the ML-EM algorithm for a more general statistical model and then particularize it to the special case of cryo-EM. Suppose we collect N observations from the model:

$$y_i = L_{\theta_i} x + \varepsilon_i, \quad i = 1, \ldots, N, \tag{5}$$

where L_θ is a linear transformation acting on the signal x, parameterized by a random variable θ, and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. The goal is to estimate x, while $\theta_1, \ldots, \theta_N$ are nuisance variables. A typical assumption is that x can be represented by a finite number of coefficients (for instance, its Fourier expansion is finite) and that these coefficients were drawn from a Gaussian distribution with mean μ and covariance matrix Σ. The key for reliable estimation is to marginalize over the nuisance variables $\theta_1, \ldots, \theta_N$. Without marginalization (i.e., when the goal is to jointly estimate x and $\theta_1, \ldots, \theta_N$), the number of parameters grows with the number of measurements indefinitely, and, thus, the ML estimator may be inconsistent. This phenomenon is exemplified by the Neyman–Scott "paradox."

In cryo-EM, the transformation L_θ is the operator described in (2). This operator rotates the volume, computes its 2D tomographic projection, and applies a 2D translation and convolution with the PSF; θ is drawn from a distribution defined over the 5D space of 3D rotations and 2D translations. The distribution of θ is generally unknown and should be estimated as part of the ML-EM algorithm.

Let us denote $\mathbf{y} = (y_1, \ldots, y_N)$. The posterior distribution $p(x|\mathbf{y})$ of (5) is proportional to the product of the prior $p(x)$ with the likelihood function

$$\begin{aligned}\mathcal{L}(x; \mathbf{y}) = p(\mathbf{y}|x) &= \prod_{i=1}^{N} \sum_{\theta_\ell \in \Theta} p(y_i|x, \theta_\ell) p(\theta_\ell) \\ &= \frac{1}{(2\pi\sigma^2)^{M/2}} \prod_{i=1}^{N} \sum_{\theta_\ell \in \Theta} p(\theta_\ell) e^{-\frac{1}{2\sigma^2}\|y_i - L_{\theta_\ell} x\|^2},\end{aligned} \tag{6}$$

where Θ is a discrete space on which θ is defined, and M is the length of each observation. If the entire posterior distribution could be computed, then one could obtain far more statistical information about x than just maximizing the likelihood. For instance, the best estimator in the minimum mean square error (MMSE) sense is obtained by marginalizing over the posterior:

$$\hat{x}_{\text{MMSE}}(\mathbf{y}) = \int_x p(x|\mathbf{y}) x dx = \mathbb{E}\{x|\mathbf{y}\}. \tag{7}$$

Unfortunately, the entire posterior can rarely be computed, especially in big data problems, such as cryo-EM. The ML-EM framework provides a simple, yet frequently very effective, iterative method that tries to compute the maximum a posteriori estimator (MAP), that is, the maximal value of $p(x|\mathbf{y})$.

Each iteration of the ML-EM algorithm consists of two steps. The first step (E-step) computes the expected value of the log of the posterior with respect to the nuisance variables, conditioned on the current estimate of x, denoted here by x_t, and the data

$$\begin{aligned}Q(x|x_t) &= \mathbb{E}_{\theta|\mathbf{y}, x_t}\{\log p(x|\mathbf{y}, \theta)\} \\ &= \text{constant} - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{\theta_\ell \in \Theta} w_{i,\ell} \|y_i - L_{\theta_\ell} x\|^2 + \log p(x),\end{aligned} \tag{8}$$

where

$$w_{i,\ell} = p(\theta_i = \theta_\ell | y_i, x_t) = \frac{e^{\frac{-1}{2\sigma^2}\|y_i - L_{\theta_\ell} x_t\|^2}}{\sum_{\theta_\ell \in \Theta} e^{\frac{-1}{2\sigma^2}\|y_i - L_{\theta_\ell} x_t\|^2}}. \tag{9}$$

If $x \sim \mathcal{N}(0, \Sigma)$, then $\log p(x) = -(1/2) x^T \Sigma^{-1} x$.

The second step (M-step) updates x by

$$x_{t+1} = \operatorname*{argmax}_x Q(x|x_t), \tag{10}$$

which is usually performed by setting the gradient of Q to zero. If x is assumed to be Gaussian, then the M-step reduces to solving a linear system of equations with respect to x. The E- and M-steps are applied iteratively. It is well known that $p(x_{t+1}|\mathbf{y}) \geq p(x_t|\mathbf{y})$; however, since the landscape of $p(x|\mathbf{y})$ is usually nonconvex, the iterations are not guaranteed to converge to the MAP estimator [25]. Usually, the ML-EM iterations halt when $(p(x_{t+1}|\mathbf{y}) - p(x_t|\mathbf{y}))/p(x_t|\mathbf{y})$ is smaller than some tolerance (but other criteria can be employed as well). The posterior distribution at each iteration can be evaluated according to (6).

The implementation details of the ML-EM algorithm for cryo-EM vary across different software packages. The popular package RELION, for example, incorporates a prior of uniform distribution of rotations over $SO(3)$ (although the distribution itself is usually nonuniform), and each Fourier coefficient of the 3D volume was drawn independently from a normal distribution [63]. The variance of the Fourier coefficients' prior is updated at each iteration by averaging over

concentric frequency shells of the current estimate (resulting in a 1D radial function). If the ML-EM algorithm is initialized with a smooth volume (which is the common practice), the variance is large only at low frequencies; hence, high frequencies are severely regularized. As a result, the first ML-EM iterations mostly update the low frequencies, while the high frequencies are only slightly affected.

As the algorithm proceeds, the structure includes more high-frequency content; thus, the variance of these frequency shells increases, and the effective resolution of the 3D map gradually improves. This strategy of initializing with a low-resolution structure and gradually increasing the resolution is called *frequency marching* and is shared by many packages in the field; see further discussion in the "Frequency Marching" section. Another popular package, cryoSPARC, also employs an ML-EM algorithm with the prior that each voxel in real space was drawn independently from a Poisson distribution with a constant parameter [57].

We mention that such statistical priors (e.g., each voxel or frequency is drawn from Poisson or normal distributions) are usually chosen out of mathematical and computational convenience and may bias the reconstruction process. It is an open challenge to replace those statistical priors by priors that are based on or inspired by biological knowledge, that is, previously reconstructed structures.

The main drawback of the ML-EM approach, especially at high resolution, is the high computational load of the E-step in each iteration. Specifically, each ML-EM iteration requires correlating each experimental image (typically a few hundred thousand) with multiple synthetic projections of the current structure estimate [sampled densely over $SO(3)$] and their 2D translations. To alleviate the computational burden, a variety of methods are employed to narrow the 5D search space. Popular techniques are based on multiscale approaches, whereby an initial search is done on a coarse grid, followed by local searches on finer grids. A more sophisticated idea was suggested in [57], based on the "branch-and-bound" methodology, that rules out regions of the search space that are not likely to contain the optimum of the objective function.

Ab initio *modeling*

In this section, we describe some of the intriguing ideas that were proposed to construct *ab initio* models, that is, models that do not require an initial guess. These methods usually result in low-resolution estimates that can be later refined as described in the previous section. To emphasize the importance of robust *ab initio* techniques, we begin by discussing model bias—a phenomenon of crucial importance in cryo-EM and statistics in general.

Model bias and validation

The cryo-EM inverse problem is nonconvex and, thus, the output of 3D reconstruction algorithms may depend on their initializations. This raises the validation problem: how can we verify that a given estimate is a faithful representation of the underlying data? Currently, a 3D model is treated as valid if its structural features meet the common biological knowledge (primary structure, secondary structure, and so on) and if it passes some computational tests based on different heuristics. For instance, the reconstruction algorithm can be initialized from many different points. If the algorithm always converges to the same or similar structures, it strengthens the confidence in the attained solution. Moreover, since the data are usually uploaded to public repositories, different researchers can examine it and compare the results against each other [61]; see a more detailed discussion in the "Verification" section.

Despite all precautions taken by researchers in the field, the verification methods are not immunized against systemic errors—this pitfall is dubbed *model bias*. One important example concerns the particle-picking stage, when one aims to detect and extract particle projections from noisy micrographs. The majority of the detection algorithms in the field are based on template-matching techniques, despite their intrinsic flaws. Specifically, choosing improper templates can lead to erroneous detection, which, in turn, biases the 3D reconstruction algorithm toward the chosen templates; see [39] and further discussion on particle-picking techniques in the "Particle Picking" section.

The model bias phenomenon is exemplified by the "Einstein from noise" experiment [67]. In this experiment, N images of pure independently identically distributed (i.i.d.) Gaussian noise are correlated with a reference image. (In the original article, the authors chose an image of Einstein as the reference and, thus, the name.) Then, each pure-noise image is shifted to best align with the reference image (based on the peak of their cross correlation), and, finally, all images are averaged.

Without bias, one would expect that averaging pure-noise images would converge toward an image of zeros as N diverges. However, in practice, the resulting image is similar to the reference image, that is, the algorithm is biased toward the reference image; see [67, Fig. 2]. In the context of cryo-EM, the lesson is that, without prudent algorithmic design, the reconstructed molecular structure may reflect the scientist's presumptions rather than the structure that best explains the data. We now turn our attention to some of the existing techniques for *ab initio* modeling.

Stochastic gradient descent over the likelihood function

Stochastic gradient descent (SGD) has gained popularity in recent years, especially due to its invaluable role in the field of deep learning [21]. The underlying idea is very simple. Suppose that we wish to minimize an objective function of the form

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x). \tag{11}$$

A gradient descent algorithm is an optimization technique used to minimize the objective function by iteratively moving in the direction opposite to its gradient. Its tth iteration takes on the form

$$x_{t+1} = x_t - \eta_t \frac{1}{N}\sum_{i=1}^{N} \nabla f_i(x_t), \tag{12}$$

where ∇ denotes a gradient and η_t is called the *step size* or *learning rate*. In SGD, we simply replace the full sum by a random element:

$$x_{t+1} = x_t - \eta_t \nabla f_i(x_t), \tag{13}$$

where i is chosen uniformly at random from $\{1, \ldots, N\}$. In expectation, the direction of each SGD step is the direction of the full gradient (12). The main advantage of SGD is that each iteration is significantly cheaper to compute than the full gradient (12). It is also easy to modify the algorithm to sum over a few random elements at each iteration, rather than just one, to improve robustness at the cost of higher computational load per iteration. In addition, it is commonly believed that the inherent randomness of SGD helps to escape local minima in some nonconvex problems.

This strategy was proven to be effective for constructing *ab initio* models using cryo-EM, achieving resolutions of up to 10 Å [57]. The objective function is the negative log of the posterior distribution. As the log posterior involves a sum over the experimental images (6), the SGD scheme chooses, at each iteration, one random image (or a subset of images) to approximate the gradient direction.

Estimating viewing directions using common lines

If the orientations and locations of the particles in the ice are known [equivalently, the 3D rotations and 2D translations in (2)], then recovering the 3D structure is a linear problem that can be solved using one of the many solutions developed for CT imaging. In this section, we survey an analytical method for estimating the viewing directions using the common-lines property.

Due to the Fourier slice theorem, any pair of projection images has a pair of central lines (one in each image) on which their Fourier transforms agree. For generic molecular structures (with no symmetry), it is possible to uniquely identify this common line, for example, by cross-correlating all possible central lines in one image with all possible central lines in the other image and choosing the pair of lines with maximum cross correlation.

The common line pins down two out of the three Euler angles associated with the relative rotation $R_i^{-1} R_j$ between images I_i and I_j. A third image is required to determine the third angle: the three common-line pairs between the three images uniquely determine their relative rotations. This technique is called *angular reconstitution*, and it was suggested, independently, by Vainshtein and Goncharov [78] and Van Heel [79]. The main drawback of this procedure is its sensitivity to noise; it requires the three pairs of common lines to be accurately identified. Moreover, estimating the rotations of additional images sequentially (using their common lines with the previously rotationally assigned images) can quickly accumulate errors.

As a robust alternative, it was proposed that the rotations $R_1, \ldots, R_N$ be estimated from the common lines of all pairs I_i, I_j simultaneously; this framework is called *synchronization*. Since this strategy takes all pairwise information into account, it has better tolerance to noise. The rotation assignment can be done using a spectral algorithm or semidefinite programming [72] and enjoys some theoretical guarantees; see, for instance, [70]. Nevertheless, as discussed in the "Main Computational Challenges" section, in a low-SNR regime, any method for estimating the rotations would fail. Therefore, the method requires as input high-SNR images that can be obtained using a procedure called *2D classification*; see further discussion in the "2D Classification" section. However, 2D classification blurs the fine details of the images, and, therefore, the attained resolution of the method is limited [34].

Three last comments are in order. First, if the structure possesses a nontrivial symmetry, there are multiple common lines between pairs of images that should be considered [56]. Second, interestingly, this technique cannot work when the underlying object is a 2D image (rather than 3D, as in cryo-EM). In this case, the Fourier transform of the tomographic projection (which is a 1D function in that case) is a line that goes through the origin of the 2D Fourier space of the underlying object. Therefore, the single common point of any two projections taken from different angles is the origin, and, thus, there is no way to find the relative angle between them. Finally, the common-lines method is effective for additional signal processing applications, such as the study of specimen populations [46].

The method of moments

Suppose that a set of parameters x characterizes a distribution $p(y|x)$ of a random variable y with L entries. We observe N i.i.d. samples of y and wish to estimate x. In the method of moments, the underlying idea is to relate the moments of the observed data with x. Those moments can be estimated from the data by averaging over the observations

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} y_i &\approx \mathbb{E}\{p(y|x)\} = M_1(x)\\
\frac{1}{N}\sum_{i=1}^{N} y_i y_i^T &\approx \mathbb{E}\{p(yy^T|x)\} = M_2(x)\\
&\vdots\\
\frac{1}{N}\sum_{i=1}^{N} y_i^{\otimes k} &\approx \mathbb{E}\{p(y^{\otimes k}|x)\} = M_k(x),
\end{aligned} \tag{14}$$

where $y^{\otimes k}$ is a tensor with L^k entries, and the entry indexed by $n = (n_1, \ldots, n_k) \in \mathbb{Z}_L^k$ is given by $\prod_{i=1}^{k} y(n_i)$. By the law of large numbers, when $N \to \infty$, the average converges almost surely to the expectation. The right-hand side underscores that the expectations are solely functions of x; the moment tensors $M_1, M_2, \ldots, M_k$ can be derived analytically. For instance, in cryo-EM, the expectation is taken against the random rotations, translations, and noise; the moments are functions of the 3D volume, the 5D distribution over the space of 3D rotations and 2D translations, and some parameters of the noise statistics (due to bias terms). The final step of the method of moments is to solve the system of equations (14), which might be highly nontrivial. Importantly, since estimating the moments requires only one pass over the observations, it can be done potentially

on the fly during data acquisition. This is especially important for cryo-EM, in which hundreds of thousands of images must be processed.

Computing the qth moment involves the product of q noisy terms and, thus, the variance of its estimation scales as σ^{2q}/N. As a result, when invoking the method of moments, it is crucial to determine the lowest-order moment that identifies the parameters x uniquely: this moment determines the *estimation rate* of the problem, that is, how many samples are required to accurately estimate the parameters. Remarkably, it was shown that under a general statistical model that is tightly related to the cryo-EM problem (as will be described in the "MRA" section), the first moment that distinguishes between different parameters determines the sample complexity of the problem (i.e., how many observations are necessary to attain an accurate estimate) in the low-SNR regime [4]. In addition, the number of equations in (14) increases with the order of moments. Thus, using lower-order moments reduces the computational complexity.

The first to suggest the method of moments for single particle reconstruction, 40 years ago, was Zvi Kam [42]. Under the simplifying assumptions that the particles are centered, the rotations are uniformly distributed, and, by ignoring the CTF, he showed that the second moment of the experimental images determines the 3D structure, up to a set of orthogonal matrices. To determine those matrices, he suggested computing a subset of the third-order moment. Recently, it was shown that under the same assumptions, the structure is indeed determined uniquely from the third moment [9]. Remarkably, if the distribution of rotations is nonuniform, then the second moment suffices [66]. Therefore, in light of [4], the sample complexity of the cryo-EM problem in the low-SNR regime, under the specified conditions, scales as σ^4 and σ^6 for nonuniform and uniform distribution of rotations, respectively. It is still an open question as to how many images are required to solve the full cryo-EM problem; see further discussion in the "Remaining Computational and Theoretical Challenges" section.

Interest in the method of moments has been recently revived, largely due to its potential application to X-ray free-electron lasers [48], [82]. We believe that the method of moments has been largely overlooked by the cryo-EM community since Kam's article for three main reasons. First, Kam's original formulation required uniform distribution of viewing directions, an assumption that typically does not hold in practice; a recent article extends the method to any distribution over $SO(3)$ [66]. Second, estimating the second- and third-order statistics accurately requires a large amount of data that was not available until recent years. Third, accurate estimation of high-order moments for high-dimensional problems requires modern statistical techniques, such as eigenvalue shrinkage in the spiked covariance model, that have been introduced only in the last decade [26].

Frequency marching

Most of the cryo-EM reconstruction algorithms start with a low-resolution estimate of the structure, which is then gradually refined to higher resolutions. This process is dubbed *frequency marching*, and it can be done explicitly by constructing a low-resolution model using an *ab initio* technique that is later refined by ML-EM [57] or, more implicitly, by an iteration-dependent regularization [63].

In [13], a deterministic and mathematically rigorous method to gradually increase the resolution was proposed. The 3D Fourier transform of the object is expanded by concentric shells. At each iteration, each experimental image is compared with many simulated projections of the current low-resolution estimate, and a viewing direction is assigned to each experimental image by template matching (hard angular assignment). Given the angular assignments, an updated structure with one more frequency shells is computed by solving a linear least squares problem, thus progressively increasing the resolution.

Building blocks in the computational pipeline

This section elaborates on some of the building blocks in the algorithmic pipeline of single-particle reconstruction using cryo-EM. For each task, we introduce the problem and discuss the underlying principles behind some of the most commonly used solutions.

Motion correction

During data collection, the electron beam induces sample motion that mitigates high-resolution information. The modern detectors (DEDs) acquire multiple frames per micrograph, allowing partial correction of the motion blur by aligning and averaging the frames. In essence, motion correction is an alignment problem (also referred to as *registration* or *synchronization*) that shares many similarities with classical tasks in signal processing and computer vision. The main challenge in alignment is the high noise levels, which hamper the precise estimation of relative shifts between frames.

Several solutions were proposed to the motion-correction problem based on a variety of methods, such as pairwise alignment of all frames and optical flow; see a survey on the subject in [58] and the references therein. The first-generation solutions aimed to estimate the movement of the entire micrograph; however, the drift is not homogeneous across the entire field of view, motivating the development of more accurate local techniques. One such strategy is implemented by the software MotionCor2 [88], which describes the motion as a block-based deformation that varies smoothly throughout the exposure.

The micrograph is first divided into patches, and motions within each patch are estimated based on cross correlation. Then, the local motions are fitted to a time-varying 2D polynomial, which is quadratic in the 2D frame coordinates and cubic in the time axis. Finally, the frames are summed, with or without dose weighting, which was determined according to a radiation-damage analysis. More recently, strategies for per-particle motion correction have been proposed that yield significant improvements in resolution [15], [92]. Figure 6 shows examples of different motion-correction strategies.

CTF estimation and correction

Each cryo-EM image is affected by the PSF of the microscope through convolution with a kernel h_i; see (2). The Fourier

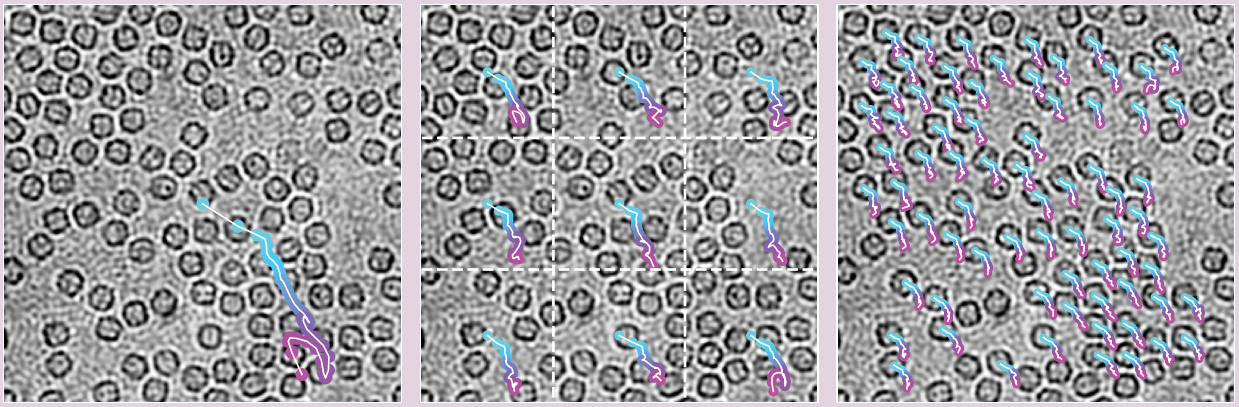

FIGURE 6. Examples of beam-induced motion correction. All movement trajectories in this figure are color coded, with cyan representing the position of the first frame and magenta indicating the position of the last frame in the sequence. (a) Strategies for global motion correction compensate for movement of the specimen across the entire field of view containing multiple particles. (Used with permission from [32].) (b) Semilocal strategies for motion correction align frames across subregions defined on a discrete grid. (Used with permission from [88].) (c) Per-particle or local drift correction allows accurate tracking of individual particles throughout the exposure to electrons. (Used with permission from [15].)

transform of the PSF—the CTF—suffers from multiple zero crossings, making its inversion (i.e., deconvolution) challenging. To compensate for the missing frequencies, the standard routine is to apply different PSFs to different micrographs. Thus, spectral information that was suppressed in one micrograph may appear in another.

The CTF cannot be specified precisely by the user. Instead, the number and location of the zero crossings is given by the specific experimental parameters of the microscope, which are hard to control. Thus, a preceding step entails estimating the CTF from the acquired data. Specifically, the CTF is modeled as

$$\mathrm{CTF}(k,\lambda,\Delta f,C_s) = -\sin\left(\pi\lambda|k|^2\Delta f - \frac{\pi\lambda^3|k|^4 C_s}{2} + \alpha\right)\cdot E(|k|), \tag{15}$$

where k is the frequency index, α is a small phase-shift term, λ is the electron wavelength, Δf is the objective defocus, C_s is the spherical aberration, and $E(|k|)$ is an exponentially decaying envelope function—specified by a parameter called *B-factor*—due to the beam energy spread, beam coherence, and sample drift [59]. The CTF mitigates the very low frequencies, and, therefore, centering projection images (i.e., finding their center of mass) is challenging. In fact, the center of mass depends on the CTF, and, thus, two projection images with the same viewing angles but different CTFs (e.g., different defocus values) will have different centers in real space.

The CTF is approximately constant along concentric rings, although, in practice, those rings might be slightly deformed; this deformation can be modeled by an additional parameter called *astigmatism*. The defocus value of the microscope has a pivotal role: high defocus values enhance low-resolution features and improve the contrast of the image (increase the SNR), whereas low defocus values enhance high-resolution features (fine details) at the cost of lower contrast (lower SNR).

CTF estimation begins by computing the power spectrum of the whole micrograph. The power spectrum exhibits "rings," called *Thon rings*, that correspond to the oscillations of the sine function (15); those rings are used to fit the CTF's parameters [see Figure 2(b)]. Two popular software packages, CTFFIND4 [59] and Gctf [85], estimate those parameters by generating multiple templates according to the postulated model (15). The template that best fits the measured Thon rings is used as the estimated CTF.

Once the CTF is estimated, the goal is to invert its action. The challenge stems from the structure of the CTF (15) because it has many small values and, thus, direct inversion is impossible. The most popular technique for CTF correction, "phase flipping," is embarrassingly simple: it disregards the information about amplitude changes and corrects the data only for the sign of the CTF, thereby obtaining the correct phases in Fourier space. Namely, if a micrograph is represented in the Fourier domain by

$$I_{\mathrm{micrograph}}(k) = \mathrm{CTF}(k)I(k), \tag{16}$$

where $I(k)$ is the micrograph before the CTF action, then the corrected image would be

$$I_{\mathrm{corrected}}(k) = \mathrm{sign}(\mathrm{CTF}(k))I_{\mathrm{micrograph}}(k) = |\mathrm{CTF}(k)|I(k). \tag{17}$$

Another approach is to apply Wiener filtering that attempts to correct both the amplitudes and the phases [29], [69]. However, this approach requires knowledge of the SNR (which is not trivial to estimate in a low-SNR environment) and cannot correct for missing frequencies.

Instead of inverting the CTF explicitly, an attractive alternative is to incorporate the CTF into the forward model of the reconstruction algorithm. This can be done naturally in

the ML-EM approach, the method of moments, and the SGD framework. Because different CTFs are applied to different micrographs, in principle, full CTF correction is feasible.

Particle picking

A micrograph consists of regions that contain only noise; those with noisy 2D projections; and those with other contaminants, such as carbon film. In particle picking, the goal is to detect and extract the 2D projections (particles) from the noisy micrographs. High-resolution reconstruction typically requires hundreds of thousands of particles, and, thus, manual picking is time consuming and tedious. In addition, it may introduce subjective bias into the procedure.

Many solutions have been proposed in the literature for the particle-picking problem, based on standard edge-detection techniques, machine learning [83], [84], [91], and template matching. For the last of these, a popular strategy is the one implemented in RELION. In this framework, the user manually selects a few hundred particles. These particle images are then 2D classified (see the next section), and the 2D classes are used as templates for an automatic particle picking based on template matching [64].

As discussed in the "Model Bias and Validation" section, a major concern of any particle-picking algorithm is a systematic bias. To alleviate the risk for model bias, a fully automated technique was proposed in [37]. Rather than templates, the algorithm automatically selects a set of reference windows that include both particle and noise windows. The selection is based on the local mean and variance: regions with particles typically have lower mean and higher variance than regions without particles. Then, regions in the micrograph are correlated with the reference windows. The regions that are most likely to contain a particle (high correlation) and those that are least likely (low correlation with all reference windows) are used to train a support-vector machine classifier. The output of this linear classifier is used to pick the particles.

In practice, a considerable number of picked particles are usually noninformative, containing adjacent particles that are too close to each other, only part of a particle, or just pure noise. Consequently, it is common to try to prune out these outliers. In [62], it was proposed to employ several particle pickers simultaneously and compute a consensus between them using a deep neural network. In [90], a simple pruning strategy is devised by viewing the output of the particle-picking algorithm as a mixture of Gaussians.

2D classification

As experimental images corresponding to similar viewing directions tend to be very much alike (due to the smoothness of the molecular structure), it is common to divide the images into several classes (i.e., clustering) and average them to increase the SNR. This process is called 2D classification, and the averaged images are referred to as *class averages*. Since the global in-plane rotation of each micrograph is arbitrary, the clustering should be invariant under in-plane rotations: two images from a similar viewing direction but with different in-plane rotations are supposed to be grouped together after appropriate alignment. The class averages are used for a variety of tasks: to construct *ab initio* models, as templates for particle picking, to provide a quick assessment of the particles, to remove picked particles that are associated with noninformative classes, and for symmetry detection.

There are three main computational aspects that make the 2D classification task quite challenging. First, as already mentioned, low SNR impedes accurate clustering and alignment. Second, the high computational complexity of finding the optimal clustering among hundreds of thousands of images may be prohibitive unless designed carefully. Third, it is not clear what is the appropriate similarity metric to accurately compare images.

Many different strategies for 2D classification were proposed; see, for instance, [73] and [80]. A popular 2D classification algorithm, implemented in RELION [65], is based on the ML-EM scheme. In this algorithm, each observed image is modeled as a sample from the statistical model

$$y_i = h_i * T_{t_i} R_{\theta_i} x_{k_i} + \varepsilon_i, \quad i = 1, \ldots, N, \tag{18}$$

where k is distributed uniformly over $\{1, \ldots, K\}$ (these are the K class averages to be estimated), θ is distributed uniformly over $\theta \in [0, 2\pi)$, t is drawn from an isotropic 2D Gaussian, h_i is the estimated PSF, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 I)$. Given this model, it is straightforward to implement an ML-EM algorithm to estimate $x_1, \ldots, x_K$ following the guidelines of the "3D Reconstruction From Projections" section.

This algorithm, although popular, suffers from three imperative weaknesses. First, the computational burden of running ML-EM is a major hurdle as the algorithm needs to go through all experimental images at each iteration. Second, it assumes that each experimental image is originated from only K possible class averages. However, this is an inaccurate model, as the orientations of particles in the ice layer are distributed continuously over $SO(3)$ (in addition to structural variability that may be present). Finally, ML-EM typically suffers from the winner-takes-all phenomenon: most experimental images would correlate well with—and, thus, be assigned to—the class averages that enjoy higher SNRs. As a result, ML-EM tends to output only a few, low-resolution classes. This phenomenon was already recognized in [73] and is also present for ML-EM-based 3D classification and refinement.

An alternative solution was proposed in [87]. In this framework, each image is averaged with its few nearest neighbors, after a proper in-plane alignment. The nearest neighbor search is executed efficiently over the bispectra of the images. The bispectrum is a rotationally invariant feature of an image—that is, it remains unchanged under an in-plane rotation. To reduce the computational complexity and denoise the data, each image is first compressed using a dimensionality-reduction technique, called *steerable principal component analysis* (*PCA*), which is the main topic of the next section.

Denoising and dimensionality-reduction techniques

PCA is a widely used tool for linear dimensionality reduction and denoising, dating back to Pearson [41], [54]. PCA computes the best (in the least squares sense) low-dimensional affine space that approximates the data by projecting the images into the space spanned by the leading eigenvectors of their covariance matrix. In cryo-EM, covariance estimation was introduced by Kam [42] and was used for the first time for dimensionality reduction and image classification by van Heel and Frank [80]. PCA is often used to denoise the experimental images and as part of the 2D classification stage.

Cryo-EM images are equally likely to appear in any in-plane rotation. Consequently, when performing PCA, it makes sense to take all in-plane rotations of each image into account. Luckily, there is a simple way to account for all in-plane rotations without rotating the images explicitly. This can be done by expanding the images in a steerable basis: a basis formed as an outer product of radial functions with an angular Fourier basis; examples are Fourier Bessel, 2D prolate spheroidal wave functions, and Zernike polynomials.

When integrating over all possible in-plane rotations, the covariance matrix of the expansion coefficients enjoys a block-diagonal structure, reducing the computational load significantly: while the covariance matrix of images of size $L \times L$ has L^4 entries, the block-diagonal structure guarantees that merely $O(L^3)$ entries are nonzero. This, in turn, reduces the computational complexity of steerable PCA from $O(NL^4 + L^6)$ to $O(NL^3 + L^4)$: the first term is the cost of computing the sample covariance over N images, and the second term is the cost of the eigendecomposition over all blocks [86].

Classical covariance-estimation techniques usually assume that the number of samples is considerably larger than the signal's dimension. However, this is not the typical case in cryo-EM, in which the dimensionality of the molecule is of the same order as the number of measurements. In this regime, one can take advantage of recent developments in high-dimensional statistics under the "spiked covariance model" [26]. These techniques, based on eigenvalue shrinkage, were recently applied successfully to denoise cryo-EM images [20].

Mathematical frameworks for cryo-EM data analysis

Inspired by the cryo-EM problem, researchers have studied a couple of abstract mathematical models in recent years. These models provide a general framework for analyzing cryo-EM from theoretical and statistical perspectives while removing some of its complications. In what follows, we introduce the models and succinctly describe some intriguing results.

MRA

The MRA model reads [10]

$$y_i = T_i(g_i \circ x) + \varepsilon_i, \quad g_i \in G, \tag{19}$$

where G and T_i are, respectively, a known compact group and linear operators, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 I)$. The signal $x \in \mathbb{R}^L$ is assumed to lie in a known space (say, the space of signals with a finite spectral expansion) on which random elements of G act. The task is to estimate the signal x from the observations $y_1, \ldots, y_N$ while the group elements $g_1, \ldots, g_N$ are unknown. Since the statistics of y are invariant under the action of a constant g on x, the recovery is possible up to left multiplication by some group element $g \in G$—that is, we wish to estimate the orbit of x under the group action.

Figure 7 shows an example of three discrete 1D MRA observations under the group of cyclic shifts $\mathbb{Z}/L$ at different noise levels. If we assume perfect particle picking, then the cryo-EM model (2) is a special case of (19) when G is the group of 3D rotations $SO(3)$ and T is the linear operator that takes the rotated structure, integrates along the z-axis (tomographic projection), convolves with the PSF, and samples it on a Cartesian grid. The MRA model formulates many additional applications, including structure from motion in computer vision [5], localization and mapping in robotics [60], study of specimen populations [46], optical and acoustical trapping [28], and denoising of permuted data [53].

While surveying all recent results about the MRA model is beyond the scope of this article, we wish to present a remarkable information/theoretic result about the sample complexity of the problem (and, therefore, also for cryo-EM under the stated assumptions). In many cases, among them when $T = I$ and the cryo-EM setup [without shifts so that $G = SO(3)$], one can estimate the group elements g_i from the measurements y_i in the high-SNR regime using one of many synchronization techniques (based on spectral methods, semidefinite programming, or nonconvex programming) [70], [72]. Once the group elements are identified, one can estimate the signal by aligning all observations (undo the group actions) and averaging.

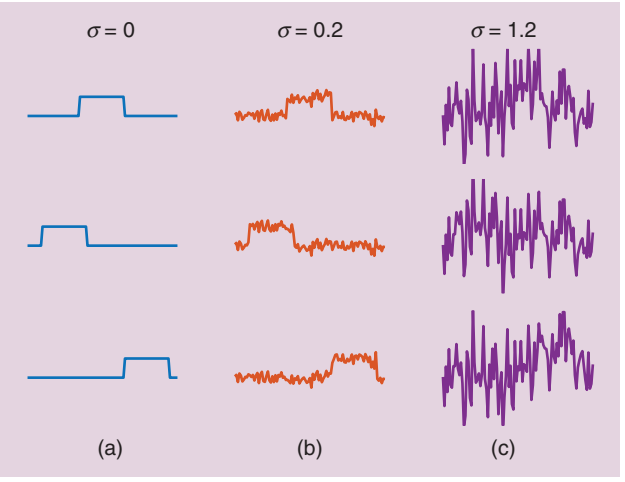

FIGURE 7. An example of MRA observations at different noise levels σ. Each column consists of three different cyclic shifts (the group actions) of a 1D periodic discrete signal (the linear operator T is the identity operator): (a) $\sigma = 0$, (b) $\sigma = 0.2$, and (c) $\sigma = 1.2$. Clearly, if the noise level is low, estimating the signal is easy: one can align the observations (i.e., undo the group action) and average out the noise. The challenge is to estimate the signal when high noise levels hinder alignment, as in (c).

The variance of averaging over i.i.d. Gaussian variables is proportional to $1/(N \cdot \mathrm{SNR})$—that is, the number of measurements should be proportional to 1/SNR to achieve an accurate estimate. In this case, we say that the sample complexity of the problem in the high-SNR regime is 1/SNR: this is a common scenario in many data processing tasks. In the low-SNR regime, the situation is radically different. Specifically, it was shown that, when $\sigma \to \infty$, the sample complexity of the MRA problem is determined by the first moment $\bar{q}$ that distinguishes between different signals. More precisely, in the asymptotic regime where N and σ diverge, the estimation error of any method is bounded away from zero if $N \cdot \mathrm{SNR}^{\bar{q}}$ is bounded from above. We then say that the sample complexity in the low-SNR regime is proportional to $1/\mathrm{SNR}^{\bar{q}}$ [4].

For the cryo-EM setup, assuming perfect particle picking and no CTF, it was shown that the sample complexity depends on the distribution of rotations: for uniform distribution the sample complexity scales as $1/\mathrm{SNR}^3$ (i.e., the structure is determined by the third moment), whereas for generic nonuniform distribution, the second moment suffices, and, thus, the sample complexity scales as $1/\mathrm{SNR}^2$ [9], [66]. Similar results were derived for simpler setups; for instance, when a discrete signal is acted on by cyclic shifts (as shown in Figure 7) [3], [9], [19].

The pivotal role of moments in sample complexity analysis led naturally to the design of algorithms based on the method of moments. Some of these algorithms hinge on the tightly related notion of invariant polynomials. A polynomial p is called *invariant* under the action of a group G if, for any signal x in the specified space and any $g \in G$, it satisfies $p(g \circ x) = p(x)$. One particularly important invariant is the third-order polynomial, the bispectrum—originally proposed by John Tukey [77]—that was used for 2D classification [87] and *ab initio* modeling [42]. We refer readers to [71] and the references therein for a more detailed account of the MRA problem.

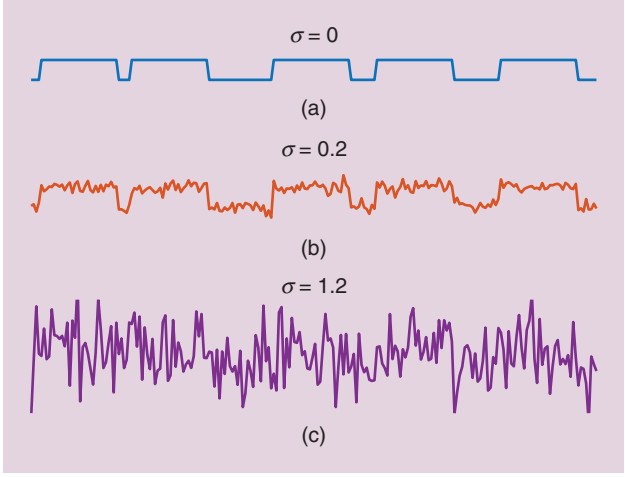

FIGURE 8. An example of an MTD observation with five signal occurrences at different noise levels (20): (a) $\sigma = 0$, (b) $\sigma = 0.2$, and (c) $\sigma = 1.2$. When (a) the noise level is low, it is easy to detect the signal occurrences and estimate the underlying signal by averaging out the noise. However, (c) when the noise level is high, reliable detection is rendered challenging. The task is, then, to estimate the underlying signal directly, without intermediate detection of its occurrences.

MTD

MTD is the problem of estimating a signal that occurs multiple times at unknown locations in a noisy measurement. In its simplest form, the MTD problem is an instance of the 1D blind deconvolution problem and can be written as

$$y = x * s + \varepsilon, \tag{20}$$

where $x \in \mathbb{R}^L$ is the target signal, $s \in \{0,1\}^{N-L+1}$ is a binary signal whose ones indicate the location of the signal copies in the measurement $y \in \mathbb{R}^N$, and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$; see Figure 8. Detecting the signal occurrences in the data (i.e., estimating the signal s) is the analog of particle picking in cryo-EM. Clearly, if the noise level is low, one can estimate s (and, analogously, detect the particle projections in the micrograph), extract the signal occurrences, and average them to suppress the noise.

However, as mentioned previously, low SNR precludes detection (and particle picking), and, therefore, one must estimate x directly, without explicit estimation of s. In [18], it was shown that under certain generative models of s, the signal x can be estimated provably, at any noise level, from the bispectrum (third-order statistics). More ambitiously, numerical experiments suggest that the bispectrum suffices to estimate multiple signals $x_1, \ldots, x_K$ from a mix of blind deconvolution problems

$$y = \sum_{i=1}^{K} x_i * s_i + \varepsilon, \tag{21}$$

without explicit estimation of the binary signals $s_1, \ldots, s_K$ that indicate the location of the corresponding signals.

The MTD model can be extended to formulate a generative model of a micrograph: the key is to treat x as a random signal that represents the tomographic projections of the molecule ϕ, rather than a deterministic signal as in (20) and (21). Specifically, locations are chosen in the 2D plane: these are the positions of the embedded samples in the micrograph. For each location, a signal is drawn from a probability distribution described by the model

$$x = h * PR_\omega \phi, \tag{22}$$

where the 3D rotation R_ω is applied to the volume ϕ according to a (possibly unknown) distribution of ω over $SO(3)$, P is a tomographic projection, and h is the microscope's PSF; the goal is to estimate ϕ. Therefore, the MTD model paves the way toward fully modeling the cryo-EM problem, including most of its important features [18]. In particular, MTD provides a mathematical and computational framework for reconstructing ϕ directly from micrographs, without intermediate particle picking [17]. A full analysis of this model is still lacking.

Remaining computational and theoretical challenges

Some interesting and important computational and theoretical challenges that lie ahead for single-particle reconstruction using cryo-EM are given next.

Conformational heterogeneity: Modeling and recovery

One of the important opportunities offered by cryo-EM is its ability to analyze different functional and conformational

states of macromolecules. Mathematically, it entails estimating multiple 3D structures simultaneously; we refer to this as the *heterogeneity problem*. There is no consensus about the proper way to model the heterogeneity problem, and the computational tools are not well established. Therefore, we believe that it offers an opportunity for researchers with strong mathematical and computational backgrounds to make a profound impact on the field of structural biology.

We now briefly survey different approaches to handling the heterogeneity problem. To this end, we extend the basic cryo-EM model (2) to account for conformational variability:

$$y_i = h_i * T_{t_i} P R_{\omega_i} \phi_i + \text{“noise,”} \quad i = 1, \ldots, N, \tag{23}$$

where the goal is to estimate the distribution from which the structures $\phi_i, \ldots, \phi_N$ are sampled. With this formulation, the problem is ill posed: there is not enough information in N 2D images to recover N 3D structures. Hence, additional assumptions about the structures must be made. The different techniques in the field can be broadly classified into two categories: discrete and continuous models. The discrete heterogeneous model assumes that the measurements stem from a few independent volumes (i.e., a discrete distribution). Under this model, each projection image can be written as

$$y_i = h_i * T_{t_i} P R_{\omega_i} \phi_{k_i} + \text{“noise,”} \quad i = 1, \ldots, N, \tag{24}$$

where ϕ_k, $k = 1, \ldots, K$, represent the K different volumes. The advantage of this approach is evident: it is easy to extend the ML-EM framework to incorporate several volumes simultaneously. Alternatively, some software packages apply a preliminary 3D classification stage, in which the 2D experimental images are being classified into different structures. CryoSPARC runs multiple trials of an SGD algorithm from different initializations. This procedure occasionally leads to multiple low-resolution estimates, and each of those is refined by an ML-EM algorithm [57]. Although the simplicity of the discrete model is a big advantage, its drawbacks are apparent: it does not scale well for large K and ignores the correlation between different functional states of the molecule and, thus, overlooks important information.

The second approach, referred to as *continuous heterogeneity*, assumes that $\phi_1, \ldots, \phi_N$ can be embedded in a low-dimensional space. For instance, one approach is to assume that the set of conformations lies in a linear subspace (that can be learned using PCA) or in more intricate low-dimensional manifolds (that could be learned by other spectral methods, e.g., diffusion maps). An alternative suggestion was to model the structure as a set of rigid domains that can move with respect to each other. We refer readers to a thorough survey on the subject in [74].

Verification

Given a 3D reconstruction, how do we verify that it is a reliable and faithful representation of the underlying molecule? This is a question of crucial importance for any scientific field. Several validation techniques were proposed in the cryo-EM literature. For example, it is possible to check the consistency of the 3D map by recording pairs of images of the same particles at different tilt angles and comparing the relative angle between orientations assigned to each projection: ideally, it should agree with the relative rotation angle of the microscope's specimen holder used during the experiment.

This approach is called the *tilt-pair validation technique* and is useful only for intermediate-resolution structures [40], [61]. In practice, structure validation is based on a set of heuristics and the experts' knowledge and experience. For instance, it is common to initialize 3D reconstruction algorithms from multiple random points; if all instances attain similar structures, it serves as a validation fidelity. If the same molecule were reconstituted by other technologies, such as X-ray crystallography and NMR, then structures can be compared. In addition, experimental data are usually uploaded to public repositories, and, thus, other researchers can process the same data, experiment with different computational techniques, and compare the results [61].

Verification is also related to the question of determining the resolution of a recovered structure. The current convention is to reconstruct two structures independently, each from one half of the data. The two subsets are chosen at random. The highest frequency for which the two structures agree (up to some tolerance) determines the resolution [75]. This process can be understood as an indication for the confidence we have in the structure at a given resolution.

However, this method is susceptible to systematic flaws; if the same refinement procedure is applied to both halves of the data, it can induce correlated blunders (although the data are independent), and the resolution determination would be unsound. In [24], a simple computational procedure was proposed to validate a structure by assessing the amount of overfitting that is present in the 3D map. Establishing computational tools that provide confidence intervals to estimated structures and are immunized against systematic errors is one of the remaining challenges in the field.

Theoretical foundations of cryo-EM

Many mathematical and statistical properties of the cryo-EM problem are still unexplained (and even unexplored). A prominent example is the sample complexity of the full cryo-EM problem: given a fixed SNR level (that might be very low) and a fixed setup, how many particles are required to achieve a desired resolution? An initial analysis was conducted in [9] and [66]; see the “MRA” section.

Another interesting question concerns computational complexity and information-computational gaps [11]. The fact that there is enough information to solve a problem (sample complexity) does not immediately imply that there is an efficient (polynomial time) algorithm to solve it. For example, it might be that for nonuniform distribution of rotations, the second moment would suffice to estimate a 3D structure from an information-theoretic perspective [66], but, at the same time, it could be computationally hard in that regime. In that case, we might need to consider the third-order moment to design a computationally efficient algorithm. Information-computational gaps

have been observed empirically in variants of the MRA and MTD models [18], [22]. Related questions—that might be even more challenging—deal with the properties of specific algorithms, such as ML-EM, the method of moments, and SGD, that are not well understood.

Another interesting research thread regards the size limit of molecular structures that can be elucidated by cryo-EM. The common belief in the community is that very small molecules cannot be visualized using cryo-EM. The logic is simple: small molecules induce low contrast, and, thus, low SNR on the micrograph, which, in turn, hinders detection (particle picking) [38]. A recent article contends this belief and suggests that it is possible, at least in principle, to reconstruct structures directly from the micrograph, without particle picking and at any SNR level, given enough data [17].

Machine learning

The groundbreaking advances in machine learning in the last decade have dramatically reshaped many computational fields and penetrated into some scientific applications. Naturally, learning techniques based on deep neural networks have been applied to cryo-EM as well. Examples are particle picking [83], [84], [91] (as discussed in the "Particle Picking" section), validation [7], 3D reconstruction [89], and particle pruning [62]. In addition, manifold-learning techniques were designed for 3D heterogeneity analysis [31], [51] and denoising [45]. Modern-learning techniques were also implemented in other cryo-EM applications that do not involve single-particle reconstruction; see, for instance, an application to feature extraction in cellular electron cryotomography [23].

Deep learning gained its popularity in applications with low noise levels. The performance of these techniques in more challenging environments, such as with highly contaminated data, is not clear yet. In addition, supervised-learning techniques are susceptible to model bias—the reconstruction will depend heavily on the training data rather than on the experimental images. This explains why the impact of deep learning on the cryo-EM field—especially on more involved tasks, such as 3D reconstruction and the heterogeneity problem—is limited at the moment. We expect that more efforts in this direction will be made in the coming years. In particular, it is still to be clarified whether this set of computational tools can outperform current tools in the field, which are based on more classical statistics.

Perspective

Single-particle reconstruction using cryo-EM is an alluring research area for investigators interested in developing modern computational tools and sophisticated mathematical models for an emerging scientific field. In this article, we introduced the problem of constituting 3D molecular structures using cryo-EM and described its unique computational characteristics and challenges. We delineated relations between the cryo-EM inverse problem and a variety of disciplines at the core of signal processing, information theory, statistics, machine learning, and group theory. We believe that contributions from these areas have the potential to drive the field forward. New ideas and solutions can, and should, be tested on experimental cryo-EM data sets publicly available online [1], [2]. We also reviewed two abstract frameworks to conveniently study the cryo-EM inverse problem from computational, statistical, and theoretical perspectives.

We believe that the challenges arising in cryo-EM research provide ample opportunities to investigate and test novel algorithms and advanced mathematical techniques to impact a task of paramount importance: to broaden our understanding of the fundamental mechanisms of life.

Acknowledgments

We thank the anonymous reviewers and Jong Chul Ye for their valuable comments and suggestions. Amit Singer was supported in part by award R01GM090200 from the National Institute of General Medical Sciences, FA9550-17-1-0291 from the Air Force Office of Scientific Research, Simons Foundation Math+X Investigator Award, the Moore Foundation Data-Driven Discovery Investigator Award, and National Science Foundation BIGDATA Award IIS-1837992.

Authors

Tamir Bendory (bendory@tauex.tau.ac.il) received his B.Sc. degree in biomedical engineering in 2010 and his M.Sc. and Ph.D. degrees in electrical engineering in 2012 and 2015, respectively, all from the Technion-Israel Institute of Technology. He then joined the program in applied and computational mathematics at Princeton University, New Jersey, as a postdoctoral research associate. In 2019, he joined Tel Aviv University as a senior lecturer (assistant professor) of electrical engineering. His research interests include theoretical and computational aspects of signal processing, data science, statistics, and optimization.

Alberto Bartesaghi (alberto@cs.duke.edu) received his B.Sc. and M.Sc. degrees in electrical engineering from the Universidad de la Republica, Montevideo, Uruguay, in 1999 and 2001 respectively, and his Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2005. He then joined the Biophysics Section of the Laboratory of Cell Biology at the National Cancer Institute/National Institutes of Health, Bethesda, Maryland, to conduct his postdoctoral studies and later became an associate scientist with the Center for Cancer Research. In 2018, he joined Duke University as an associate professor of computer science, biochemistry, and electrical and computer engineering.

Amit Singer (amits@math.princeton.edu) received his B.Sc. degree in physics and mathematics and his Ph.D. degree in applied mathematics from Tel Aviv University, Israel, in 1997 and 2005, respectively. He is a professor of mathematics and core member of the Program in Applied and Computational Mathematics and of the Center for Statistics and Machine Learning at Princeton University, New Jersey. He joined Princeton in 2008 as an assistant professor. From 2005 to 2008, he was a Gibbs Assistant Professor in Applied Mathematics at Yale University. His current research interests in applied mathematics

include theoretical and computational aspects of data science and developing computational methods for structural biology.

References

[1] "EMPIAR: Electron Microscopy Public Image Archive," European Bioinformatics Institute, Cambridgeshire, UK. [Online]. Available: https://www.ebi.ac.uk/pdbe/emdb/empiar/

[2] "The Electron Microscopy Data Bank (EMDB) at PDBe," European Bioinformatics Institute, Cambridgeshire, UK. [Online]. Available: http://www.ebi.ac.uk/pdbe/emdb/

[3] E. Abbe, T. Bendory, W. Leeb, J. M. Pereira, N. Sharon, and A. Singer, "Multireference alignment is easier with an aperiodic translation distribution," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3565–3584, 2018. doi: 10.1109/TIT.2018.2889674.

[4] E. Abbe, J. M. Pereira, and A. Singer, "Estimation in the group action channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2018, pp. 561–565. doi: 10.1109/ISIT.2018.8437646.

[5] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, pp. 105–112, 2011. doi: 10.1145/2001269.2001293.

[6] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, "Fundamental limits in multi-image alignment," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5707–5722, 2016. doi: 10.1109/TSP.2016.2600517.

[7] T. K. Avramov, D. Vyenielo, J. Gomez-Blanco, S. Adinarayanan, J. Vargas, and D. Si, "Deep learning for validating and estimating resolution of cryo-electron microscopy density maps," *Molecules*, vol. 24, no. 6, p. 1181, 2019. doi: 10.3390/molecules24061181.

[8] X.-C. Bai, G. McMullan, and S. H. Scheres, "How cryo-EM is revolutionizing structural biology," *Trends Biochem. Sci.*, vol. 40, no. 1, pp. 49–57, 2015. doi: 10.1016/j.tibs.2014.10.005.

[9] A. S. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, and A. S. Wein, Estimation under group actions: Recovering orbits from invariants. 2017. [Online]. Available: arXiv:1712.10163

[10] A. S. Bandeira, Y. Chen, and A. Singer, Non-unique games over compact groups and orientation estimation in cryo-EM. 2015. [Online]. Available: arXiv:1505.03840

[11] A. S. Bandeira, A. Perry, and A. S. Wein, "Notes on computational-to-statistical gaps: Predictions using statistical physics," *Port. Math.*, vol. 75, no. 2, pp. 159–186, 2018. doi: 10.4171/PM/2014.

[12] S. Banerjee, A. Bartesaghi, A. Merk, P. Rao, S. L. Bulfer, Y. Yan, N. Green, B. Mroczkowski et al., "2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition," *Science*, vol. 351, no. 6275, pp. 871–875, 2016. doi: 10.1126/science.aad7974.

[13] A. Barnett, L. Greengard, A. Pataki, and M. Spivak, "Rapid solution of the cryo-EM reconstruction problem by frequency marching," *SIAM J. Imaging Sci.*, vol. 10, no. 3, pp. 1170–1195, 2017. doi: 10.1137/16M1097171.

[14] A. H. Barnett, J. Magland, and L. af Klinteberg, "A parallel nonuniform fast Fourier transform library based on an 'exponential of semicircle' kernel," *SIAM J. Sci. Comput.*, vol. 41, no. 5, pp. C479–C504, 2019. doi: 10.1137/18M120885X.

[15] A. Bartesaghi, C. Aguerrebere, V. Falconieri, S. Banerjee, L. A. Earl, X. Zhu, N. Grigorieff, J. L. Milne et al., "Atomic resolution cryo-EM structure of β-galactosidase," *Structure*, vol. 26, no. 6, pp. 848–856, 2018. doi: 10.1016/j.str.2018.04.004.

[16] A. Bartesaghi, A. Merk, S. Banerjee, D. Matthies, X. Wu, J. L. Milne, and S. Subramaniam, "2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor," *Science*, vol. 348, no. 6239, pp. 1147–1151, 2015. doi: 10.1126/science.aab1576.

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

[18] 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, 2019. doi: 10.1088/1361-6420/ab2aec. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6420/ab2aec/meta

[19] T. Bendory, N. Boumal, C. Ma, Z. Zhao, and A. Singer, "Bispectrum inversion with application to multireference alignment," *IEEE Trans. Signal Process.*, vol. 66, no. 4, pp. 1037–1050, 2018. doi: 10.1109/TSP.2017.2775591.

[20] T. Bhamre, T. Zhang, and A. Singer, "Denoising and covariance estimation of single particle cryo-EM images," *J. Struct. Biol.*, vol. 195, no. 1, pp. 72–81, 2016. doi: 10.1016/j.jsb.2016.04.013.

[21] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *Siam Rev.*, vol. 60, no. 2, pp. 223–311, 2018. doi: 10.1137/16M1080173.

[22] N. Boumal, T. Bendory, R. R. Lederman, and A. Singer, "Heterogeneous multireference alignment: A single pass approach," in *Proc. 52nd IEEE Annu. Conf. Information Sciences and Systems (CISS)*, 2018, pp. 1–6. doi: 10.1109/CISS.2018.8362313.

[23] M. Chen, W. Dai, S. Y. Sun, D. Jonasch, C. Y. He, M. F. Schmid, W. Chiu, and S. J. Ludtke, "Convolutional neural networks for automated annotation of cellular cryo-electron tomograms," *Nat. Meth.*, vol. 14, no. 10, pp. 983–985, 2017. doi: 10.1038/nmeth.4405.

[24] S. Chen, G. McMullan, A. R. Faruqi, G. N. Murshudov, J. M. Short, S. H. Scheres, and R. Henderson, "High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy," *Ultramicroscopy*, vol. 135, pp. 24–35, Dec. 2013. doi: 10.1016/j.ultramic.2013.06.004.

[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B, Methodol.*, vol. 39, no. 1, pp. 1–22, 1977. doi: 10.1111/j.2517-6161.1977.tb01600.x.

[26] D. L. Donoho, M. Gavish, and I. M. Johnstone, "Optimal shrinkage of eigenvalues in the spiked covariance model," *Ann. Stat.*, vol. 46, no. 4, pp. 1742, 2018. doi: 10.1214/17-AOS1601.

[27] J. Dubochet, J. Lepault, R. Freeman, J. Berriman, and J.-C. Homo, "Electron microscopy of frozen water and aqueous solutions," *J. Microsc.*, vol. 128, no. 3, pp. 219–237, 1982. doi: 10.1111/j.1365-2818.1982.tb04625.x.

[28] P. Elbau, M. Ritsch-Marte, O. Scherzer, and D. Schmutz, Inverse problems of trapped objects. 2019. [Online]. Available: arXiv:1907.01387

[29] J. Frank and P. Penczek, "On the correction of the contrast transfer function in biological electron microscopy," *Optik*, vol. 98, no. 3, pp. 125–129, 1995.

[30] J. Frank, "Advances in the field of single-particle cryo-electron microscopy over the last decade," *Nat. Protoc.*, vol. 12, no. 2, pp. 209–212, 2017. doi: 10.1038/nprot.2017.004.

[31] J. Frank and A. Ourmazd, "Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM," *Methods*, vol. 100, pp. 61–67, May 2016. doi: 10.1016/j.ymeth.2016.02.007.

[32] T. Grant and N. Grigorieff, "Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6," *Elife*, vol. 4, p. e06980, May 2015. doi: 10.7554/eLife.06980.

[33] T. Grant, A. Rohou, and N. Grigorieff, "cisTEM, user-friendly software for single-particle image processing," *Elife*, vol. 7, p. e35383, 2018. doi: 10.7554/eLife.35383.

[34] 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. doi: 10.1016/j.jsb.2017.09.007.

[35] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast Fourier transform," *SIAM Rev.*, vol. 46, no. 3, pp. 443–454, 2004. doi: 10.1137/S003614450343200X.

[36] T. W. Guo, A. Bartesaghi, H. Yang, V. Falconieri, P. Rao, A. Merk, E. T. Eng, A. M. Raczkowski et al., "Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR-Cas surveillance complex," *Cell*, vol. 171, no. 2, pp. 414–426, 2017. doi: 10.1016/j.cell.2017.09.006.

[37] A. Heimowitz, J. Andén, and A. Singer, "APPLE picker: Automatic particle picking, a low-effort cryo-EM framework," *J. Struct. Biol.*, vol. 204, no. 2, pp. 215–227, 2018. doi: 10.1016/j.jsb.2018.08.012.

[38] R. Henderson, "The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules," *Q. Rev. Biophys.*, vol. 28, no. 2, pp. 171–193, 1995. doi: 10.1017/S003358350000305X.

[39] R. Henderson, "Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 45, pp. 18,037–18,041, 2013. doi: 10.1073/pnas.1314449110.

[40] R. Henderson, S. Chen, J. Z. Chen, N. Grigorieff, L. A. Passmore, L. Ciccarelli, J. L. Rubinstein, R. A. Crowther et al., "Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy," *J. Mol. Biol.*, vol. 413, no. 5, pp. 1028–1046, 2011. doi: 10.1016/j.jmb.2011.09.008.

[41] I. Jolliffe, *Principal Component Analysis*. Berlin: Springer-Verlag, 2011.

[42] Z. Kam, "The reconstruction of structure from electron micrographs of randomly oriented particles," *J. Theor. Biol.*, vol. 82, no. 1, pp. 15–39, 1980. doi: 10.1016/0022-5193(80)90088-0.

[43] Y. Kang, O. Kuybeda, P. W. de Waal, S. Mukherjee, N. Van Eps, P. Dutka, X. E. Zhou, A. Bartesaghi et al., "Cryo-EM structure of human rhodopsin bound to an inhibitory G protein," *Nature*, vol. 558, no. 7711, pp. 553–558, 2018. doi: 10.1038/s41586-018-0215-y.

[44] W. Kühlbrandt, "The resolution revolution," *Science*, vol. 343, no. 6178, pp. 1443–1444, 2014. doi: 10.1126/science.1251652.

[45] B. Landa and Y. Shkolnisky, "The steerable graph Laplacian and its application to filtering image datasets," *SIAM J. Imaging Sci.*, vol. 11, no. 4, pp. 2254–2304, 2018. doi: 10.1137/18M1169394.

[46] A. Levis, Y. Y. Schechner, and R. Talmon, "Statistical tomography of microscopic life," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2018, pp. 6411–6420. doi: 10.1109/CVPR.2018.00671.

[47] M. Liao, E. Cao, D. Julius, and Y. Cheng, "Structure of the TRPV1 ion channel determined by electron cryo-microscopy," *Nature*, vol. 504, no. 7478, pp. 107, 2013. doi: 10.1038/nature12822.

[48] H. Liu, B. K. Poon, D. K. Saldin, J. C. Spence, and P. H. Zwart, "Three-dimensional single-particle imaging using angular correlations from X-ray laser data," *Acta Crystallogr. A, Found. Crystallogr.*, vol. 69, no. 4, pp. 365–373, 2013. doi: 10.1107/S0108767313006016.

[49] D. Matthies, C. Bae, G. E. Toombes, T. Fox, A. Bartesaghi, S. Subramaniam, and K. J. Swartz, "Single-particle cryo-EM structure of a voltage-activated potassium channel in lipid nanodiscs," *Elife*, vol. 7, p. e37558, 2018. doi: 10.7554/eLife.37558.

[50] A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer et al., "Breaking cryo-EM resolution barriers to facilitate drug discovery," *Cell*, vol. 165, no. 7, pp. 1698–1707, 2016. doi: 10.1016/j.cell.2016.05.040.

[51] A. Moscovich, A. Halevi, J. Andén, and A. Singer, "Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes," *Inverse Probl.*, to be published. doi: 10.1088/1361-6420/ab4f55. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6420/ab4f55/meta

[52] E. Nogales and S. H. Scheres, "Cryo-EM: A unique tool for the visualization of macromolecular complexity," *Mol. Cell*, vol. 58, no. 4, pp. 677–689, 2015. doi: 10.1016/j.molcel.2015.02.019.

[53] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Denoising linear models with permuted data," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, 2017, pp. 446–450. doi: 10.1109/ISIT.2017.8006567.

[54] K. Pearson, "On lines and planes of closest fit to systems of points in space," *London, Edinburgh, Dublin Philosoph. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, 1901. doi: 10.1080/14786440109462720.

[55] P. A. Penczek, R. Renka, and H. Schomberg, "Gridding-based direct Fourier inversion of the three-dimensional ray transform," *J. Opt. Soc. Am. A*, vol. 21, no. 4, pp. 499–509, 2004. doi: 10.1364/JOSAA.21.000499. [Online]. Available: https://www.osapublishing.org/josaa/abstract.cfm?uri=JOSAA-21-4-499

[56] G. Pragier and Y. Shkolnisky, "A common lines approach for ab initio modeling of cyclically symmetric molecules," *Inverse Problems*, Nov. 19, 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6420/ab2fb2

[57] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination," *Nat. Meth.*, vol. 14, no. 3, pp. 290–296, 2017. doi: 10.1038/nmeth.4169.

[58] Z. Ripstein and J. Rubinstein, "Processing of cryo-EM movie data," *Meth. Enzymol.*, vol. 579, pp. 103–124, 2016. doi: 10.1016/bs.mie.2016.04.009.

[59] 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. doi: 10.1016/j.jsb.2015.08.008.

[60] D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard, "SE-sync: A certifiably correct algorithm for synchronization over the special Euclidean group," *Int. J. Robot. Res.*, vol. 38, nos. 2–3, pp. 95–125, 2019. doi: 10.1177/0278364918784361.

[61] P. B. Rosenthal and J. L. Rubinstein, "Validating maps from single particle electron cryomicroscopy," *Curr. Opin. Struct. Biol.*, vol. 34, pp. 135–144, 2015. doi: 10.1016/j.sbi.2015.07.002.

[62] R. Sanchez-Garcia, J. Segura, D. Maluenda, J. M. Carazo, and C. O. S. Sorzano, "Deep consensus, a deep learning-based approach for particle pruning in cryo-electron microscopy," *IUCrJ*, vol. 5, no. 6, pp. 854–865, 2018. doi: 10.1107/S2052252518014392.

[63] 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. doi: 10.1016/j.jsb.2012.09.006.

[64] S. H. Scheres, "Semi-automated selection of cryo-EM particles in RELION-1.3," *J. Struct. Biol.*, vol. 189, no. 2, pp. 114–122, 2015. doi: 10.1016/j.jsb.2014.11.010.

[65] S. H. Scheres, M. Valle, R. Nuñez, C. O. Sorzano, R. Marabini, G. T. Herman, and J.-M. Carazo, "Maximum-likelihood multi-reference refinement for electron microscopy images," *J. Mol. Biol.*, vol. 348, no. 1, pp. 139–149, 2005. doi: 10.1016/j.jmb.2005.02.031.

[66] 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 Probl.,* vol. 34, no. 7, 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6420/ab6139/meta

[67] M. Shatsky, R. J. Hall, S. E. Brenner, and R. M. Glaeser, "A method for the alignment of heterogeneous macromolecules from electron microscopy," *J. Struct. Biol.*, vol. 166, no. 1, pp. 67–78, 2009. doi: 10.1016/j.jsb.2008.12.008.

[68] F. Sigworth, "A maximum-likelihood approach to single-particle image refinement," *J. Struct. Biol.*, vol. 122, no. 3, pp. 328–339, 1998. doi: 10.1006/jsbi.1998.4014.

[69] C. V. Sindelar and N. Grigorieff, "An adaptation of the Wiener filter suitable for analyzing images of isolated single particles," *J. Struct. Biol.*, vol. 176, no. 1, pp. 60–74, 2011. doi: 10.1016/j.jsb.2011.06.010.

[70] A. Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 1, pp. 20–36, 2011. doi: 10.1016/j.acha.2010.02.001.

[71] A. Singer, "Mathematics for cryo-electron microscopy," in *Proc. Int. Congr. Mathematicians*, 2018, pp. 3995–4014.

[72] A. Singer and Y. Skhkolnisky, "Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming," *SIAM J. Imaging Sci.*, vol. 4, no. 2, pp. 543–572, 2011. doi: 10.1137/090767777.

[73] C. Sorzano, J. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, R. Melero, G. Caffarena-Fernández, M. Li, G. Xu et al., "A clustering approach to multireference alignment of single-particle projections in electron microscopy," *J. Struct. Biol.*, vol. 171, no. 2, pp. 197–206, 2010. doi: 10.1016/j.jsb.2010.03.011.

[74] C. Sorzano, A. Jiménez, J. Mota, J. Vilas, D. Maluenda, M. Martínez, E. Ramírez-Aportela, T. Majtner et al., "Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy," *Acta Crystallogr. F, Struct. Biol. Commun.*, vol. 75, no. 1, pp. 19–32, 2019. doi: 10.1107/S2053230X18015108.

[75] C. Sorzano, J. Vargas, J. Otón, V. Abrishami, J. de la Rosa-Trevín, J. Gómez-Blanco, J. Vilas, R. Marabini et al., "A review of resolution measures and related aspects in 3D electron microscopy," *Prog. Biophys. Mol. Biol.*, vol. 124, pp. 1–30, Mar. 2017. doi: 10.1016/j.pbiomolbio.2016.09.005.

[76] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "EMAN2: An extensible image processing suite for electron microscopy," *J. Struct. Biol.*, vol. 157, no. 1, pp. 38–46, 2007. doi: 10.1016/j.jsb.2006.05.009.

[77] J. Tukey, "The spectral representation and transformation properties of the higher moments of stationary time series," in D. R. Brillinger, *The Collected Works of John W. Tukey*, vol. 1. Belmont, CA: Wadsworth, 1953, pp. 165–184.

[78] B. Vainshtein and A. Goncharov, "Determination of the spatial orientation of arbitrarily arranged identical particles of unknown structure from their projections," *Sov. Phys. Dokl.*, vol. 31, p. 278, Apr. 1986.

[79] M. Van Heel, "Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction," *Ultramicroscopy*, vol. 21, no. 2, pp. 111–123, 1987. doi: 10.1016/0304-3991(87)90078-7.

[80] M. Van Heel and J. Frank, "Use of multivariates statistics in analysing the images of biological macromolecules," *Ultramicroscopy*, vol. 6, no. 2, pp. 187–194, 1981. doi: 10.1016/0304-3991(81)90059-0.

[81] K. R. Vinothkumar and R. Henderson, "Single particle electron cryomicroscopy: Trends, issues and future perspective," *Q. Rev. Biophys.*, vol. 49, p. e13, Jan. 2016. doi: 10.1017/S0033583516000068.

[82] B. von Ardenne, M. Mechelke, and H. Grubmüller, "Structure determination from single molecule X-ray scattering with three photons per image," *Nat. Commun.*, vol. 9, p. 2375, June 2018. doi: 10.1038/s41467-018-04830-4.

[83] T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel et al., "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM," *Commun. Biol.*, vol. 2, p. 218, June 2019. doi: 10.1038/s42003-019-0437-z.

[84] 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," *J. Struct. Biol.*, vol. 195, no. 3, pp. 325–336, 2016. doi: 10.1016/j.jsb.2016.07.006.

[85] K. Zhang, "Gctf: Real-time CTF determination and correction," *J. Struct. Biol.*, vol. 193, no. 1, pp. 1–12, 2016. doi: 10.1016/j.jsb.2015.11.003.

[86] Z. Zhao, Y. Shkolnisky, and A. Singer, "Fast steerable principal component analysis," *IEEE Trans. Comput. Imaging*, vol. 2, no. 1, pp. 1–12, 2016. doi: 10.1109/TCI.2016.2514700.

[87] Z. Zhao and A. Singer, "Rotationally invariant image representation for viewing direction classification in cryo-EM," *J. Struct. Biol.*, vol. 186, no. 1, pp. 153–166, 2014. doi: 10.1016/j.jsb.2014.03.003.

[88] S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard, "MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy," *Nat. Meth.*, vol. 14, no. 4, pp. 331–332, 2017. doi: 10.1038/nmeth.4193.

[89] E. D. Zhong, T. Bepler, J. H. Davis, and B. Berger, Reconstructing continuously heterogeneous structures from single particle cryo-EM with deep generative models. 2019. [Online]. Available: arXiv:1909.05215

[90] Y. Zhou, A. Moscovich, T. Bendory, and A. Bartesaghi, "Unsupervised particle sorting for high-resolution single-particle cryo-EM," *Inverse Probl.,* to be published. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6420/ab5ec8/meta

[91] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy," *BMC Bioinformatics*, vol. 18, no. 1, p. 348, 2017. doi: 10.1186/s12859-017-1757-y.

[92] J. Zivanov, T. Nakane, and S. H. Scheres, "A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis," *IUCrJ*, vol. 6, no. 1, pp. 5–17, 2019. doi: 10.1107/S205225251801463X.

[93] "Press release: The Nobel Prize in Chemistry 2017," Oct. 4, 2017. [Online]. Available: https://www.nobelprize.org/prizes/chemistry/2017/press-release/

SP