JOURNAL OF STRUCTURAL BIOLOGY **122,** 328–339 (1998)
ARTICLE NO. SB984014

A Maximum-Likelihood Approach to Single-Particle Image Refinement

F. J. Sigworth

Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06520-8026

Received April 9, 1998

The alignment of single-particle images fails at low signal-to-noise ratios and small particle sizes, because noise produces false peaks in the cross-correlation function used for alignment. A maximum-likelihood approach to the two-dimensional alignment problem is described which allows the underlying structure to be estimated from large data sets of very noisy images. Instead of finding the optimum alignment for each image, the algorithm forms a weighted sum over all possible in-plane rotations and translations of the image. The weighting factors, which are the probabilities of the image transformations, are computed as the exponential of a cross-correlation function. Simulated data sets were constructed and processed by the algorithm. The results demonstrate a greatly reduced sensitivity to the choice of a starting reference, and the ability to recover structures from large data sets having very low signal-to-noise ratios. © 1998 Academic Press

Key Words: **electron microscopy; maximum likelihood; single-particle alignment.**

INTRODUCTION

An important approach to the determination of macromolecular structures is the analysis of cryoelectronmicroscopic images of randomly oriented single particles (Frank, 1996). Paradoxically, it is the larger macromolecular assemblies, with molecular weights above about 1 MDa, for which the single-particle approach works best because particles of this size produce images that can be unambiguously aligned so that the necessary signal-averaging can be performed (Henderson, 1995). Smaller particles cannot be aligned because the signal-to-noise ratio of low-dose images does not allow the position and orientation of particles to be determined unambiguously. However, low-dose images of small particles obviously do contain some structural information. Can this information be somehow used to obtain high-resolution structures from large ensembles of small-particle images? This question was the impetus for the present study.

In many cases where there are ambiguous or unknown values of important variables, maximum-likelihood approaches have allowed the estimation of parameters of a model. For example, a maximum-likelihood approach has been applied to the phase extension of models in X-ray and electron-crystallographic analyses (Dong *et al.*, 1992; Doublie *et al.*, 1994); maximum-likelihood approaches have also been applied to the estimation of parameters for hidden Markov models of DNA sequences (Krogh *et al.*, 1994) and ion channel currents (Chung *et al.*, 1991; Venkataramanan *et al.*, 1998). This paper considers a maximum-likelihood approach to the problem of recovering 2D structural information from a set of small-particle images. We assume that these data images have been obtained from a population of identical particles that are oriented normal to the support film, so that they present an identical view. The orientation of the particles in the plane of the support film (described by an angle α) is, however, random. Thus each individual data image contains the projection of the particle whose position relative to the center of the image and whose rotation angle are not precisely known. The goal is to obtain a reconstructed projection image from a large number of such data images.

REVIEW OF THE ALIGNMENT PROBLEM

Maximizing the cross-correlation of images, followed by averaging of images thus aligned, is the standard method for combining the images of randomly positioned particles (Saxton and Frank, 1977). Summarized here is the “refinement” step of the reference-free alignment procedure described by Penczek *et al.* (1992), as applied to a two-dimensional object. In this step a tentative model for the projection structure of a particle is improved, making use of data in the form of a set of images. We shall assume that each of the N images is a noisy, translated, and rotated copy of the underlying true structure W, to which white Gaussian noise has been added. Examples of such images are shown in Fig. 1.

1047-8477/98 $25.00

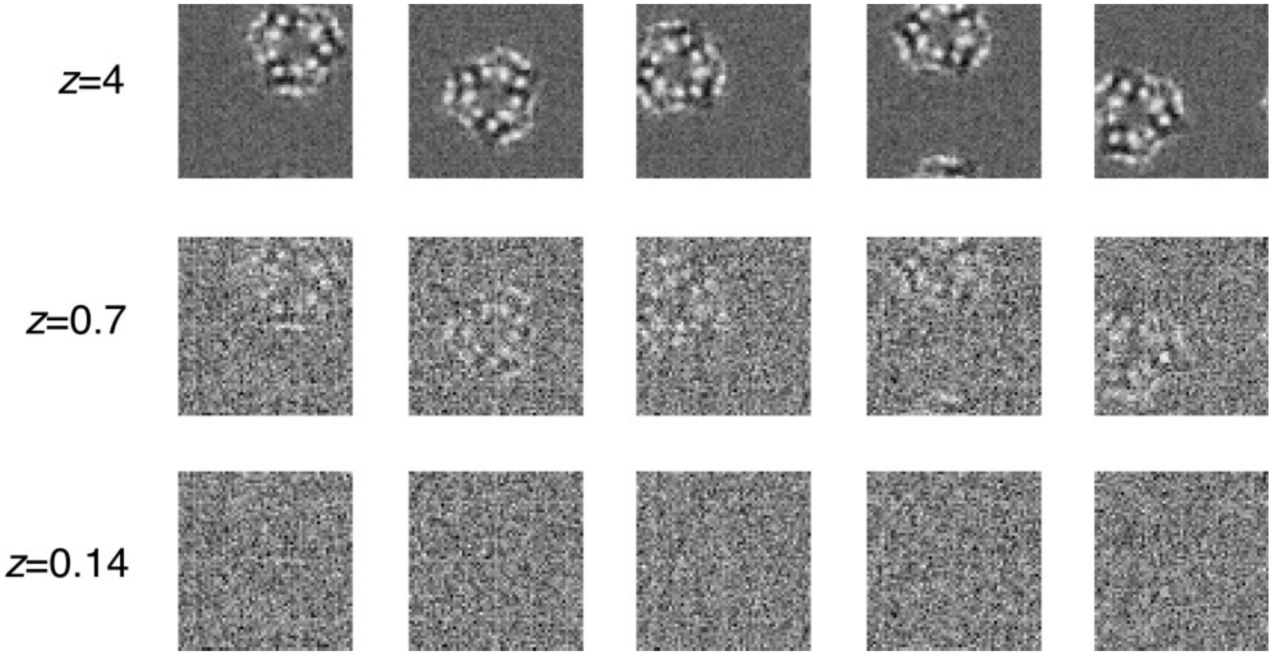

FIG. 1. Examples of synthetic images generated as in Eq. (1). The "true" structure W was derived from the z projection of the nonhydrogen atom density of a bacteriorhodopsin trimer (Grigorieff *et al.*, 1996). This projection was filtered with a Gaussian filter having half-power spatial frequency 0.1 Å^{-1}, scaled to have an rms amplitude of unity within the molecular boundary and sampled with 1.5-Å square pixels. The structure was scaled (by the factor z), subjected to random translations and rotations, and superimposed on Gaussian noise of unity variance. Examples of data with signal amplitudes $z = 4$, 0.7 and 0.14 are shown, which have SNR values of 16, 0.5, and 0.02, respectively.

The ith image is described by

$$X_i = zW(-\varphi_i) + G_i; \quad i = 1 \ldots N, \tag{1}$$

where z is a scaling factor and G_i is the noise. The signal-to-noise ratio, defined as the ratio of signal variance to noise variance (Frank, 1996), is proportional to z^2. The notation $W(-\varphi_i)$ describes that the structure W is mapped onto a transformed coordinate system, where $-\varphi_i$ defines the transformation (translation and rotation) applied to the structure. The inverse transformation φ_i, which returns the structure to its standard position, is parameterized by the rotation angle and the x and y translations so that we write $\varphi_i = (s_{\alpha i}, s_{xi}, s_{yi})$. Letting j be the index of a pixel, the value of an individual pixel in X_i is x_{ij}.

The goal of the refinement is to determine for the data set $\mathbf{X} = \{X_i; i = 1 \ldots N\}$ a set of transformation parameters $\Phi = \{\phi_i; i = 1 \ldots N\}$, where each $\phi_i = (q_{\alpha i}, q_{xi}, q_{yi})$ is the triplet of parameters for optimal alignment of image X_i. Given the set of parameters, the average image

$$A = \frac{1}{N}\sum_{i=1}^{N} X_i(\phi_i)$$

will be an optimal estimate of the original structure if the ϕ_i are equal to the true transformations φ_i in Eq. (1). The alignment parameters can be found by optimizing any of a number of related quantities (Penczek *et al.*, 1992). One of these is the squared magnitude of the summed images,

$$L'_1(\mathbf{X}, \Phi) = \left|\sum_{i=1}^{N} X_i(\phi_i)\right|^2. \tag{2}$$

A local maximum of this function can be obtained by iteration, as follows. Given the alignment after n iterations $\Phi^{(n)}$ and the corresponding average $A^{(n)}$, an improved alignment set $\Phi^{(n+1)}$ is obtained by maximizing the cross-correlation between each image X_i and the average,

$$\phi_i^{(n+1)} = \operatorname{argmax}_{\phi}[X_i(\phi) \cdot A^{(n)}], \quad i = 1 \ldots N. \tag{3}$$

Here the dot indicates an inner product between the images, such that

$$X \cdot A \equiv \sum_j x_j a_j$$

with the summation being taken over all M pixels in each image. The squared norm of an image X is defined as

$$|X|^2 \equiv (X \cdot X).$$

To make the result less dependent on the initial alignment parameters, Penczek *et al.* (1992) first subtract the old contribution of X_i to $A^{(n)}$ before performing the cross-correlation,

$$\phi_i^{(n+1)} = \mathrm{argmax}_{\phi}\left(X_i(\phi) \cdot \left[A^{(n)} - \frac{1}{N} X_i(\phi_i^{(n)})\right]\right), \quad i = 1 \ldots N. \tag{4}$$

Given the improved alignment set $\Phi^{(n+1)}$, a refined average $A^{(n+1)}$ is obtained as

$$A^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} X_i(\phi_i^{(n+1)}). \tag{5}$$

By repeating steps (4) and (5) a local maximum of the function in Eq. (2) is approached. If the signal-to-noise ratio is sufficiently high, the correct alignment parameters will be found and the global maximum will correspond to an estimate A that is closest to the true structure in a least-squares sense. If however the signal-to-noise ratio is low, false peaks in the cross-correlation function result in errors in the determination of the alignment parameters in Eq. (4) and the algorithm may converge to a maximum of Eq. (2) that does not correspond to the true structure.

MAXIMUM-LIKELIHOOD REFINEMENT

Consider an alternative approach that provides a maximum-likelihood estimate of the original structure W. As before, a function of the reconstructed image A is maximized; however, in this case no alignment set Φ is determined. Instead, the transformation variables ϕ_i are treated as hidden, random variables. The goal is to find the most likely parameter values for a model that describes the data set **X**. An appropriate model contains not only the estimate A of the underlying structure but also a formal description of the noise (e.g., its standard deviation σ) and parameters ξ describing the statistics of the transformations ϕ_i. We assume here a model which describes the data images X_i as

$$X_i = A(-\phi_i) + \sigma R_i \tag{6}$$

This model has the set of parameters $\Theta = (A, \sigma, \xi)$. Because this model has the same form as that which gives rise to the data (Eq. (1)), an estimated structure A should approach the true structure zW. The parameter σ is the standard deviation of each pixel of the added Gaussian noise, taking the R_i, $i = 1 \ldots N$ to be independent, white Gaussian noise images of unit variance.

The individual transformation variables ϕ_i; $i = 1 \ldots N$ are taken to be independent triplets of random variables whose probability distribution is parameterized by ξ. The data set is imagined to arise from a selection process in which a large micrograph is scanned to detect particles, and small images X_i are extracted with one particle roughly centered in each image. We assume that this approximate centering of the particles leaves residual random transformations ϕ_i which are Gaussian-distributed in x and y and uniformly distributed in the rotation angle α. Letting $\phi = (q_\alpha, q_x, q_y)$ and $\xi = (\xi_\sigma, \xi_x, \xi_y)$, the probability density of the transformation variables ϕ is then given by

$$f(\phi|\Theta)d\phi = \frac{1}{2\pi\xi_\sigma^2} \exp\left[-\frac{(q_x - \xi_x)^2 + (q_y - \xi_y)^2}{2\xi_\sigma^2}\right] \frac{dq_\alpha}{2\pi}\, dq_x dq_y. \tag{7}$$

The function to be optimized is the likelihood, which is the probability that a particular data set **X** arises from the model with parameter set Θ,

$$\mathrm{Lik}(\Theta) = P(\mathbf{X}|\Theta). \tag{8}$$

The likelihood is treated as a function of Θ with the data **X** treated as constant. The parameter set $\hat{\Theta}$ that maximizes the likelihood is an asymptotically efficient estimate of the true values of these parameters (Rice, 1995; Cramér, 1946). This means that, in the limit of very large data sets, no other way of estimating the parameters gives smaller errors than maximum-likelihood estimation.

Maximizing the likelihood is equivalent to maximizing its logarithm L. Because we assume the independence of the random variables associated with the individual images, L will be the sum of the log likelihoods of the individual images,

$$L(\Theta) = \sum_{i=1}^{N} \ln P(X_i|\Theta). \tag{9}$$

In order to evaluate the probability $P(X_i|\Theta)$ of observing the data image X_i given the model parameters Θ, we expand it as an integral over all possible values of the transformation ϕ as

$$L(\Theta) = \sum_{i=1}^{N} \ln \int \gamma_i(\phi; \Theta)\, d\phi, \tag{10}$$

where

$$\gamma_i(\phi; \Theta) = P(X_i|\phi, \Theta) f(\phi|\Theta). \qquad (11)$$

Here $f(\phi|\Theta)$ is the probability density given in Eq. (7). The other factor is the probability of obtaining an individual image X_i given ϕ and Θ. Taking the value of each pixel to be a continuous variable, we express this probability as a probability density. With the assumption of white Gaussian noise in each data image, the probability density is a product of Gaussians representing the noise distribution in each of the M pixels:

$$\begin{aligned} P(X_i|\phi, \Theta) &= \prod_{j=1}^{M} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_{ij}(\phi) - a_j)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^M \exp\left[-\frac{|X_i(\phi) - A|^2}{2\sigma^2}\right] \end{aligned} \qquad (12)$$

Equations (7) and (10–12) therefore allow the log likelihood to be evaluated for a given value of the parameters of the model.

Refinement of the Image

A local maximum of the likelihood is attained by finding those parameter values at which the partial derivatives of the log likelihood are zero. For example, optimum values of the individual pixels a_j of the structure estimate A are found by setting

$$\frac{\partial L}{\partial a_j} = 0, j = 1 \ldots M. \qquad (13)$$

These partial derivatives are given by

$$\frac{\partial L}{\partial a_j} = \sum_{i=1}^{N} \frac{\int \left(\frac{x_{ij}(\phi) - a_j}{\sigma^2}\right) \gamma_i(\phi; \Theta)\, d\phi}{\int \gamma_i(\phi; \Theta)\, d\phi}. \qquad (14)$$

An iterative numerical procedure is required to solve Eq. (13) along with the corresponding equations for the other model parameters. Because the system of equations is very large, a simple steepest-ascent algorithm is used here; it is equivalent to the Baum-Welsh algorithm for the optimization of hidden Markov model parameters (Levinson *et al.*, 1983), which increases the likelihood at each step. Given an estimate $\Theta^{(n)}$ of the parameters (which include the pixel values $a_j^{(n)}$) from the nth iteration, an improved set of pixel values $a_j^{(n+1)}$ is obtained by solving

$$\sum_{i=1}^{N} \frac{\int \left(\frac{x_{ij}(\phi) - a_j^{(n+1)}}{\sigma^2}\right) \gamma_i(\phi; \Theta^{(n)})\, d\phi}{\int \gamma_i(\phi; \Theta^{(n)})\, d\phi} = 0,\ j = 1 \ldots M. \qquad (15)$$

The solution is

$$A^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\int X_i(\phi)\gamma_i(\phi; \Theta^{(n)})\, d\phi}{\int \gamma_i(\phi; \Theta)^{(n)})\, d\phi}. \qquad (16)$$

The new estimate of the structure $A^{(n+1)}$ is therefore a weighted average of the transformed data images, the weights being the probability function γ_i evaluated with the old parameters $\Theta^{(n)}$.

Reestimating Other Model Parameters

The other model parameters can be optimized concurrently with the pixel values. The noise standard deviation σ can be reestimated by approximating $\partial L/\delta\sigma = 0$, yielding the formula

$$\sigma^{(n+1)} = \left(\frac{1}{NM} \sum_{i=1}^{N} \frac{\int |X_i(\phi) - A^{(n)}|^2 \gamma_i(\phi; \Theta^{(n)})\, d\phi}{\int \gamma_i(\phi; \Theta^{(n)})\, d\phi}\right)^{1/2}. \qquad (17)$$

It is possible also to reestimate parameters of the probability density $f(\phi|\Theta)$. Taking the functional form given in Eq. (7) and approximating the solutions to $\partial L/\partial\xi_x = 0$ and $\partial L/\partial\xi_y = 0$, we obtain reestimated values of ξ_x and ξ_y. The reestimation formula for ξ_x is

$$\xi_x^{(n+1)} = \frac{1}{N} \sum_{i=1}^{N} \frac{\int q_x \gamma_i(\phi; \Theta^{(n)})\, d\phi}{\int \gamma_i(\phi; \Theta^{(n)})\, d\phi}, \qquad (18)$$

where q_x is the x-displacement component of the transformation ϕ. Similarly a reestimation formula for the standard deviation of the translations is given by

$$\xi_\sigma^{(n+1)} = \left(\frac{1}{2N} \sum_{i=1}^{N} \frac{\int (q_x^2 + q_y^2)\gamma_i(\phi; \Theta^{(n)})\, d\phi}{\int \gamma_i(\phi; \Theta^{(n)})\, d\phi}\right)^{1/2}. \qquad (19)$$

The maximum-likelihood refinement algorithm makes use of Eqs. (16)–(19). Given an estimate $\Theta^{(n)}$ of the model parameters, these equations provide a set of parameters $\Theta^{(n+1)}$ which increases the likelihood of the model. The set includes not only the improved reconstructed image $A^{(n+1)}$ but also improved values of the other model parameters, whose optimization is necessary to ensure the fidelity of the estimated structure. By repeated iterations of this sort, a local maximum of the likelihood is approached.

A Note about Implementation

Reestimation of the parameters can be made more computationally efficient by noticing that the function $\gamma_i(\phi; \Theta)$ is related to the cross-correlation of the data and the reference images. The exponential function in Eq. (12) can be rewritten as

$$\exp\left(-\frac{|X_i(\phi) - A|^2}{2\sigma^2}\right) = \exp\left(-\frac{|X_i(\phi)|^2 + |A|^2}{2\sigma^2}\right)\exp\left(\frac{X_i(\phi)\cdot A}{\sigma^2}\right),$$

which is seen to be the product of two terms. The first term is independent of ϕ since the coordinate transformation ϕ is expected to leave the image power unchanged, i.e.,

$$|X_i(\phi)|^2 = |X_i|^2$$

for all ϕ. In this case only the second term needs to be evaluated as a function of ϕ. It is seen to be an exponential function of the cross-correlation of the image X_i with the reference A. Thus the function γ_i can be evaluated simply as

$$\gamma_i(\phi; \Theta) = k_i \exp\left(\frac{X_i(\phi)\cdot A}{\sigma^2}\right) f(\phi|\Theta), \qquad (20)$$

where k_i is a constant that is independent of ϕ.

RELATIONSHIP OF THE MAXIMUM-LIKELIHOOD AND ALIGNMENT ALGORITHMS

The maximum-likelihood (ML) refinement algorithm becomes equivalent to the cross-correlation alignment algorithm (CCA; Eqs. (3) and (5)) at high signal-to-noise ratios. In the ML algorithm, the reestimated image is computed as a weighted integral over all possible transformations of the data images (Eq. (16)); the weighting function is $\gamma_i(\phi; \Theta)$. It can be seen from Eq. (20) that when σ is small (the case of high signal-to-noise ratio) γ_i as a function of ϕ is strongly peaked near its maximum value. The maximum occurs at

$$\phi_i = \text{argmax}\,[X_i(\phi)\cdot A + \sigma^2 \ln f(\phi; \Theta)]. \qquad (21)$$

Ignoring for the moment the second term involving $f(\phi; \Theta)$, this assignment of ϕ_i is seen to be the same as that given by CCA Eq. (3). At a high signal-to-noise ratio γ_i approaches the form of a delta function in ϕ, and Eq. (16) becomes equivalent to Eq. (5).

The second term in Eq. (21) suggests a way in which a priori information about the distribution of particles can be incorporated into the cross-correlation alignment algorithm. Given prior information about the positions or orientations of the particles, formally expressed as model parameters for the pdf of the random transformations $f(\phi; \Theta)$, the cross-correlation can be corrected by this additional term before the maximum is found.

As we shall see, the advantage of the maximum-likelihood algorithm appears at low signal-to-noise ratios. In these cases the transformations ϕ_i obtained by maximizing the cross-correlation do not reliably reflect the actual transformations φ_i; instead, there may be many false peaks in the cross-correlation function due to noise. The maximum-likelihood algorithm allows for the presence of these ambiguities in the cross-correlation function.

SIMULATIONS

Methods

To compare the performance of the cross-correlation alignment and maximum-likelihood algorithms, several large synthetic data sets were constructed and analyzed. Shown here are the results from a data set like those illustrated in Fig. 1. The synthetic images were derived from the projected structure of a bacteriorhodopsin trimer. The scaling of this "true structure" W and the noise was chosen so that the resulting "data images," formed as in Eq. (1), had signal to noise ratios equal to z^2. In these images the random transformations φ were uniformly distributed in angle α, quantized in 7.5° steps; the translations in x and y were Gaussian distributed with a standard deviation of 10 pixels, quantized in 1-pixel steps. Periodic boundary conditions were employed in both the generation and analysis of the images, as can be seen in some of the images in Fig. 1. All image rotations were carried out using cubic interpolation of pixel values.

Refinement of the estimated structure was carried out by iteratively evaluating Eqs. (3) and (5) for CCA or Eqs. (16)–(19) for the ML algorithm. In some cases a modified cross-correlation alignment (MCCA) was performed using Eq. (21) instead of Eq. (3) to bias the search in view of the known distribution of transformations. It should be noted that the "reference-free" variant of the cross-correlation alignment (Eq. (4)) was not used here because of the large increase in computer resources required; because of the very large N, the effect of the modification is expected to be very small.

The search for the maximum of the cross-correlation in Eqs. (3) or (21) and the evaluation of the integrals in Eqs. (16)–(19) covered the entire ϕ space. As in the data synthesis, α was quantized in 16 steps of 7.5° to cover the range of angles from 0 to 120°. At each value of α, translations were searched using two-dimensional FFTs; in all, a total of $64 \times 64 \times 16$ possible transformations were searched for each data image X_i in each iteration. At the end of each iteration, threefold symmetry was imposed by threefold rotational averaging of A. As implemented in the MATLAB environment, each iteration of either the CCA or ML algorithm required about 30 min of CPU time on a Power Macintosh G3 computer for the $N = 4000$ images. In the analyses of the data set, each algorithm was carried out through many iterations until the refined structures A became stationary.

Comparison of CCA and ML Refinement

At sufficiently high signal-to-noise ratios, the maximum-likelihood and cross-correlation alignment algorithms give similar results. Figure 2 compares the behavior of the ML and MCCA algorithms for a data set with $z = 0.14$. Although this corresponds to a signal-to-noise ratio (SNR) of only 0.02 for each pixel, the fact that the cross-correlation is evaluated over a large number of pixels makes the alignment fairly reliable. At this SNR the best-case cross-correlation, which is the cross-correlation between W and a data image X_i, results in the exact identification of the correct transformation φ_i in 64% of the cases; success rates above 80% result if slight misalignments are allowed. This SNR is near the minimum for successful MCCA refinement in the problem shown, where the relatively poor starting reference $A^{(0)}$ is itself a data image created with $z = 0.14$. (Other simulations showed that if the standard CCA algorithm is used, which does not bias the search according to the distribution of transformations, a twofold larger SNR of 0.04 is required for convergence to the correct structure.) Meanwhile the ML refinement is seen to yield a very similar result using a similar number of iterations.

It is well known that the cross-correlation alignment is sensitive to the starting estimate and tends to retain the starting estimate's features when the data have a low signal-to-noise ratio. Figure 3 shows the behavior of CCA refinement based on a data set with SNR = 0.005, one-fourth that used Fig. 2. In the analysis two different starting estimates were used; one was a data image with $z = 0.2$. The other starting estimate was a motif (the "radiation symbol") having the same symmetry and approximate area as the molecular boundary. As might be expected, convergence of the CCA algorithm resulted in refined structures that mainly resemble the starting estimates and bear little resemblance to the true structure.

It is not surprising that the CCA algorithm per-

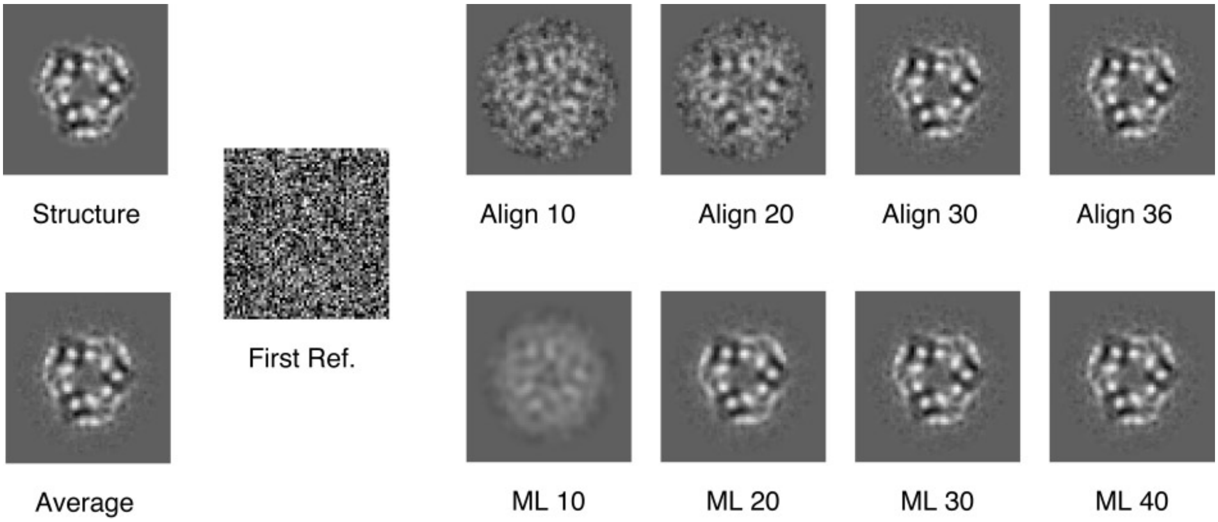

FIG. 2. Cross-correlation and maximum-likelihood refinement based on a data set of $N = 1000$ images with $z = 0.14$. The starting reference $A^{(0)}$ was in each case the single data image shown. Intermediate reconstructions $A^{(n)}$ are shown for $n = 10, 20, \ldots$ for MCCA (top row) and ML algorithms. The MCCA algorithm made use of Eqs. (21) and (5); the probability density $f(\varphi; \Theta)$ was evaluated with the parameters ξ equal to those used in the simulation. MCCA iterations after $n = 36$ produced no change. The final reconstructions are nearly indistinguishable from the true average image (lower left) which was formed using the known transformations φ_i.

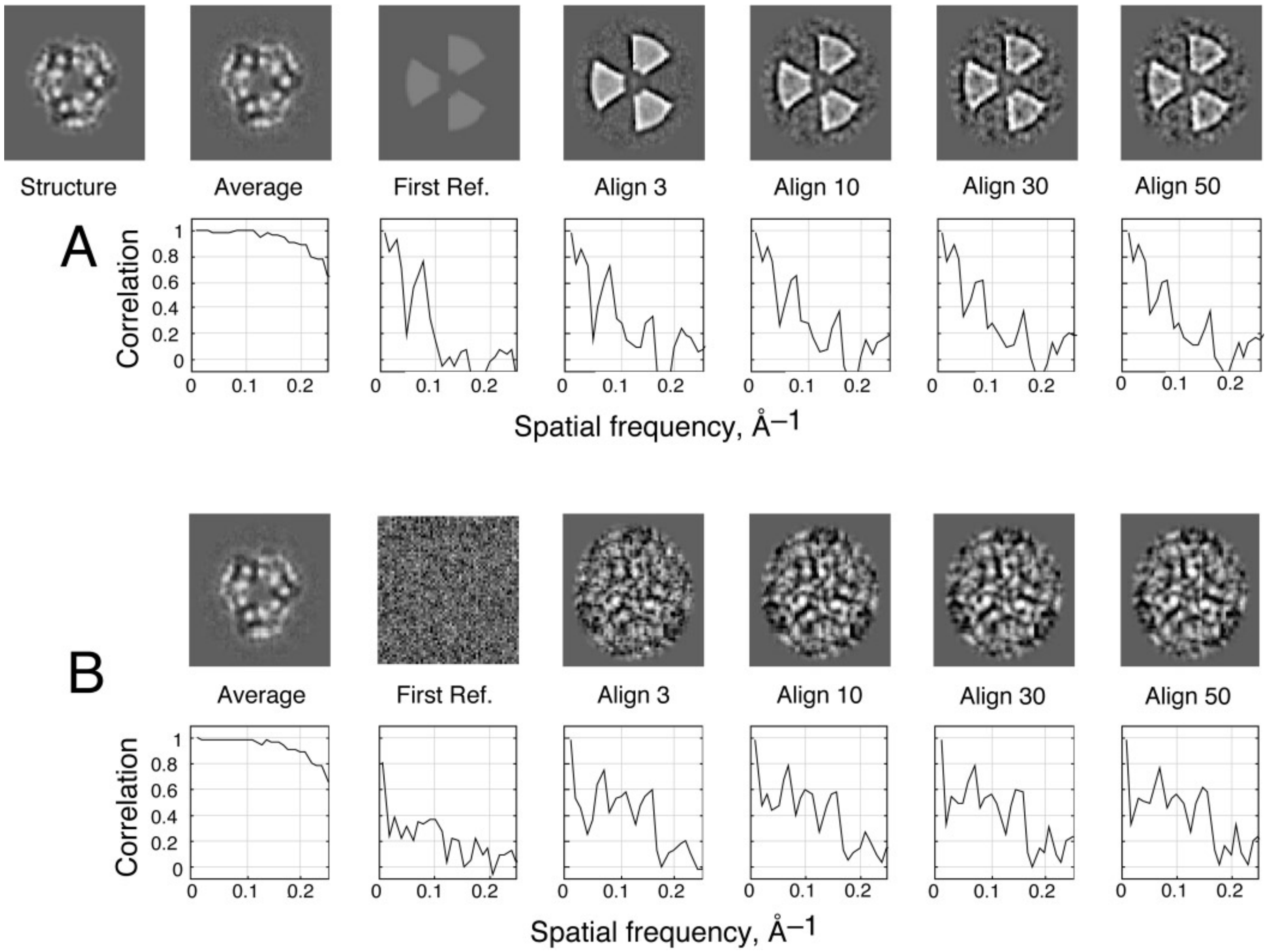

FIG. 3. Cross-correlation alignment of a simulated data set of 4000 images with signal amplitude $z = 0.07$. Shown are images and the Fourier ring correlation (FRC) for the correctly aligned average (obtained using the original φ_i values), the first reference image $A^{(0)}$, and the average aligned image after various numbers of iterations, using as the initial reference in (A) the "radiation symbol" or in (B) a scaled ($z = 0.2$) copy of the true structure with added noise. Little change was seen in the aligned averages after 20 iterations, and the final alignments bear little resemblance to the true structure. The FRC was computed between the true structure and each image shown. It was obtained from the Fourier transforms F_A and F_W of A and W, respectively, according to (van Heel, 1987)

$$C = \frac{\sum F_A F_W^*}{\sqrt{\sum |F_A|^2 \sum |F_W|^2}},$$

where the sums are taken over an annulus of the Fourier plane corresponding to the given spatial frequency magnitude; values of C are plotted as a function of spatial frequency in the graphs. The FRC shows that the true average agrees well with the original structure to a spatial frequency of 0.25 Å^{-1}; however, the agreement of the aligned images is very poor.

formed poorly at this SNR, because the best-case cross-correlation yields the correct alignment in only 4% of the cases. The probability of obtaining the correct alignment of data images with the starting estimates that were used is even lower.

ML analysis was performed on this data set and with the same two starting structures. Other model parameters were set to incorrect initial values (e.g., $\xi_\sigma = 5$ instead of 10) to check for the robustness of the reestimation. Convergence occurred quite slowly, but after about 200 iterations the structures refined from the two starting estimates are essentially identical (Fig. 4), and the other model parameters also converged to the correct values. The Fourier ring

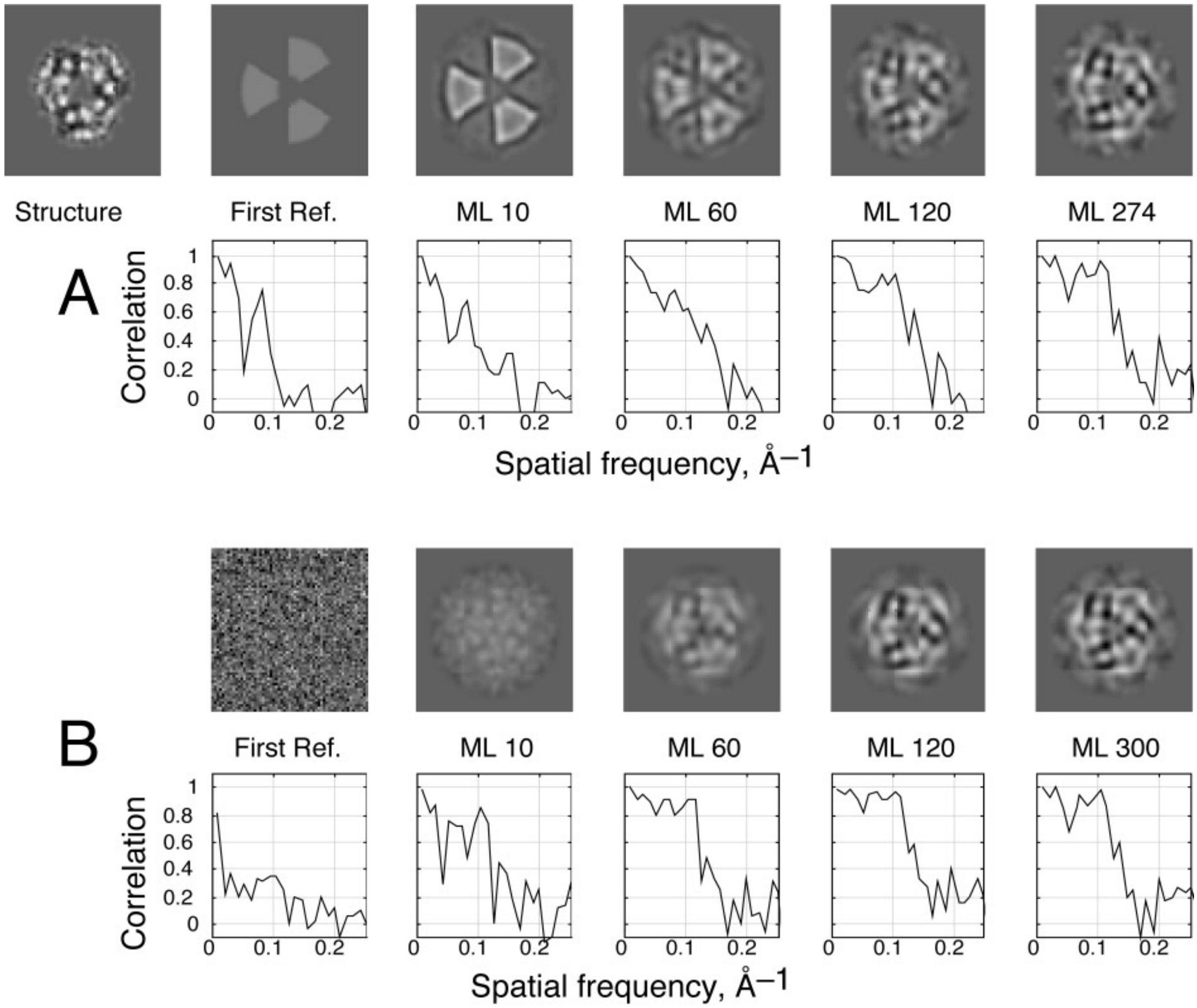

FIG. 4. Maximum-likelihood refinement using the simulated data set. Using the same initial references and data set as described in the legend to Fig. 3, ML refinement was carried out to 274 and 300 iterations in the two examples shown. Gradual convergence to an image very similar to the true structure is seen from both (A) the "radiation symbol" or (B) the noisy structure as the starting reference. The Fourier ring correlations show convergence in each case to a refined image, having good agreement with the original structure to about 0.12 Å^{-1} resolution.

correlation was computed between W and the refined structures. This function gives a measure of the similarity of images as a function of spatial frequency (van Heel, 1987), and shows good reconstruction of the original to resolutions beyond 0.1 Å^{-1}.

The examples in Fig. 4 use a data set containing $N = 4000$ images having a structure amplitude of $z = 0.07$. For a data set of this size this SNR appears to represent the lower limit of useful information for the ML algorithm. In other simulations it was seen that convergence from the two starting structures did not occur from 4000 images having the lower value $z = 0.05$. The size of the data set makes a difference, however; convergence was seen for $N = 8000$ and $z = 0.05$ after about 500 iterations. Thus having a large data set is important in cases where the SNR is very small; also in these cases the convergence of the steepest-ascent algorithm becomes quite slow.

Analysis of Pure-Noise Images

It is interesting to compare the behavior of the CCA and ML algorithms when presented with a data set consisting only of noise. As has been pointed out by van Heel *et al.* (1992), cross-correlation alignment tends to "align" the noise in a data set to reproduce the starting reference. This is demonstrated in Fig. 5A, where the CCA algorithm reproduces the starting "radiation symbol" motif; with iteration the motif remains, while the rms pixel amplitude actually increases to a maximum (dotted curve in Fig. 5C). In the ML algorithm on the other hand, iteration

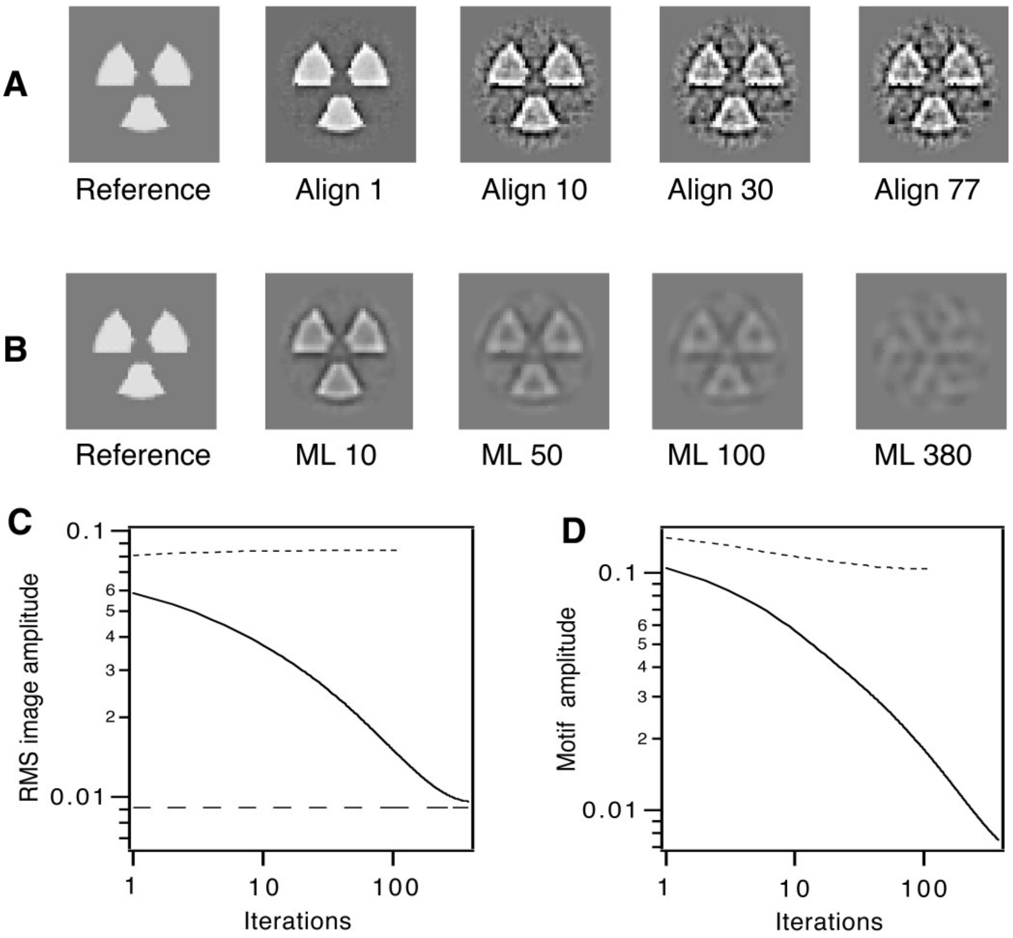

FIG. 5. Analysis of a data set containing no underlying structure. A set of $N = 4000$ images containing only Gaussian noise (independent pixels, unity standard deviation) was used to refine the "radiation symbol" initial structure, with threefold symmetry enforced. (A) The original reference and the result after the given numbers of iterations of cross-correlation refinement are shown. Iterations after the 77th produced no change; the final structure is similar to the initial reference. (B) The ML refinement shows a gradual decay of the original motif into noise. (C) The progress of refinement is shown by plotting the rms pixel amplitude (averaged over the central disc of the floated image) as a function of iteration number for the CCA algorithm (dotted curve) and MA algorithm (solid curve). The dashed line represents the expected noise level in an average of the N images (with threefold averaging). (D) The normalized motif amplitude $m = A^{(n)} \cdot A^{(0)}/|A^{(0)}|$, where $A^{(0)}$ is the reference image, is plotted as a function of iteration number n. CCA and ML results are represented by dotted and solid curves, respectively.

results in a monotonic decrease in the amplitude of the false motif. The rms amplitude of the refined structure decays to approach the value expected for the average noise in the 4000 images (Fig. 5C), while the correlation with the original motif continues to decay even after 380 iterations (Fig. 5D).

The contrasting behaviors seen in Fig. 5 reflect the fact that the two algorithms optimize different quantitities. The CCA algorithm maximizes the squared magnitude of the reconstruction (Eq. (2)), leading to the undesirable result of increasing rms amplitudes in the refined image. The ML algorithm optimizes the likelihood; asymptotically this should yield unbiased estimates of the various model parameters, and in this example it yields very low pixel amplitudes in its estimate of the underlying structure.

Asymptotic Behavior of ML Refinement

When alignment can be carried out reliably, the refined image A obtained from either the CCA or ML algorithms is simply an average of the aligned images. In this case the statistics of the refined image are readily estimated. For example, given a fixed number of images and fixed noise variance, the mean-squared error in the refined structure is inversely proportional to the SNR of the original images. But what happens to the error in the ML-refined structure when the signal amplitude decreases to a level where reliable alignment is no longer possible? Presumably the increased uncertainty in alignment results in increased noise or bias in the refined image. This can be seen to be the case in the ML reconstructions in Fig. 4. Although they resemble the true structure, their Fourier ring correlations show a lower useful resolution than that obtained from the "true average," obtained by averaging the 4000 data images after correct alignment (see the "Average" panels in Fig. 3). The lower quality of the ML reconstructions would be approximately matched if only 200 correctly aligned data images were to be averaged.

To test the asymptotic behavior of the ML alignment, a very simple test structure was used to synthesize up to 200 000 data images, each 8 × 8 pixels in size, at various signal-to-noise ratios. The underlying structure W was a pattern of five ones in a background of 59 zeros (Fig. 6A). The probability of correct alignment of these data images, based on cross-correlation with W, is seen to drop steeply for SNR < 5 in this case (Fig. 6B). Above this threshold, the SNR of the refined image is proportional to that of the data images as is expected from the statistics of averaging. Below the threshold, however, the SNR of the refined image decreases much more steeply than linearly with decreases in the data SNR (Fig. 6C), roughly as the third power.

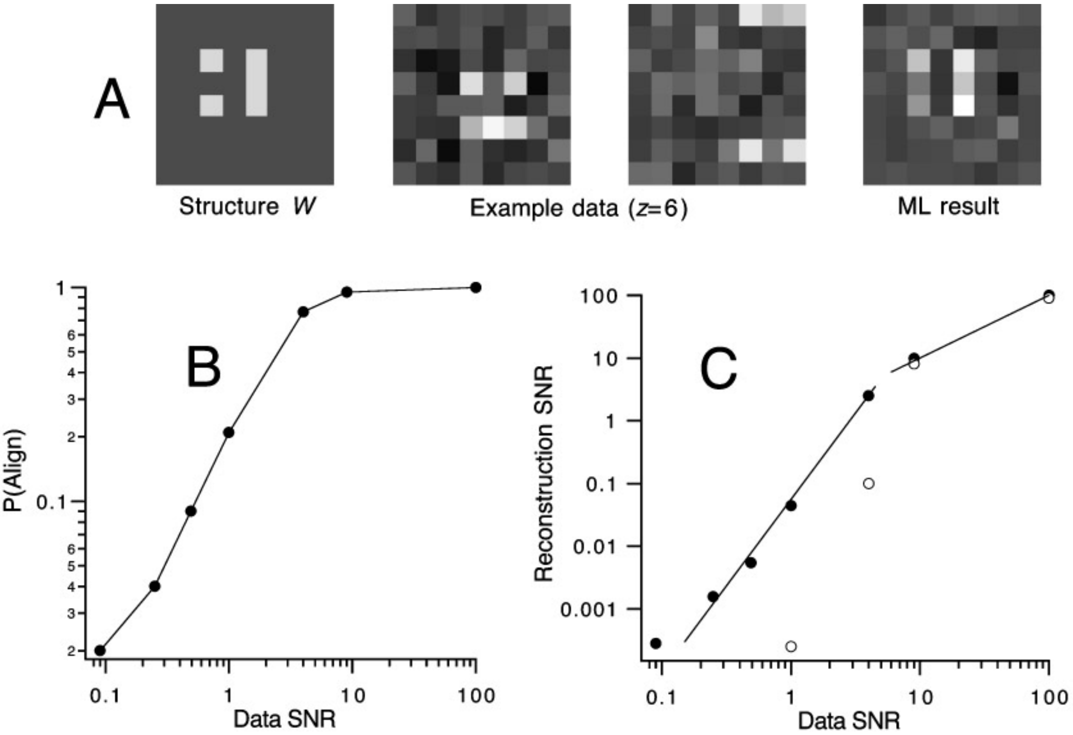

FIG. 6. ML refinement of a simple structure, consisting of 5 pixels having value 1 and 59 pixels with value 0. (A) The true structure W and two examples of simulated images obtained as in Eq. (1) with amplitude $z = 6$. Also shown is the result from ML refinement (700 iterations) from a data set of 200 000 images having $z = 0.3$. In the data simulation, all x and y translations have equal probability, rotations are quantized to multiples of 90°, and periodic boundary conditions are used. (B) Probability of correct alignment of the 8 × 8 pixel object is shown as a function of the data signal-to-noise ratio z^2. (C) Quality of reconstructed image as a function of the signal-to-noise ratio of the data images (filled circles). The reconstruction signal-to-noise ratio s was computed from the refined image A according to

$$s = \frac{|A - zW|^2}{NM},$$

where N is the number of data images and $M = 64$ is the number of pixels. In each simulation N was chosen large enough to ensure convergence of the refinement. Open circles are from corresponding CCA analyses of the same data sets. Lines represent a linear relationship between input and output signal-to-noise at high data SNR (upper right) and a third-power relationship at low SNR (lower left).

DISCUSSION

Presented here is a maximum-likelihood approach to the refinement of a two-dimensional structure from a data set of noisy images of randomly positioned and oriented particles. The maximum-likelihood algorithm appears to be much less sensitive to the choice of the starting estimate and functions well at low signal-to-noise ratios where cross-correlation-based alignment is unreliable. As implemented, the ML algorithm requires a summation over all possible translations and rotations of each data image. This contrasts with the more limited (and faster) alignment search that is typically performed as sequential translational and rotational alignments are carried out (Frank, 1996). However, in the simulations shown here it is seen that for the two-dimensional refinement problem a large data set can be processed at tolerable speed by a single-processor computer.

The traditional cross-correlation alignment algorithm has two properties that limit its usefulness for data having very low signal-to-noise ratios. First, the quantity being maximized by this algorithm (the power in the refined image; Eq. (2)) is not necessarily the correct one. In the simulation shown in Figs. 5A and 5C, the CCA iterations actually decrease the quality of the refinement. Second, the CCA algorithm involves a discontinuous mapping between a structure estimate $A^{(n)}$ and its refinement $A^{(n+1)}$. For each data image a particular alignment is chosen as the best one, on the basis of maximizing the cross-correlation with $A^{(n)}$. It can occur that a small change in $A^{(n)}$ makes no difference in the alignment parameters, and therefore leaves $A^{(n+1)}$ unchanged. One result of this discontinuous mapping is the fact that the algorithm typically halts after a finite number of iterations.

The maximum-likelihood algorithm performs better than CCA at low signal-to-noise ratios because it optimizes a better measure of refinement (the likelihood) and because the optimization is done through a refinement step that involves a continuous mapping. In the limit of infinitely large data sets, maximization of the likelihood yields unbiased estimates of parameters (Rice, 1995; Cramér, 1946); in the examples shown here the bias in the estimated structures is indeed seen to be much lower than that from the CCA algorithm. It should, however, be kept in mind that the ML algorithm is still susceptible to convergence to false local maxima in the likelihood function.

An important advantage of the maximum-likelihood approach is that it can make use of prior information about the probability density function $f(\phi)$ of transformations. In the case of partial disorder in a two-dimensional crystal, for example, $f(\phi)$ would reflect the distribution of particle positions about the corresponding lattice points. Maximum-likelihood refinement could then be performed to increase the resolution of the refined structure beyond the quality that is otherwise obtained. As shown in Eq. (21), this sort of prior information can be incorporated into the conventional cross-correlation alignment as well.

The theory presented here can be extended in various ways. For computational ease the image noise has been assumed here to be Gaussian, but the calculation of the probability (Eq. (12)) can be changed to reflect the Poisson statistics of electron counting. The theory can also be extended to incorporate multiple references, so that multiple particle types (or views) can be refined simultaneously. The theory can in principle be extended to the refinement of 3D structures from projections, although in this case the exhaustive search of 3D transformations may be computationally prohibitive.

I thank Dr. R. MacKinnon for his generous hospitality during my sabbatical stay at Rockefeller University, where much of this work was done. I also thank Dr. S. Darst for introducing me to cryo-EM techniques, and Drs. P. Penczek and D. DeRosier for helpful discussions. This work was supported by NIH Grants NS-21501 and NS-35282. The Matlab programs implementing the algorithms described in the paper are available from the author on request.

REFERENCES

Chung, S. H., Krishnamurthy, V., and Moore, J. B. (1991) Adaptive processing techniques based on hidden Markov models for characterizing very small channel current buried in noise and deterministic interferences, *Phil. Trans. R. Soc. Lond. B* **334,** 357–384.

Cramér, H. (1946) Mathematical Methods of Statistics, Princeton Univ. Press, Princeton.

Dong, W., Baird, T., Fryer, J. R., Gilmore, C. J., MacNicol, D. D., Bricogne, G., Smith, D. J., O'Keefe, M. A., and Hövmoller, S. (1992) Electron microscopy at 1-Å resolution by entropy maximization and likelihood ranking, *Nature* **355,** 605–609.

Doublie, S., Xiang, S., Gilmore, C. J., Bricogne, G., and Carter, C., Jr. (1994) Overcoming non-isomorphism by phase permutation and likelihood scoring: Solution of the TrpRS crystal structure, *Acta Crystallogr. A* **50,** 164–182.

Frank, J. (1996) Three-Dimensional Electron Microscopy of Macromolecular Assemblies, Academic Press, San Diego.

Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M., and Henderson, R. (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin, *J. Mol. Biol.* **259,** 393–421.

Henderson, R. (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules, *Q. Rev. Biophys.* **28,** 171–193.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994) Hidden Markov models in computational biology: Applications to protein modeling, *J. Mol. Biol.* **235,** 1501–1531.

Levinson, S. E., Rabiner, L. R., and Sondhi, M. M. (1983) An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition, *Bell System Tech. J.* **62,** 1035–1074.

Penczek, P., Radermacher, M., and Frank, J. (1992) Three-dimensional reconstruction of single particles embedded in ice, *Ultramicroscopy* **40,** 33–53.

Rice, J. A. (1995) Mathematical Statistics and Data Analysis, Duxbury Press, Belmont, CA.

Saxton, W. O., and Frank, J. (1977) Motif detection in quantum noise-limited electron micrographs by cross-correlation, *Ultramicroscopy* **2,** 219–227.

van Heel, M. (1987) Similarity measures between images, *Ultramicroscopy* **21,** 95–100.

van Heel, M., Schatz, M., and Orlova, E. (1992) Correlation functions revisited, *Ultramicroscopy* **46,** 307–316.

Venkataramanan, L., Kuc, R., and Sigworth, F. J. (1998) Identification of hidden Markov models for ion channel currents. II. State-dependent excess noise., *IEEE Trans. Sig. Proc.* **46,** 1916–1929.