MODEL CALCULATIONS FOR JOINT PATTERN RECOGNITION AND SIGNAL RECONSTRUCTION IN CRYO ELECTRON MICROSCOPY*

ZHYE YIN†, PETER C. DOERSCHUK†‡, AND SAUL B. GELFAND†

Dedicated to Sanjoy Mitter on the occasion of his 70th birthday.

Abstract. Cryo electron microscopy is a measurement modality which provides images from which 3-D reconstructions of biological particles such as viruses can be estimated. When the specimen is composed of mixtures of particles of different types, the 3-D reconstruction problem must be solved jointly with a pattern classification problem. The performance of the estimators is not well understood because the computations are not suitable for analytical results and are too large for extensive Monte Carlo results. The problem formulation typically has nuisance parameters and different treatments of the nuisance parameters lead to different estimators. In this paper two types of estimators and two model problems are studied with the conclusion that it is difficult to improve upon maximum likelihood estimators based on integrating out the nuisance parameters.

1. Introduction. Cryo electron microscopy (cryo EM) is a biophysical method for determining the 3-D structure of a particle, for example, a virus. There are many different types of cryo EM reconstruction problems and we focus on problems where an ensemble of identical particles is available, the particles do not associate to form crystalline structures, and each particle has internal symmetries (such as the icosahedral symmetry that is present in a broad class of viral particles). The measurements are images of single particles. Each image is basically a linearly-filtered 2-D projection of the 3-D electron scattering intensity of the particle where the linear filter is due to the electron optics of the microscope and is called the contrast transfer function (CTF). There are several major challenges in computing a 3-D reconstruction from a set of cryo EM images. The particles are highly sensitive to the electron beam. Therefore only one projection image is recorded from any particular particle (hence the need for an ensemble of identical particles) and, because the beam current is minimized, the image is highly noisy, with an SNR less than 1. Because the SNR is low, it is difficult to determine the center of the particle in the image. Because the particles

†Purdue University, School of Electrical and Computer Engineering, 465 Northwestern Avenue, West Lafayette, IN 47907-2035, E-mail: [zhye,doerschu,gelfand]@ecn.purdue.edu

‡Purdue University, Department of Biomedical Engineering.

are small, a typical virus measures 10^2–10^3Å in diameter, the particle is placed in the microscope in a random orientation and an image is then recorded. Therefore the orientation of the projection is not known. Under standard physics assumptions the functional form of the CTF is known [1] and typically the parameter values are at least approximately known. However, in the 2-D spatial frequency domain, the CTF has zeros in the range of spatial frequencies necessary for computing a 3-D reconstruction. In summary, the challenges include unknown projection orientation, unknown location of the center of the particle, zeros in the CTF, and low SNR.

The origin of the quantitative image formation model on which the 3-D reconstructions are based is as follows: Let the real-valued positive scattering density of the particle be denoted by $\rho(\mathbf{x})$ with 3-D Fourier transform denoted by $P(\mathbf{k})$. We consider only the first order image formation theory [2, 3, 4]. Because the sample is thin, the EM image depends only on the projection of the 3-D scattering density onto the 2-D object plane. This projection is denoted by $\sigma(\boldsymbol{\chi})$ with Fourier transform $\Sigma(\boldsymbol{\kappa})$. Let the image be denoted by $\sigma^i(\boldsymbol{\chi})$ with Fourier transform $\Sigma^i(\boldsymbol{\kappa})$. The key result [2, Eq. 11c] is that Σ and Σ^i are closely related. Let the contrast transfer function (CTF) be denoted by $G(\kappa)$, noting that it only depends on the magnitude of the 2-D spatial frequency vector $\boldsymbol{\kappa}$ [1]. Then, in terms of these definitions, the Fourier transforms of the measured image and the projection are related by $\Sigma^i(\boldsymbol{\kappa}) = \exp(-i\boldsymbol{\kappa}^T\boldsymbol{\chi}_0)G(\kappa)\Sigma(\boldsymbol{\kappa})$ where $\boldsymbol{\chi}_0$ is the origin location in the image. This theory can be elaborated to include the effects of specimen thickness (leading to varying levels of defocus), chromatic aberration, partial coherence, *etc.* In order to relate the projection σ with 2-D Fourier transform Σ to the 3-D Fourier transform P of the scattering density ρ we use the 3-D version of the projection slice theorem [5, Eq. 6.2.5, p. 130]. For projections in the z direction, it is easy to see from the definition of the 3-D Fourier transform in rectangular coordinates that $\Sigma(\boldsymbol{\kappa}) = P((\boldsymbol{\kappa}^T, 0)^T)$. In the case of general projection orientation there is a set of Euler angles, denoted by α, β, and γ, such that the desired projection is the z direction projection of the rotated scattering density $\rho'(\mathbf{x}) = \rho(R_E^{-1}(\alpha, \beta, \gamma)\mathbf{x})$. Since the Fourier transform of ρ', denoted by P', is $P'(\mathbf{k}) = P(R_E^{-1}(\alpha, \beta, \gamma)\mathbf{k})$ it follows that $\Sigma(\boldsymbol{\kappa}) = P'((\boldsymbol{\kappa}^T, 0)^T) = P(R_E^{-1}(\alpha, \beta, \gamma)(\boldsymbol{\kappa}^T, 0)^T)$ which is the 3-D projection slice theorem. Therefore the Fourier transform of the measured image is

$$(1) \qquad \Sigma^i(\boldsymbol{\kappa}) = \exp(-i\boldsymbol{\kappa}^T\boldsymbol{\chi}_0)G(\kappa)P(R_E^{-1}(\alpha, \beta, \gamma)(\boldsymbol{\kappa}^T, 0)^T).$$

2. Current methods for computing 3-D reconstructions. A large body of research has been devoted to methods for computing 3-D reconstructions. Detailed discussions, with extensive references, include Refs. [6, 7, 8, 9, 10, 11]. Assume that $G(\kappa)$ is known or even that $G(\kappa) = 1$. The key complexity in Eq. 1 is then that the values of $\boldsymbol{\chi}_0 \in \mathbb{R}^2$, $\alpha \in [0, 2\pi)$, $\beta \in [0, \pi]$, and $\gamma \in [0, 2\pi)$, which differ from image to image, are unknown. If these five numbers were known, then Eq. 1 describes a linear relationship between the 2-D Fourier transform of the image and the 3-D

Fourier transform of the particle's electron scattering intensity evaluated on a 2-D plane through the origin so standard tomographic ideas can be applied. Therefore an important theme in Refs. [6, 7, 8, 9, 10, 11] is to estimate values for $z = (\boldsymbol{\chi}_0, \alpha, \beta, \gamma)$, use these values with Eq. 1 to compute a 3-D reconstruction, and then iterate where iteration involves improving the accuracy of the z values, including additional images that were initially excluded, increasing the range of image spatial frequencies that are used in the reconstruction, and increasing the spatial resolution of the reconstruction. Because $\boldsymbol{\chi}_0$ can be estimated by methods as simple as human examination of each image, we will focus on methods for estimating (α, β, γ). In particular, we will focus on the so-called common lines methods and correlation methods.

The common lines method (e.g., Ref. [12] and the references therein) exploits the symmetry of the particle. Let $R_g \in \mathbb{R}^{3 \times 3}$, $R_g^{-1} = R_g^T$, $\det(R_g) = +1$, $g \in \{0, \dots, N_g - 1\}$, and $R_0 = I_3$ (the 3×3 identity matrix) be matrices that describe the symmetry, i.e., $\rho(\mathbf{x}) = \rho(R_g^{-1}\mathbf{x})$ for all $g \in \{0, \dots, N_g - 1\}$ and all $\mathbf{x} \in \mathbb{R}^3$. By direct computation it can be shown that the Fourier transform $P(\mathbf{k})$ has the same symmetry: $P(\mathbf{k}) = P(R_g^{-1}\mathbf{k})$ for all $g \in \{0, \dots, N_g - 1\}$ and all $\mathbf{k} \in \mathbb{R}^3$. Ignoring the CTF ($G(\kappa)$) and origin uncertainty ($\boldsymbol{\chi}_0$), the image formation process is $\Sigma_0(\boldsymbol{\kappa}) = P(R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T, 0)^T)$. By the symmetry of $P(\cdot)$ it follows that $\Sigma_0(\boldsymbol{\kappa}) = P(R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T, 0)^T) = P(R_g^{-1}R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T, 0)^T) = \Sigma_g(\boldsymbol{\kappa})$. $\Sigma_0(\cdot)$ and $\Sigma_g(\cdot)$ are two identical images that come from evaluating $P(\cdot)$ on different slices through $\mathbf{k}$ space, specifically, the slices defined by $R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T, 0)^T$ and $R_g^{-1}R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T, 0)^T$. Both slices include the origin in $\mathbf{k}$ space and intersect in a line. On the line of intersection, the two images are equal. However, and this is the key fact clearly illustrated in Ref. [10, Figs. 7.68–7.69, pp. 245–246], the intersection line in $\Sigma_0(\cdot)$ is a certain $\boldsymbol{\kappa}$ line while the intersection line in $\Sigma_g(\cdot)$ is a different $\boldsymbol{\kappa}$ line. Since $\Sigma_0(\boldsymbol{\kappa}) = \Sigma_g(\boldsymbol{\kappa})$, this construction has located two lines in the image (whose location is a function of g, α, β, γ), along which the image has the same value. These are the so-called common lines. The common lines are used to determine (α, β, γ) for a particular image by the following type of procedure: Define a measure of difference for the values of $\Sigma(\cdot)$ along two radially directed lines. This measure usually depends on the phases of the complex-valued $\Sigma(\cdot)$ image. Then estimate (α, β, γ) to have that value which results in a minimum value for the measure when evaluated on the two lines determined by $(g, \alpha, \beta, \gamma)$. Because the images are small (e.g., $10^2 \times 10^2$ pixels) and noisy (e.g., SNR less than 1), this would be an inaccurate estimator if applied to only one pair of common lines. Therefore, it is applied jointly to all possible pairs of common lines which, in the common case of icosahedral symmetry, is 37 pairs of common lines [12, p. 49]. The common lines are computed in terms of standard matrix operations as follows: Define $S_{g,\alpha,\beta,\gamma} = R_{\alpha,\beta,\gamma}R_g$. Define a partition of $S_{g,\alpha,\beta,\gamma}$ such that the first two rows are $U_{g,\alpha,\beta,\gamma} \in \mathbb{R}^{2 \times 3}$ and the third row is $\mathbf{v}_{g,\alpha,\beta,\gamma}^T \in \mathbb{R}^{1 \times 3}$. $\Sigma_0(\cdot)$ and $\Sigma_g(\cdot)$ are identical images which, however, depend on $P(\cdot)$ evaluated on different planes in

$\mathbf{k}$ space. The two planes can be described by $\mathbf{v}_{0,\alpha,\beta,\gamma}^T\mathbf{k} = 0$ and $\mathbf{v}_{g,\alpha,\beta,\gamma}^T\mathbf{k} = 0$, respectively. Both planes include the origin in $\mathbf{k}$ space and, in the general case, intersect in a line[1]. In order to describe this line, define

$$V_{g,\alpha,\beta,\gamma} = \left[\begin{array}{c} \mathbf{v}_{0,\alpha,\beta,\gamma}^T \\ \mathbf{v}_{g,\alpha,\beta,\gamma}^T \end{array} \right] \in \mathbb{R}^{2\times 3}$$

which has a 1-D null space described by the unit vector $\mathbf{u}_{g,\alpha,\beta,\gamma} \in \mathbb{R}^3$. Then, for the first image, $\mathbf{k} = R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow R_{\alpha,\beta,\gamma}\mathbf{k} = (\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \mathbf{v}_{0,\alpha,\beta,\gamma}^T\mathbf{k} = 0$ and, for the second image, $\mathbf{k} = R_g^{-1}R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \mathbf{k} = (R_{\alpha,\beta,\gamma}R_g)^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow R_{\alpha,\beta,\gamma}R_g\mathbf{k} = (\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \mathbf{v}_{g,\alpha,\beta,\gamma}^T\mathbf{k} = 0$ so the line of intersection is determined by $V_{g,\alpha,\beta,\gamma}\mathbf{k} = (0,0)^T$ which has the solution $\mathbf{k} = \mathbf{u}_{g,\alpha,\beta,\gamma}t$ for $t \in \mathbb{R}$. This line in $\mathbf{k}$ space corresponds to different lines in $\boldsymbol{\kappa}$ space in the two images. Specifically, for the first image, $\mathbf{k} = R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \mathbf{u}_{0,\alpha,\beta,\gamma}t = R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow R_{\alpha,\beta,\gamma}\mathbf{u}_{0,\alpha,\beta,\gamma}t = (\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \boldsymbol{\kappa} = U_{0,\alpha,\beta,\gamma}\mathbf{u}_{0,\alpha,\beta,\gamma}t$ and, for the second image, $\mathbf{k} = R_g^{-1}R_{\alpha,\beta,\gamma}^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \mathbf{k} = (R_{\alpha,\beta,\gamma}R_g)^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \mathbf{u}_{g,\alpha,\beta,\gamma}t = (R_{\alpha,\beta,\gamma}R_g)^{-1}(\boldsymbol{\kappa}^T,0)^T \Leftrightarrow R_{\alpha,\beta,\gamma}R_g\mathbf{u}_{g,\alpha,\beta,\gamma}t = (\boldsymbol{\kappa}^T,0)^T \Leftrightarrow \boldsymbol{\kappa} = U_{g,\alpha,\beta,\gamma}\mathbf{u}_{g,\alpha,\beta,\gamma}t$ so the two lines are $\boldsymbol{\kappa} = U_{0,\alpha,\beta,\gamma}\mathbf{u}_{0,\alpha,\beta,\gamma}t$ and $\boldsymbol{\kappa} = U_{g,\alpha,\beta,\gamma}\mathbf{u}_{g,\alpha,\beta,\gamma}t$ for $t \in \mathbb{R}$.

The correlation method (e.g., Ref. [13] and the references therein) is based on a 3-D mathematical model for the electron scattering intensity $\rho(\mathbf{x})$. From this model, a library of projections at different projection orientations (i.e., different values of (α,β,γ)) is computed. In order to determine the projection orientation (i.e., a value of (α,β,γ)) for a particular image, the image is correlated with each projection in the library and the estimated projection orientation is chosen to be the orientation of the library member with highest correlation. This search procedure can be accelerated by using so-called polar Fourier transforms [13]. Once a projection orientation for each image has been estimated, a 3-D reconstruction can be computed and the resulting reconstruction used as a new model in an iterative procedure. An important issue is how to determine the initial model. The sources that have been used include [13] (1) idealized models, such as an icosahedrally-symmetric collection of spheres with the correct total scattering intensity and the correct quasisymmetry [14]; (2) 3-D reconstructions from any source (e.g., crystal x-ray diffraction or cryo electron microscopy) of a virus expected to have a similar structure to the virus under investigation; and (3) a 3-D reconstruction of the virus under investigation based on a single image.

3. Our approach to computing 3-D reconstructions. We are developing a system for computing 3-D reconstructions from cryo EM images [15, 16, 17, 18, 19]. Related work includes the "ROSE" algorithm [20, 21, 22] and the algorithm of Ref. [23]. Please see Ref. [15] for a detailed comparison. Our approach is fundamentally statistical, in particular, it is a maximum likelihood estimator. The particle (we

---

[1] We will not consider all of the details needed to deal with the special cases.

have focused on viruses) is represented as a weighted sum of basis functions which are products of harmonics and radial functions and therefore 3-D reconstruction is equivalent to determining the vector of weights. This approach is particularly attractive for viruses which generally have rotational symmetries because the symmetry operations then effect only the harmonic factor of the basis function. The most common symmetry is icosahedral symmetry, which is the symmetry we have focused on and which we exactly incorporate without needing constraints on the weights by using icosahedral harmonics [24, 25].

After sampling the image in $\boldsymbol{\kappa}$ space and grouping the samples from the ith image as a column vector denoted by y_i, the measurement model, based on Eq. 1, is [17]

$$(2) \qquad\qquad y_i = L(z_i)c + n_i, i \in \{1, \ldots, N_v\}$$

where c is the vector of weights, y_i is the ith image represented as a vector, z_i is the projection orientation (three Euler angles) and particle center location in the image (a two-component vector in a Euclidean space) for the ith image, $L(z_i)$ is the resulting linear transformation from weights to noise-free image, and n_i is the noise represented as a vector. Typical dimensions [15] are images of 117×117 pixels, 720 weights, and $N_v = 583$ images. We assume that c is deterministic but unknown. We assume that z_i are i.i.d. random variables where the projection orientation components are independent of the particle center location components, the orientations are uniform over the product of the unit sphere and the unit circle, and the center location is uniform over a 2-D disk. We assume that the n_i are i.i.d. zero-mean Gaussian random variables with a diagonal covariance matrix which we adaptively estimate from the image data. We assume that z_i and n_i are independent.

Using the statistical model of the previous paragraph, we solve a maximum likelihood (ML) estimation problem for the c weights after integrating out the z_i variables which we regard as nuisance parameters. Our primary software solves the ML optimization problem using an expectation maximization algorithm, where the z_i are the "missing" data, although we are currently investigating alternatives such as conjugate gradients. This is quite different from most existing systems [6, 7, 8, 9, 10, 11, 26, 27, 28, 29] which iterate between estimating orientations and center locations (i.e., z_i) and estimating 3-D structure (i.e., c) and biologists are sometimes even surprised that it is possible to compute reconstructions without estimating the projection orientation of each image.

In this paper three aspects of this problem are investigated in an attempt to contrast our work with more standard approachs: First, the ML estimator of the previous paragraph is compared with an ML-MAP (maximum *a posteriori*) estimator where the c weights are estimated by ML and the z_i nuisance parameters are estimated by MAP. The ML-MAP estimator is roughly the customary approach to these problems in the biological literature. Note that in the cryo EM problem, there is no *a priori*

probabilistic information available about c and hence performance has to be determined as a function of c. The second aspect considered is the performance of both ML and ML-MAP estimators when the *a priori* probability distribution used in the MAP calculations is not accurate. This problem arises in practice for the following reason. The orientation of the particle in the electron microscope is random but the probability distribution is complicated due to the physics. In particular, the specimen is a thin film, rapidly frozen, of particles in water and the orientation of particles in the film is, in part, controlled by their structure by mechanisms such as the interaction of the particle with the water-air interface. In practice, the histogram of estimated projection directions (which is equivalent to the histogram of particle orientations) is not uniform but rather shows peaks [11, p. 871]. This nonuniform probability distribution is difficult to model, for instance, because of its dependence on the unknown structure, and so we have used a uniform distribution but desire to have at least a qualitative understanding of the effect of that assumption. Finally, the third aspect considered is estimator performance: for the ML estimator described in Ref. [17], we have not been able to compute its performance analytically and, because each 3-D reconstruction computation described in Ref. [15] takes a day or two on a 32-node PC cluster, we have also not been able to compute performance by Monte-Carlo methods. Therefore, in this paper, simpler model problems are considered which allow performance computations (bias and estimation error variance) to be performed by Monte-Carlo methods.

The problem described in Eq. 2 is actually simpler than the main problem described in Refs. [15, 16, 17]. In particular the problem described in Eq. 2 assumes that all images are of the same type of particle. However, viruses undergo a sequences of changes, called maturation, as they self-assemble into an infectious particle. This process is of great biological interest and experiments have been done (and it is straightforward to conceive of further experiments of this type) where the thin film of particles in water contains a mixture of particles in different maturation states and therefore the cryo EM images show a mixture of particle types. This motivates the generalization of the problem described in Eq. 2 to

$$(3) \qquad\qquad y_i = L(z_i)c_{\eta_i} + n_i, i \in \{1, \ldots, N_v\}$$

where $\eta_i \in \{1, \ldots, N_\eta\}$ is a label describing the type of particle (e.g., maturation state) shown in the ith image and N_η is known. The goal is to estimate the 3-D reconstruction of each type of particle. Thus this is a joint pattern recognition and 3-D signal reconstruction problem. In Refs. [15, 16, 17] the η_i are modeled as i.i.d. with a known probability mass function and the η_i are merged with the z_i to form a more complicated nuisance parameter. The resulting nuisance parameter is then integrated out and an ML problem is solved for c. However, it would be of interest to consider ML-MAP estimators which treat 3-D reconstruction and pattern recognition

more equally. The reason the simpler problem of Eq. 2 is emphasized in this paper is that it is difficult to find model problems corresponding to Eq. 3 for which analytical calculations are possible.

4. The definition of the estimators. If r is the data and ω are the parameters then a ML estimate is $\hat{\omega} = \arg\max_{\omega} p(r|\omega)$ and a MAP estimate is $\hat{\omega} = \arg\max_{\omega} p(\omega|r) = \arg\max_{\omega} p(\omega, r)$. When ω can be partitioned into ω_1, which is deterministic but unknown (corresponding to the 3-D virus structure) and ω_2, which is Bayesian (corresponding to the projection orientations and boxed image origin locations) then there are two natural approaches to estimating ω_1:

$$\hat{\omega}_1^{\mathrm{ML}} = \arg\max_{\omega_1} \int p(r|\omega_1, \omega_2)p(\omega_2)\mathrm{d}\omega_2$$

$$\hat{\omega}_1^{\mathrm{ML-MAP}} = \arg\max_{\omega_1} \max_{\omega_2} p(r|\omega_1, \omega_2)p(\omega_2).$$

The properties of both of these estimators are considered in this paper.

5. Model problems.

5.1. Random (real-valued) gain model. The first model problem has a collection of N_v real-valued measurements that are related to the structure by a random gain:

$$y_i = z_i c + n_i$$

$$i \in \{1, \ldots, N_v\}$$

$$z_i \sim \text{i.i.d. } N(m_z, \sigma_z^2)$$

$$n_i \sim \text{i.i.d. } N(0, \sigma^2),$$

where $N(m, \sigma^2)$ denotes the Gaussian probability density function with mean m and variance σ^2 and the sequences $\{z_i\}$ and $\{n_i\}$ are independent. By direct computation it can be shown that

$$(4) \qquad p(y_i|z_i, c)p(z_i) = \frac{s(c)}{\sqrt{2\pi}\sigma\sigma_z} \exp\left(-\frac{1}{2}\delta_i(c)\right) N\left(\mu_i(c), s^2(c)\right)(z_i)$$

$$\mu_i(c) = \frac{y_i c \sigma_z^2 + m_z \sigma^2}{c^2 \sigma_z^2 + \sigma^2}$$

$$s^2(c) = \frac{\sigma^2 \sigma_z^2}{c^2 \sigma_z^2 + \sigma^2}$$

$$\delta_i(c) = \frac{(y_i - c m_z)^2}{c^2 \sigma_z^2 + \sigma^2}.$$

After further computing, it can be shown that the ML estimator is determined by the real roots of the cubic polynomial

$$0 = c^3 \left(\sigma_z^4\right) + c^2 \left(m_z \sigma_z^2 \bar{y}\right) + c^1 \left[\sigma^2 \left(\sigma_z^2 + m_z^2\right) - \sigma_z^2 r_y\right] + c^0 \left(m_z \sigma^2 \bar{y}\right)$$

$$\bar{y} = \frac{1}{N_v} \sum_{i=1}^{N_v} y_i$$

$$r_y = \frac{1}{N_v} \sum_{i=1}^{N_v} y_i^2.$$

The particular root is that root which maximizes the log likelihood function

$$l_{\mathrm{ML}}(c) = -\frac{N_v}{2} \ln \left(c^2 \sigma_z^2 + \sigma^2\right) - \frac{N_v}{2} \ln (2\pi) - \frac{N_v}{2} \frac{r_y - 2m_z \bar{y} c + m_z^2 c^2}{c^2 \sigma_z^2 + \sigma^2}.$$

After further computing based on Eq. 4, it can be shown that the ML-MAP estimator is determined by the roots of the quadratic polynomial

$$0 = c^2 \left(\bar{y} m_z \sigma_z^2\right) + c^1 \left(m_z^2 \sigma^2 - \sigma_z^2 r_y\right) + c^0 \left(-\bar{y} m_z \sigma^2\right).$$

Since the discriminant is $\left(m_z^2 \sigma^2 - \sigma_z^2 r_y\right)^2 + (2\bar{y} m_z)^2 \sigma_z^2 \sigma^2 \geq 0$, it follows that the roots are always real. The particular root is that root which maximizes the maximum (with respect to z_i, $i \in \{1, \ldots, N_v\}$) of the log likelihood function

$$l_{\mathrm{ML-MAP}}(c) = -N_v \ln (2\pi\sigma\sigma_z) - \frac{N_v}{2} \frac{r_y - 2\bar{y} c m_z + c^2 m_z^2}{c^2 \sigma_z^2 + \sigma^2}.$$

The measure of SNR used in cryo EM is the square root of the ratio of energy in the 2-D projection of the 3-D scattering intensity to energy in the noise. Therefore, the analogous measure in the model problem is

$$\mathrm{SNR}_1(c) = \sqrt{\mathrm{E}[(z_i c)^2]/\mathrm{E}[n_i^2]}$$

(5)
$$= c\sqrt{\sigma_z^2 + m_z^2}/\sigma.$$

5.2. Random phase model. The second model problem has a collection of N_v complex-valued measurements that are related to the structure by a random phase (i.e., a rotation in the 2-D complex plane):

$$y_i = \exp(j\phi_i)c + n_i$$

$$i \in \{1, \ldots, N_v\}$$

$$\phi_i \sim \text{i.i.d. uniform on } [-\pi, +\pi]$$

$$(\Re\{n_i\}, \Im\{n_i\})' \sim \text{i.i.d. } N\left(\mathbf{0}, \sigma^2 I_2\right),$$

where the sequences $\{\phi_i\}$ and $\{n_i\}$ are independent, $\mathbf{0} = (0,0)'$, I_2 is the 2×2 identity matrix, and $j = \sqrt{-1}$. We allow c to be complex although we will only be able to estimate $|c|$ from this data.

By direct computation it can be shown that the log likelihood function for the ML estimate of c is

$$(6) \qquad l_{\mathrm{ML}}(c) = -N_v \ln \left(2\pi\sigma^2\right) - \frac{N_v}{2\sigma^2} \left(r_y + |c|^2\right) + \sum_{i=1}^{N_v} \ln I_0 \left(\frac{|c||y_i|}{\sigma^2}\right),$$

where $I_\nu(\cdot)$ is the modified Bessel function of the first kind of order ν [30, p. 376, Eq. 9.6.19]. Notice that there is no sufficient statistic for this problem and that the log likelihood function is a function of $|c|$ alone. We have not been able to analytically maximize this expression with respect to c and thereby determine the estimator analytically. Therefore we use numerical optimization based on the log likelihood and, optionally, its gradient which for $c \in \mathbb{R}$ and $c \geq 0$ is [30, p. 376, Eq. 9.6.27]

$$\frac{\mathrm{d}l_{\mathrm{ML}}}{\mathrm{d}c}(c) = -\frac{N_v}{\sigma^2}c + \sum_{i=1}^{N_v} \frac{|y_i|}{\sigma^2} \frac{I_1\left(\frac{c|y_i|}{\sigma^2}\right)}{I_0\left(\frac{c|y_i|}{\sigma^2}\right)}.$$

In order to check for a boundary maximum, it is also necessary to know that

$$l_{\mathrm{ML}}(0) = -N_v \ln \left(2\pi\sigma^2\right) - \frac{N_v}{2\sigma^2}r_y$$

which follows from $I_0(0) = 1$. Finally, it can be shown that the log likelihood function for the ML-MAP estimator is

$$l_{\mathrm{ML-MAP}}(c) = -N_v \ln \left(2\pi\sigma^2\right) - N_v \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{N_v} \left(|y_i| - |c|\right)^2$$

so that the ML-MAP estimator is

$$(7) \qquad \hat{c}_{\mathrm{ML-MAP}} = m_y \exp(j\theta)$$

$$(8) \qquad m_y = \frac{1}{N_v} \sum_{i=1}^{N_v} |y_i|$$

for arbitrary $\theta \in \mathbb{R}$.

The origin of the SNR definition used in cryo EM was described in Subsection 5.1. Applying that definition to the second model problem gives the result that

$$\mathrm{SNR}_2(c) = \sqrt{\mathrm{E}[|\exp(j\phi_i)c|^2]/\mathrm{E}[|n_i|^2]}$$

$$(9) \qquad\qquad\qquad = c/(\sqrt{2}\sigma).$$

6. Numerical results. All performance calculations were done by Monte Carlo using 1000 trials. Since c is deterministic but unknown, performance is described as a function of the true value of c. For each value of c, the computations used the same pseudo-random numbers in order to achieve variance reduction. For the second model problem, the ML estimate is found using Matlab function `fminbnd` which is

a scalar optimization algorithm that exploits bounds on the feasible set which were $\max(0, \min(c-1, .8c))$ and $\max(c+1, 1.2c)$. The other three estimators are computed directly from the formulae given in Section 5.

We have explored the performance of these four estimators over a range of parameter values for both matched and mismatched cases. Here we can only give typical examples and have chosen examples close to the cryo EM application, especially in terms of the SNR value of 1. In Figure 1 the matched performance is shown. In the first model problem with gain z_i the ML-MAP estimator is clearly inferior to the ML estimator since the ML-MAP estimator has both larger bias and larger variance. In the second model problem with phase ϕ_i the situation is more ambiguous since the ML-MAP estimator has larger bias but smaller variance than the ML estimator. Therefore, in an *ad hoc* manner, we attempted to find an estimator with a more attractive bias-variance tradeoff by subtracting the bias of the ML-MAP estimator at the estimated value of c from the estimate. Specifically, define $b_{\mathrm{ML-MAP}}(c) = E[\hat{c}_{\mathrm{ML-MAP}}(y)|c] - c$ which is the bias. [$b_{\mathrm{ML-MAP}}(\cdot)$ is computed by Monte-Carlo on a 1-D grid with step size 0.05 and then is linearly interpolated between grid points.] Define a new estimator $\hat{\hat{c}}_{\mathrm{ML-MAP}}(y) = \hat{c}_{\mathrm{ML-MAP}}(y) - b(\hat{c}_{\mathrm{ML-MAP}})$. As shown in Figure 1, the $\hat{\hat{c}}_{\mathrm{ML-MAP}}(y)$ estimator has lower bias but greater variance than $\hat{c}_{\mathrm{ML-MAP}}(y)$. Neither $\hat{c}_{\mathrm{ML}}(y)$, $\hat{c}_{\mathrm{ML-MAP}}(y)$, nor $\hat{\hat{c}}_{\mathrm{ML-MAP}}(y)$ has both lower bias and lower variance so a choice among them must be based on application issues, including computational complexity which is markedly lower for $\hat{c}_{\mathrm{ML-MAP}}(y)$ since no numerical optimization is required to compute $\hat{c}_{\mathrm{ML-MAP}}(y)$. Note also that the $\hat{\hat{c}}_{\mathrm{ML-MAP}}(y)$ estimate cannot be computed in the cryo EM problem because the bias must be evaluated by Monte-Carlo methods which are far too costly in the cryo EM problem.

In Figure 2 the mismatched probability distribution and the resulting performance is shown. Recall that the mismatch is in the *a priori* probability distribution used in the MAP calculations, i.e., in the distribution for z_i or for ϕ_i. This is a severe mismatch. For instance, based on the strongly localized histograms seen in cryo EM problems, the probability distributions used to generate the data are discrete even though the distributions used to derive the estimators are continuous (Gaussian or uniform). Except for different scaling of the independent variable, the same mismatched distribution is used in both model problems. In the second model problem with phase ϕ_i the results for the mismatched case are similar to the results for the matched case. The reason is that the noise n_i has a circularly-symmetric probability density function (pdf). Specifically, the model is $y_i = \exp(j\phi_i)c + n_i$ which is equivalent to $y_i = \exp(j\phi_i)[c + \exp(-j\phi_i)n_i]$ which is equivalent to $y_i = \exp(j\phi_i)(c + n_i)$ because of the circular symmetry of the pdf of n_i. Then, since the estimators depend only on $|y_i|$ (Eq. 6 or Eqs. 7 and 8), it follows that performance is not altered if the pdf is mismatched. However, in the first model problem with gain z_i, results are significantly degraded by the mismatched pdf on z_i in comparison with the matched

case. In particular, typical biases are three times larger though the standard deviations are much less increased. Also, unlike the matched case where the ML estimator had both smaller bias and smaller variance, in the mismatched case the ML estimate has greater bias and smaller variance. It is difficult to know which model problem and what type of mismatch correctly models the cryo EM problem and so the conclusion regarding mismatched calculations seems not to favor one estimator over the other.

Finally, in order to show the connection with the cryo EM problem, examples of experimental cryo EM images and a surface rendering of the resulting 3-D reconstruction for Nudaurelia Omega Capensis Virus [31] are shown in Figure 3. Note the low SNR of the two typical cryo EM images that are shown. It is clearly difficult to precisely determine the center of the virus particle in these images and certainly difficult to extract much information about the projection orientation. These problems motivate our study of comprehensive statistical approaches to this 3-D reconstruction problem.

7. Discussion. Two model problems and two types of estimator have been studied. Unlike the situation in the motivating cryo EM problem, for these model problems it is possible to do both analytical and Monte-Carlo calculations. In these model problems the ML estimator appears to have a modest performance advantage over the ML-MAP estimator which, however, vanishes when the probabilistic assumptions on the nuisance parameters are mismatched. In order to make concrete the connection with cryo EM, an example of a 3-D reconstruction of Nudaurelia Omega Capensis Virus is also given. Finally, as described in detail in the final paragraph of Section 3, the most interesting cryo EM problem combines 3-D reconstruction with pattern classification in order to determine which image comes from which type of virus particle. However, even the simplest model problem for such joint pattern classification and signal reconstruction problems is a very challenging problem from the point of view of analytical calculations and so we have not discussed them in this paper.

8. Acknowledgements. We would like to thank Professor John E. Johnson (The Scripps Research Institute) for motivating this work and for providing the cryo EM images used in Figure 3.

REFERENCES

[1] O. SCHERZER, *The theoretical resolution limit of the electron microscope*, J. Appl. Phys., 20(1949), pp. 20–29.

[2] HAROLD P. ERICKSON, *The Fourier transform of an electron micrograph—First order and second order theory of image formation*, In: R. Barer and V. E. Cosslett, editors, Advances in Optical and Electron Microscopy (Volume 5), pages 163–199. Academic Press, London and New York, 1973.

[3] J. LEPAULT AND T. PITT, *Projected structure of unstained, frozen-hydrated T-layer of Bacillus brevis*, EMBO J., 3:1(1984), pp. 101–105.
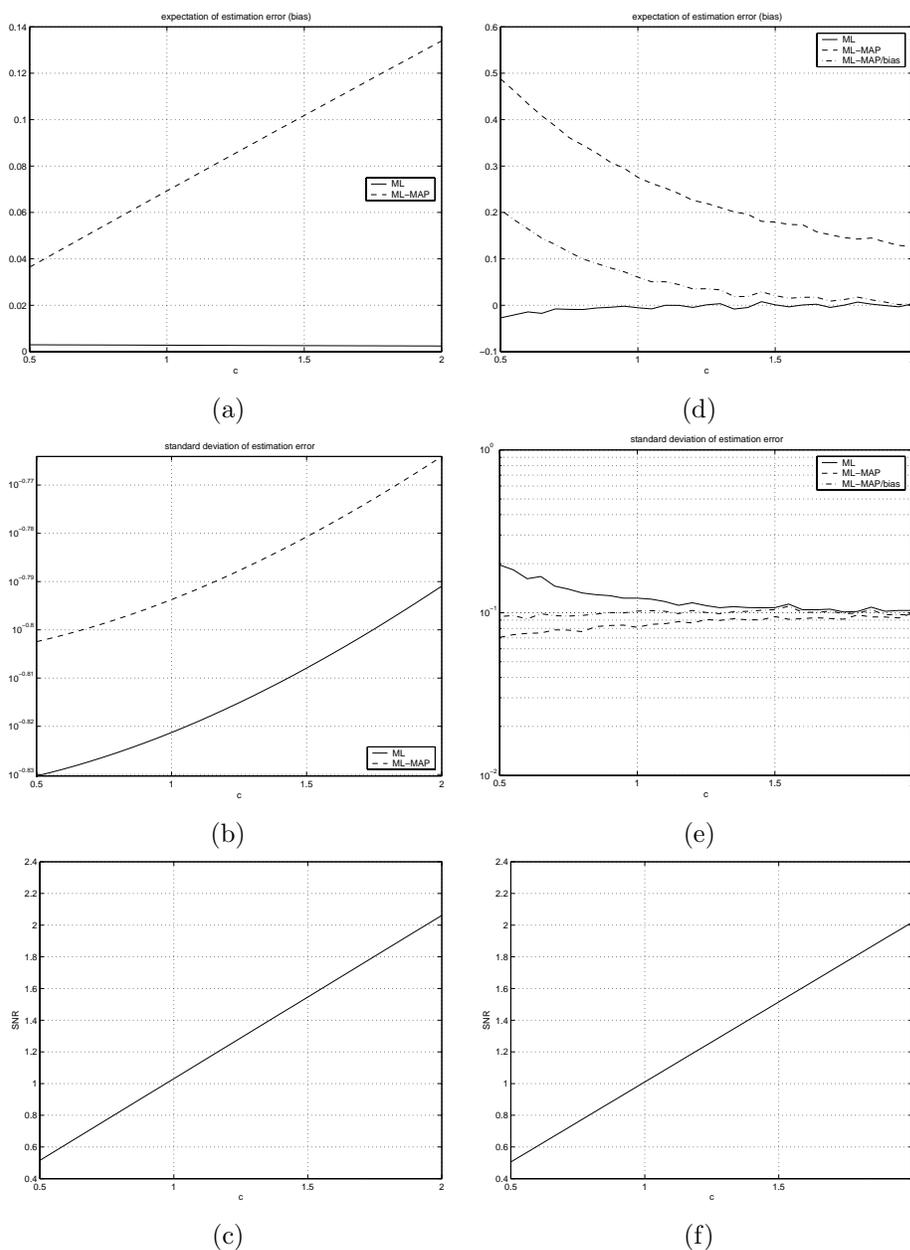
FIG. 1. *Matched case: all assumptions used to derive the estimator are also used to cre-ate the data. The legend labels ML, ML-MAP, and ML-MAP/bias refer to estimators $\hat{c}_{\mathrm{ML}}(y)$, $\hat{c}_{\mathrm{ML-MAP}}(y)$, and $\hat{\hat{c}}_{\mathrm{ML-MAP}}(y)$, respectively. Panels (a)–(c): The first model problem (i.e., real-valued gain): bias for c, estimation error standard deviation for c, and SNR (Eq. 5), respectively. The parameters are $N_v = 50$, $m_z = 1.0$, $\sigma_z = 0.25$, and $\sigma_n = 1.0$. Panels (d)–(f): The second model problem (i.e., phase): bias for c, estimation error standard deviation for c, and SNR (Eq. 9), respectively. The parameters are $N_v = 50$ and $\sigma = 0.7$.*
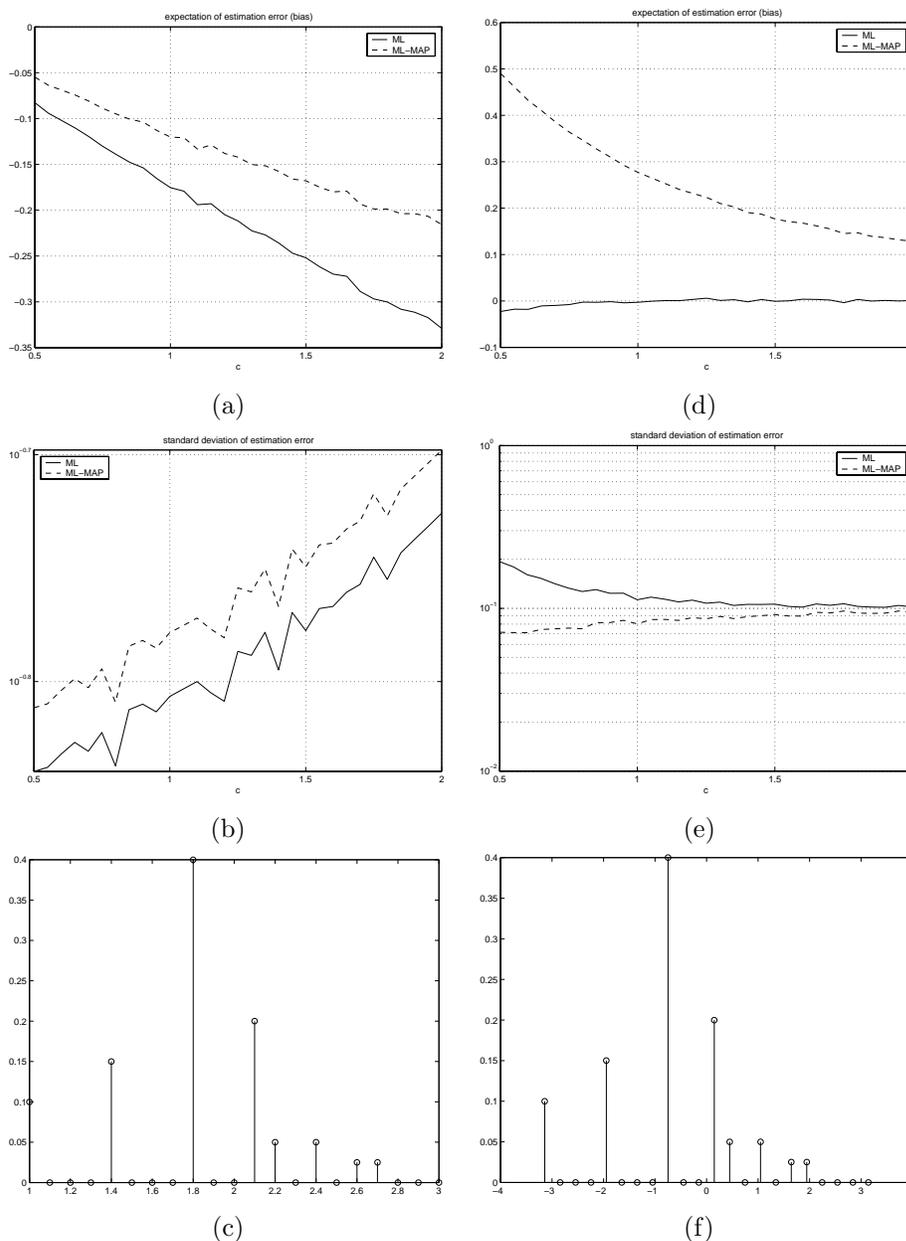
(a)

(d)

(b)

(e)

(c)

(f)

FIG. 2. *Mismatched case: the probability density function (pdf) for the gain z_i or phase ϕ_i used to derive the estimator and used to create the data are different. The legend labels ML and ML-MAP refer to estimators $\hat{c}_{\mathrm{ML}}(y)$ and $\hat{c}_{\mathrm{ML-MAP}}(y)$, respectively. Panels (a)–(c): The first model problem (i.e., real-valued gain): bias for c, estimation error standard deviation for c, and actual pdf (it is discrete) on the gain z_i, respectively. The parameters used in the estimator are $N_v = 50$, $m_z = 1.0$, $\sigma_z = 0.25$, and $\sigma_n = 1.0$. Panels (d)–(f): The second model problem (i.e., phase): bias for c, estimation error standard deviation for c, and actual pdf (it is discrete) on the phase ϕ_i, respectively. The parameters used in the estimator are $N_v = 50$ and $\sigma = 0.7$.*

[4] Chikashi Toyoshima and Nigel Unwin, *Contrast transfer for frozen-hydrated specimens: Determination from pairs of defocused images*, Ultramicroscopy, 25:4(1988), pp. 279–292.

[5] Stanley R. Deans, *The Radon Transform and Some of Its Applications*, Krieger Publishing Company, 1993.

[6] B. Carragher and P. R. Smith, *Special issue: Advances in computational image processing for microscopy*, J. Struct. Biol., 116:1(1996), pp. 1–249.

[7] Wah Chiu and Steven Ludtke, *Special issue: Single-particle analysis*, J. Struct. Biol., 133:2/3(2001), pp. 89–266.

[8] Bridget Carragher and Pawel Penczek, *Special issue: Analytical methods and software tools for macromolecular microscopy*, J. Struct. Biol., 133:1/2(2003), pp. 1–252.

[9] Joachim Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, San Diego, 1996.

[10] M. F. Moody, *Image analysis of electron micrographs*, In: P. W. Hawkes and U. Valdrè, editors, Biophysical Electron Microscopy: Basic Concepts and Modern Techniques, chapter 7, pages 145–287. Academic Press, 1990.

[11] T. S. Baker, N. H. Olson, and S. D. Fuller, *Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs*, Microbiology and Molecular Biology Reviews, 63:4(1999), pp. 862–922.

[12] S. D. Fuller, S. J. Butcher, R. H. Cheng, and T. S. Baker, *Three-dimensional reconstruction of icosahedral particles—The uncommon line*, J. Struct. Biol., 116(1996), pp. 48–55.

[13] Timothy S. Baker and R. Holland Cheng, *A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy*, J. Struct. Biol., 116(1996), pp. 120–130.

[14] D. L. D. Caspar and A. Klug, *Physical principles in the construction of regular viruses*, Cold Spring Harbor Symposium on Quantitative Biology, 27(1962), pp. 1–24.

[15] Zhye Yin, Yili Zheng, Peter C. Doerschuk, Padmaja Natarajan, and John E. Johnson, *A statistical approach to computer processing of cryo electron microscope images: Virion classification and 3-D reconstruction*, J. Struct. Biol., 144(2003), pp. 24–50.

[16] Zhye Yin, Yili Zheng, and Peter C. Doerschuk, *An ab initio algorithm for low-resolution 3-D reconstructions from cryo electron microscopy images*, J. Struct. Biol., 133:2/3(2001), pp. 132–142.

[17] Peter C. Doerschuk and John E. Johnson, *Ab initio reconstruction and experimental design for cryo electron microscopy*, IEEE Trans. Info. Theory, 46:5(2000), pp. 1714–1729.

[18] Yili Zheng, *Parallel implementations of 3-D reconstruction algorithms for cryo electron microscopy: A comparative study*, Master's thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayeete, Indiana, USA, July 2002.

[19] Zhye Yin, *Maximum likelihood 3-D virus reconstruction from projections of unknown orientation and cryo electron microscopy application*, PhD thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA, May, 2003.

[20] R. H. Vogel, S. W. Provencher, C.-H. von Bonsdorff, M. Adrian, and J. Dubochet, *Envelope structure of Semliki Forest virus recosntructed from cryo-electron micrographs*, Nature, 320(1986), pp. 533–535.

[21] Stephen W. Provencher and Robert H. Vogel, *Three-dimensional reconstructions from electron micrographs of disordered specimens: I. Method*, Ultramicroscopy, 25(1988), pp. 209–222.

[22] Robert H. Vogel and Stephen W. Provencher, *Three-dimensional reconstructions from electron micrographs of disordered specimens: II. Implementation and results*, Ultramicroscopy, 25(1998), pp. 223–240.

[23] Z. Kam and I. Gafni, *Three-dimensional reconstruction of the shape of human wart virus*

using spatial correlations, Ultramicroscopy, 17(1985), pp. 251–262.

[24] Yibin Zheng and Peter C. Doerschuk, *Explicit computation of orthonormal symmetrized harmonics with application to the identity representation of the icosahedral group*, SIAM Journal on Mathematical Analysis, 32:3(2000), pp. 538–554.

[25] Yibin Zheng and Peter C. Doerschuk, *Explicit orthonormal fixed bases for spaces of functions that are totally symmetric under the rotational symmetries of a Platonic solid*, Acta Cryst., A52(1996), pp. 221–235.

[26] Peter D. Lauren and N. Nandhakumar, *Estimating the viewing parameters of random, noisy projections of asymmetric objects for tomographic reconstruction*, IEEE Trans. PAMI, 19:5(1997), pp. 417–430.

[27] Ioana M. Martin and Dan C. Marinescu, *Concurrent computation and data visualization for spherical-virus structure determination*, IEEE Comput. Sci. Eng., pp. 40–52, October–December 1998.

[28] Christopher J. Lanczycki, Calvin A. Johnson, Benes L. Trus, James F. Conway, Alasdair C. Steven, and Robert L. Martino, *Parallel computing strategies for determining viral capsid structure by cryoelectron microscopy*, IEEE Comput. Sci. Eng., pp. 76–91, April–June 1998.

[29] Z. H. Zhou, W. Chiu, K. Haskell, H. Spears, Jr., J. Jakana, F. J. Rixon, and L. R. Scott, *Refinement of Herpesvirus B-capsid structure on parallel supercomputers*, Biophys. J., 74(1998), pp. 576–588.

[30] Milton Abramowitz and Irene A. Stegun, editors, *Handbook of Mathematical Functions*, volume 55 of *Applied Mathematics Series*, National Bureau of Standards, December 1972. Tenth Printing.

[31] S. Munshi, L. Liljas, J. Cavarelli, W. Bomu, B. McKinney, V. Reddy, and J. E. Johnson, *The 2.8Å structure of a T = 4 animal virus and its implications for membrane translocation of RNA*. J. Mol. Bio., 261:1(1996), pp. 1–10.
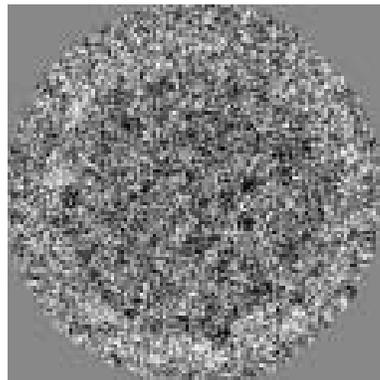
(a)
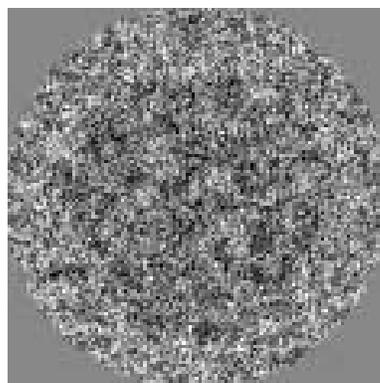
(b)
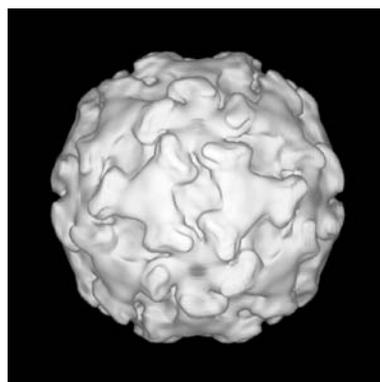
(c)

Fig. 3. *Two cryo EM images [Panels (a) and (b)] and a surface rendering of the resulting 3-D reconstruction [Panel (c)] for Nudaurelia Omega Capensis Virus [31]*