7

Algorithms for Three-dimensional Reconstruction From the Imperfect Projection Data Provided by Electron Microscopy

Jose-Maria Carazo, Gabor T. Herman, Carlos O. S. Sorzano and Roberto Marabini

Jose-Maria Carazo • Centro Nacional de Biotecnología (CSIC), Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain

Gabor T. Herman • Department of Computer Science, The Graduate Center, City University of New York, New York, NY 10016, USA

Carlos O. S. Sorzano • Escuela Politécnica Superior, Universidad San Pablo-CEU, 28668 Boadilla del Monte, Madrid, Spain

Roberto Marabini • Escuela Politécnica Superior, Universidad Autónoma, 28049 Cantoblanco, Madrid, Spain

1. INTRODUCTION

Since the 1970s, it has become increasingly evident that transmission electron microscopy (TEM) images of typical thin biological specimens carry a large amount of information on 3D macromolecular structure. It has been shown many times how the information contained in a set of TEM images (2D signals) can determine a useful estimate of the 3D structure under study.

In its most general form, the 3D reconstruction problem in TEM can be defined by the statement: given a collection of projection images (2D data) g, determine the 3D structure f that produced the images g. This problem has to be solved under conditions in which the image data, as well as the information about the geometry of data collection that relates g to f, are imperfect; in particular, both the gray level information in the images and the information regarding the direction of projections are corrupted by substantial noise. We are interested in knowing under which conditions g is adequate for producing an $\hat{f}$ that is close to f in some sense. For experimental reasons, we always have three basic limitations on the collection of the image data set g: the image gray level noise, the imperfect information about the data collection geometry and, finally, the restriction to a finite number of images. Can we do something to ameliorate the situation? Can we qualify our confidence in the reproducibility of the reconstructed structures (in other words, do small changes in g produce radical changes in the estimate $\hat{f}$?). Formulated in this way, it becomes clear that the topics covered under 3D reconstruction in TEM belong to the broad class of signal-recovery/inverse problems. It is within this framework that we discuss how the physical limits directly affect the fidelity of the estimated 3D structures produced by various reconstruction algorithms.

2. OVERVIEW OF OUR APPROACH

In this section, we present the concepts that we consider fundamental to a careful discussion of our topic. We illustrate these concepts as we go along on the relatively simple problem of recovering a 2D structure from its 1D images.

In the area of inverse problems, it is assumed that we have some understanding of the data collection process. Usually we describe it by some mathematical idealization, combined with a description of how the actual data collection process differs from the ideal one. For example, in the field of (2D) image reconstruction from (1D) projections, the ideal data consist of the line integrals, for all lines, of the image f to be reconstructed. In practice, we can have data corresponding to only finitely many lines, and the data are likely to be contaminated with noise (see Fig. 1). To approximate f from such data, there are two basic approaches that we call transform methods and series expansion methods, respectively.

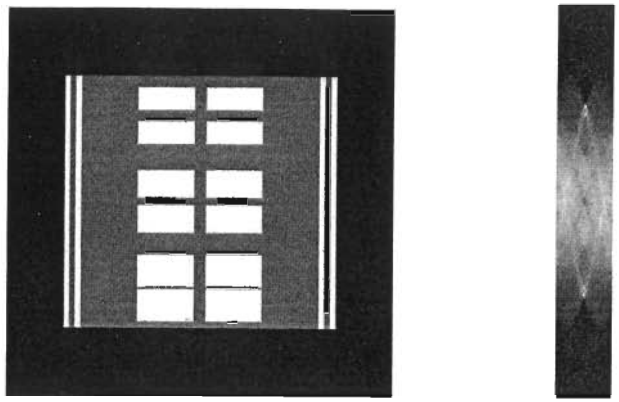


FIGURE 1. A 341 × 341 pixel representation of a 2D image (left) and its 1D projections (right). Each projection consists of noisy versions of line integrals for 495 parallel lines (the distance between them is the same as the length of an edge of a pixel in the image); they correspond to individual columns in the picture on the right. The angle between the lines corresponding to one column and the next one is 2°; there are 70 columns corresponding to direction angles ranging from −69° to 69° with the vertical. The noise is additive zero-mean Gaussian noise, with a standard deviation that is the integral for a vertical line through the largest of the 12 central rectangular shapes in the image. (The image is approximately a cross-section through a 3D phantom, designed by Fernández *et al.* (2002), of mitochondria; the just discussed rectangular shapes are sections through the simulated cristae.)

The essence of transform methods is to find a mathematical procedure that describes the recovery of f from its ideal data, and then implement this procedure (as best as one can) making use of the actual data. In the case of 2D image reconstruction from 1D projections, the ideal data consist of all $p(l,\theta)$ that are the line integrals of f along lines at a distance l from the origin, making an angle θ in radians with the vertical. In that case, we have the formula (Herman (1980), p. 95)

$$f(r,\phi)=\frac{1}{2\pi^2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\int_{-E}^{E}\frac{1}{r\cos(\theta-\phi)-l}p_1(l,\theta)\mathrm{d}l\mathrm{d}\theta \qquad (1)$$

for f in polar coordinates with ϕ in radians, where $p_1(l, \theta)$ denotes the partial derivative of $p(l, \theta)$ with respect to l, and E is a number large enough so that $p(l, \theta) = 0$, for all $l > E$ and all θ. This formula is derived, based on the pioneering work of Radon (1917), in Section 16.3 of Herman (1980). The inner integral in equation (1) is referred to as a convolution, while the outer integral is referred to as a back-projection.

For the data collection method described in Fig. 1, we have only (noisy versions of) the data $p(nd, \Delta\theta)$, where d is the length of an edge of a pixel

in the image on the left of Fig. 1, the n are integers in the range from –247 to 247, the m are integers in the range from –34 to 35, and Δ_m is $(2m - 1)°$ in radians. We now discuss one way of finding an approximation $\hat{f}$ to the f in equation (1) based on such data. What we describe is referred to as the convolution method by Herman (1980); it is more popularly known as the filtered back-projection method (FBP), and it is very similar in nature to the weighted back-projection method (WBP) for the reconstruction of 3D structures from 2D projection images that is discussed below in Section 4.1.

To evaluate the convolution, we do the following. We define, for $-247 \le \bar{n} \le 247$ and $-34 \le m \le 35$,

$$p_c(\bar{n}d, \Delta_m) = d \sum_{n=-247}^{247} p(nd, \Delta_m) q((\bar{n} - n)d), \tag{2}$$

(see p. 128 of Herman, 1980), where q is the so called filter function. There are a number of ways to define q; they will lead to different reconstructions $\hat{f}$. In the illustration given here, we use

$$q(u) = \frac{2A}{\pi} \int_0^{\frac{A}{2}} \sin\frac{\pi U}{A} \cos(2\pi U u) dU, \tag{3}$$

see Herman (1980), p. 126, with the *Sinc* window from Table 8.1 chosen for F_A. Note that q depends on the parameter A. Choosing A small has a smoothing effect on the reconstruction; the ideal choice of A depends on the relative nature of the structure we are attempting to reconstruct (the signal) and of the data collection (including the noise).

The back-projection, for the data collection geometry described in Fig. 1, is approximated by

$$\begin{aligned}\hat{f}(r,\phi) = 11\Delta[p_c(r\cos(\Delta_{-34} - \phi), \Delta_{-34}) + p_c(r\cos(\Delta_{35} - \phi), \Delta_{35})] \\ + \Delta \sum_{m=-33}^{34} p_c(r\cos(\Delta_m - \phi), \Delta_m)\end{aligned} \tag{4}$$

where Δ is 2° in radians, and values of $p_c(r \cos(\Delta_m - \phi), \Delta_m)$ are evaluated from values provided by equation (2) using interpolation (typically linear). Bigger weights have been given to the first and the last angles in this approximation to the outer integral of equation (1), due to the fact that there is a big gap between these two angles for which no measurements are taken. (An alternative way of handling such missing data problems is discussed below in Section 4.1.)

The reconstruction based on equations (2–4) is shown on the left of Fig. 2. Pixel values were calculated by evaluating $\hat{f}(r,\phi)$ at the centers of the pixels. The value of A was selected to minimize the norm of the difference between the reconstruction and the original image in a central region; its optimal value is 0.1.

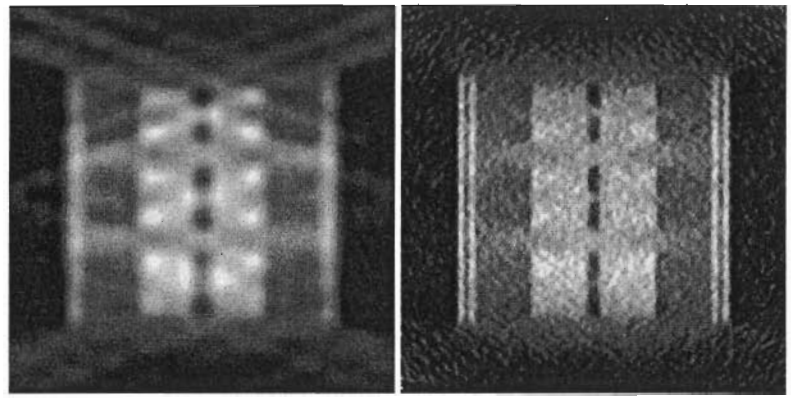


FIGURE 2. Reconstructions from the data shown in Fig. 1; by FBP on the left and by ART with blobs on the right.

Now we give an example, in the same context, of a series expansion method. Such a method is basically different from a transform method, in as much as no attempt is made to find a mathematical expression for the solution of (an idealized version) of the original problem, but rather the problem is converted into a finite-dimensional one by restricting our search for an approximation $\hat{f}$ to only those that can be expressed as a linear combination of some fixed basis functions, i.e. it is assumed *a priori* that

$$\hat{f}(r,\phi) = \sum_{j=1}^{J} x_j b_j(r,\phi) \tag{5}$$

(Herman, 1980, p. 98), and the task becomes that of estimating an appropriate J-dimensional vector $\mathbf{x}$ (whose jth component is x_j). If the data collection method is linear (as is indeed the case with line integrals), then we can say that the ith measurement ($1 \le i \le I$) is

$$y_i \simeq \sum_{j=1}^{J} r_{i,j} x_j \tag{6}$$

(Herman, 1980, p. 100), where $r_{i,j}$ is what the ith measurement would be if the structure consisted of only the jth basis function. Our understanding of the data collection procedure usually allows us to calculate (or, at least, to estimate) the $r_{i,j}$. (For the data collection scheme described in Fig. 1, $I = 70 \times 495 = 34{,}650$.) Our task is now as follows: given a data vector $\mathbf{y}$ (whose ith component is y_i) and knowing the system matrix R (whose (i,j)th entry is $r_{i,j}$), find an $\mathbf{x}$ that 'satisfies' equation (6).

The algebraic reconstruction techniques (ART) are a family of algorithms designed to solve such problems. We now describe and illustrate a specific ART algorithm, called the relaxation method for systems of equalities by Herman (1980), p. 187. (Variants of this algorithm will be discussed and illustrated later in Section 4 for the reconstruction of 3D structures from 2D projection images.)

The algorithm produces a sequence of J-dimensional vectors $\mathbf{x}^{(0)}$, $\mathbf{x}^{(1)}, \ldots$. Typically we choose $\mathbf{x}^{(0)}$ to be the vector of all zeros, and we stop the process after we have cycled through all the data some integer number of times. In the step going from $\mathbf{x}^{(k)}$ to $\mathbf{x}^{(k+1)}$, we pick the next equality from equation (6) to be considered; we denote the index i associated with that equality by i_k. (The order in which the equalities are picked can be of great practical importance; see Herman and Meyer (1993). In the illustrations below, we follow the data access ordering recommended in that paper; its implication for the reconstruction of 3D structures from 2D projection images is discussed in Section 4.) Then

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda \frac{y_{i_k} - \mathbf{r}_{i_k} \cdot \mathbf{x}^{(k)}}{\mathbf{r}_{i_k} \cdot \mathbf{r}_{i_k}} \mathbf{r}_{i_k}, \tag{7}$$

where $\mathbf{r}_i$ (for $1 \leq i \leq I$) denotes the J-dimensional vector whose jth component is $r_{i,j}$ and $\cdot$ denotes the usual inner product. A property of this algorithm (follows from Section 16.8 of Herman (1980)) is that if there is an $\mathbf{x}$ satisfying equation (6) exactly, then the algorithm will converge to such an $\mathbf{x}$, provided that $0 < \lambda < 2$. However, the choice of λ is important for efficient early behavior, especially in the case of noisy data.

Note that the algorithm, as described above, does not depend on the choice of the basis functions. However, this choice is important in practical applications. For the illustration in Fig. 2, we have chosen the so-called blob basis functions of Lewitt (1990, 1992), following the recommendations of Matej and Lewitt (1995, 1996). A comment to be made is that using blobs as basis functions is efficacious in noisy situations (Marabini *et al.*, 1998; Matej *et al.*, 1994). The blobs that we used for the reconstruction of 3D structures from 2D projection images are defined in Section 4.2.2 below.

We applied this algebraic reconstruction technique to the data described in Fig. 1, cycling through the data 10 times. Using the same criterion as in the case of the FBP for optimizing A, we found that the optimal value of λ is 0.05. On the right of Fig. 2, we show the resulting reconstruction.

The figures of merit (FOMs), measuring the quality of the reconstructions (defined in this case as one minus a normalized squared distance between the reconstruction and the original image in a central region), are 0.18 and 0.23 for the FBP and the ART reconstructions, respectively. While this numerical difference is small, the appearances of the reconstructions

are strikingly different, indicating that this particular FOM is not good at capturing the usefulness of a reconstruction for a specific task.

3. DATA COLLECTION, MISSING INFORMATION AND IMPERFECT DATA

3.1. General Considerations

To perform a 3D reconstruction of a macromolecular structure from TEM images, it is necessary to collect a set of views of the structure from different directions. The way in which these different views are obtained varies with the type of aggregate into which the specimens are arranged.

The simplest case is presented by specimens that are aggregated into a helical superstructure (for the purpose of this presentation, it does not matter whether these aggregates are natural or artificially induced). In this case, it is clear that a single image of the helix already contains views from different directions of the individual specimens; this is so because the relative orientation between such a specimen and the electron beam changes along the helix in a well-defined way. This property makes it possible to perform a complete 3D reconstruction of the helix from one image of the helical aggregate (at least up to some aliasing-limited resolution, see DeRosier and Moore, 1970).

There are other types of aggregations that are also highly symmetrical; this is the situation, for instance, for icosahedral viruses, where a general view of the specimen already provides 59 other symmetry-related views around the specimen (Crowther *et al.*, 1970). For the general case, however, we cannot count on any type of symmetries, since we may well have only non-aggregated specimens of an asymmetric structure. In this latter situation, the set of experimentally obtained projection images determining the reconstruction must be collected explicitly for a number of projection directions around the structure. The same situation occurs for specimens arranged in a 2D crystal, since most of the symmetry elements are not orientation-dependent in the tilt direction perpendicular to the grid plane (except for a possible screw axis).

In the rest of this chapter, we focus on the case of reconstructing a structures with no translational symmetries (such as 2D crystals) or point symmetries (such as icosahedral particles). A differentiation has to be made between the 'single-particle reconstruction problem' and the 'tomographic reconstruction problem'. The key conceptual difference is that in the latter case, we aim at reconstructing 'unique' specimens, while in the former case we assume that we have multiple copies of structurally identical specimens, and the task is the reconstruction of the common structure.

The differentiation introduced in the previous paragraph has very strong implications for the approaches that are used to collect data for the 3D reconstruction, as well as for the degree of completeness of the data so

obtained. For the tomographic case, we obtain different views of the specimen from different projection directions by tilting the specimen in the microscope with the help of a goniometer. The projection directions are known, at least in an approximate way, by the angles that are set at the goniometer. For the single-particle case, however, where the different views from identical copies of the macromolecular structure are encountered, the differences between the images are due to the differences in projection directions relative to the common structure. The experimental views from different copies are combined in the reconstruction as if they were coming from the same specimen at different projection directions. Clearly, if a sufficiently large number of views is obtained, then there is no need to tilt the specimens in the microscope. On the other hand, the directions of the projections are not known *a priori* and have to be determined before (or possibly as part of) the reconstruction process.

There are technical limits to the maximum tilt angle that commercial goniometers can achieve, usually around ±60°. Even if specially designed goniometers are constructed, providing, in principle, unlimited tilting (Chalcroft and Davey, 1984), there are physical limits to the maximum tilt angle that can be reached while still obtaining useful images. These limits arise from the increase in the effective specimen thickness that is proportional to $1/(\cos(\textit{tilt angle})$, which makes multiple scattering events more probable at high tilt. For the single-particle case, these limitations do not apply in principle, since the different views can be obtained by combining images from different particles, but the problem then becomes that of 'discovering' the projection direction of each view.

The general problem of reconstructing an asymmetric object is thus posed as the 3D reconstruction of a structure from a finite set of noisy 2D projection images over a possibly restricted angular range and with an approximate knowledge of the data collection geometry. It is due to the combined presence of noise in the image pixels' values, limited angular coverage and lack of information of the projection directions that it becomes necessary to consider reconstruction algorithms capable of dealing with imperfect data.

3.2. Fourier Space Formulation

Much insight into how imperfect data may affect a reconstruction is gained by introducing the Fourier transform (FT) operator. The FT F of a function f is defined by

$$F(\mathbf{R}) = \int_{\mathbb{R}^n} f(\mathbf{r}) exp(-2\pi i \mathbf{r}\cdot\mathbf{R}) d\mathbf{r}. \quad (8)$$

In Cartesian coordinates for the case $n = 3$, this definition becomes (where now f represents a 3D structure)

$$F(X,Y,Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y,z) exp(-2\pi i(xX + yY + zZ)) dx dy dz. \quad (9)$$

The inverse FT f of F is defined by

$$f(\mathbf{r}) = \int_{\mathbb{R}^n} F(\mathbf{R}) exp(2\pi i \mathbf{r}\cdot\mathbf{R}) d\mathbf{R}. \quad (10)$$

We thus obtain pairs of functions (f, F), where F is the FT of f, and f is the inverse FT of F. The convention of denoting such pairs by lower case and upper case letters (lower case associated with real space and upper case with Fourier space) will be used throughout this chapter, in relation to both functions and variables. After applying the FT operator to the original function f, the coordinate system associated with the variable $\mathbf{R}$ of the new function F is different from the coordinate system of the variable r of the original function f. In the following, we will use the term *Fourier space* to refer to the coordinate system of F, and the term *real space* to refer to the original coordinate system of f.

One of the most important results relating f and F to each other is the so-called central section theorem, which states that the FT of an ideal projection image (containing noiseless line integrals) of a 3D structure is equal to a central section of the 3D FT of the structure (e.g. Crowther *et al.*, 1970; others refer to the same result as the projection theorem, e.g. Herman, 1980). Figure 3 shows this relationship. Let us consider the z-axis pointing in the direction of the electron beam. Without loss of generality, let us place the tilt axis along the x-axis. If the ideal projection image g is obtained by tilting the object by θ, then its FT G will be a central section tilted around the X-axis by θ of the 3D FT of the structure. It is common practice to measure tilt angles between +90° and –90° with reference to the z- (or Z-) axis; so, if θ is, say, +60° (with the usual convention that clockwise rotations are negative), then G is a central section that includes the X-axis and makes an angle of $\theta' = -30°$ with the Z-axis.

The central section theorem is fundamental to the following discussion, helping us to visualize how the different practical data collection procedures necessarily result in the partial absence of information regarding the macromolecular structure.

3.3. Single-axis Tilt, Conical Tilt and Generalized Geometries

Single-axis tilt is a common data collection procedure in electron tomography. The specimen is tilted by small increments in the microscope around a fixed axis, and a micrograph of it is obtained in each orientation (Fig. 4a). By calculating the FTs of each of these projections and using the central section theorem, we obtain a set of tilted planes in Fourier space that have the tilt axis in common (Fig. 4b and c). Note that for tilting around a single axis, the 3D problem can be divided into a set of 2D problems, each being the reconstruction in a plane perpendicular to the tilt axis, which can be done by the methods discussed in Section 2. The maximum tilt angle achievable in a microscope is limited, and this limit defines the maximum

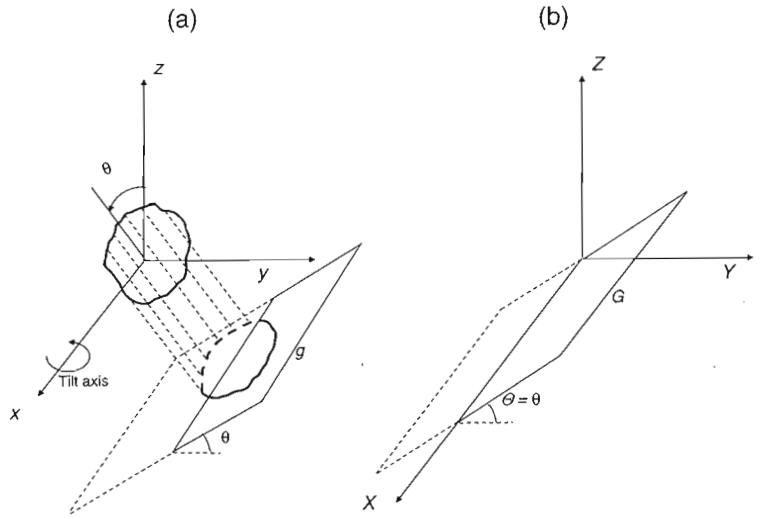


FIGURE 3. The central section theorem. Real space coordinate axes are labeled (*x, y, z*), while Fourier space coordinate axes are labeled (*X, Y, Z*). (a) An object is shown in real space placed near the origin, together with a representation of its projection image *g*, formed by integrals along parallel lines that are all orthogonal to the *x*-axis and are tilted with respect to the *z*-axis by θ. (b) The FT of the object extends over the entire Fourier space. The FT *G* of the projection *g* coincides with the FT of the object restricted to a plane through the origin of the Fourier space (a central section) that contains the *X*-axis and is tilted by θ with respect to the *XY* plane.

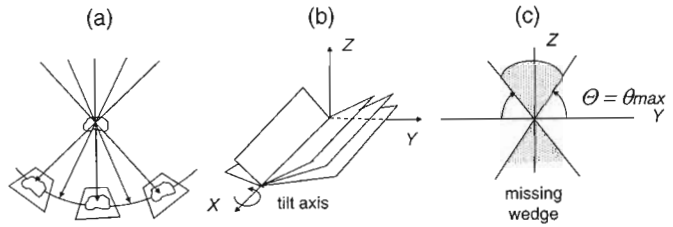


FIGURE 4. Single-axis tilt data collection geometry. (a) Tilting in real space around an axis perpendicular to the plane of this page (which is assumed to be the *x*-axis). (b) Fourier space representation of the information presented in (a); the different projections provide values in the central sections through the 3D FT of the object with the *X*-axis in common. (c) Close-up view of (b) from the the direction of the positive *X*-axis, showing the missing wedge. ((a) is from Radermacher (1980), reproduced with permission.)

tilt angle of the corresponding sections in Fourier space (Fig. 4c). It is then clear that there exists a wedge-shaped region in Fourier space for which no data can be measured (Fig. 4c), and it is consequently termed the missing wedge region. Correspondingly, the measurable area is also wedge shaped, and it is usually termed the *data wedge*.

The missing wedge is centered on the *Z*-axis in Fourier space and, assuming a maximum tilt angle of 60°, has a width of 60° (from –30° to +30° off the *Z*-axis), while the data wedge is centered around the tilt axis in the *XY* plane and has a width of 120°. Being more precise, there are two missing wedges and two data wedges (Fig. 4c).

A generalization of single-axis tilt results in the so-called conical tilt schema for data collection. Conceptually, one can get to this mode of data collection by combining single-axis tilts for tilt axes in many directions. In practice, the same geometry of data collection can be achieved by first tilting the specimen in the microscope by the maximum attainable tilt angle, θ_{max}, and then rotating it in the tilted plane by small angular increments (Fig. 5a); see Radermacher (1988) and Radermacher *et al.* (1987).

Following the same line of reasoning that was used for the single-axis tilt data collection, it is clear that there exists a portion of the FT of the structure for which no data can be measured. The shape of this missing region in Fourier space is a cone centered at the *Z*-axis with a half-width of $90° - |\theta_{max}|$ (Fig. 5b); consequently, it is referred to as the missing cone region.

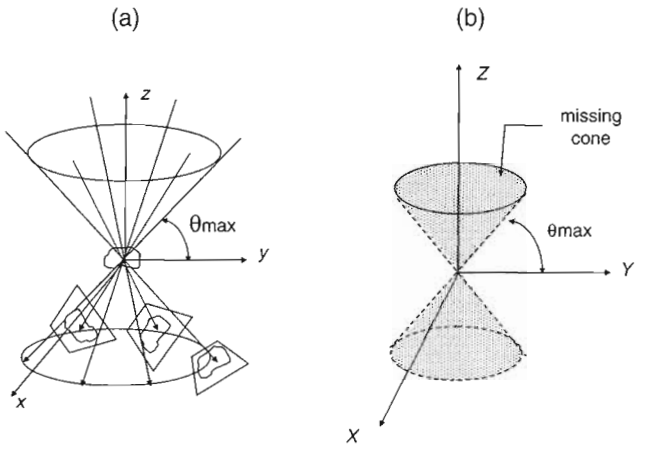


FIGURE 5. Conical tilt data collection geometry. (a) Tilting in real space by θ_{max}, around multiple tilt axes in the *xy* plane. (b) Fourier space representation of the information collected in (a); no experimental data are available in a 'missing cone' around *Z*. (From Radermacher (1980), reproduced with permission.)

Having considered single-axis tilt and conical tilt, the concept of 'generalized geometries' can now be introduced. It refers to those cases in which the pattern of projection directions is not fixed *a priori*. The most common experimental situation illustrating this schema happens in the field of single-particle reconstruction, where we have typically thousands of images that are considered to be views of the same structure from a random set of projection directions whose orientations have to be found before (or possibly during) the reconstruction process.

4. SAMPLE ALGORITHMS

4.1 Weighted Back-projection

There are inversion formulae for the reconstruction of 3D structures from 2D projection images that are very similar in nature to equation (1) for the reconstruction of 2D images from 1D projections; see, for example, Theorem 2.16 of Natterer and Wübbeling (2001). Such formulae consist of an inner integral (corresponding to the convolution in equation (1), but this time the integrations are over the 2D projection planes rather than the 1D projection lines) and a back-projection (which is this time a back-projection into 3D space of the processed 2D projection images). As in all transform methods, a computational procedure has to be designed to implement such ideal inversion formulae for the reconstruction of a 3D structure from the noisy samples of its projection images that is provided by TEM. WBP is one such computational procedure.

Before going into the details of the WBP method, we demonstrate the need for weighting (or, in terms of the approach of Section 2, convolution) with the back-projection. If we do a simple back-projection without weighting (or convolution), then all we do is just 'smear' each of the projections back along the direction from which it was taken. The result is a blurry approximation of the real structure (see Fig. 6a). The purpose of the weighting in WBP (and of the convolution in FBP) is to process the measured projection images so that when the images processed thus are back-projected, the blurring produced by a simple back-projection is avoided, or at least considerably reduced (see Fig. 6b).

In FBP, we convolve with the same filter function q in all projections (see equations (2) and (3)), irrespective of the arrangement of the projection directions. This arrangement is taken into consideration only in the back-projection, and that is why different weights are assigned to different convolved projections in equation (4). The WBP is based on an alternative philosophy: it attempts to design filters (different ones for different projections, each dependent on the geometrical arrangement of all the projections) so that when the convolved projections are simply back-projected (each with the same weight), then the the blurring seen in Fig. 6a (bottom right) will not be observed.

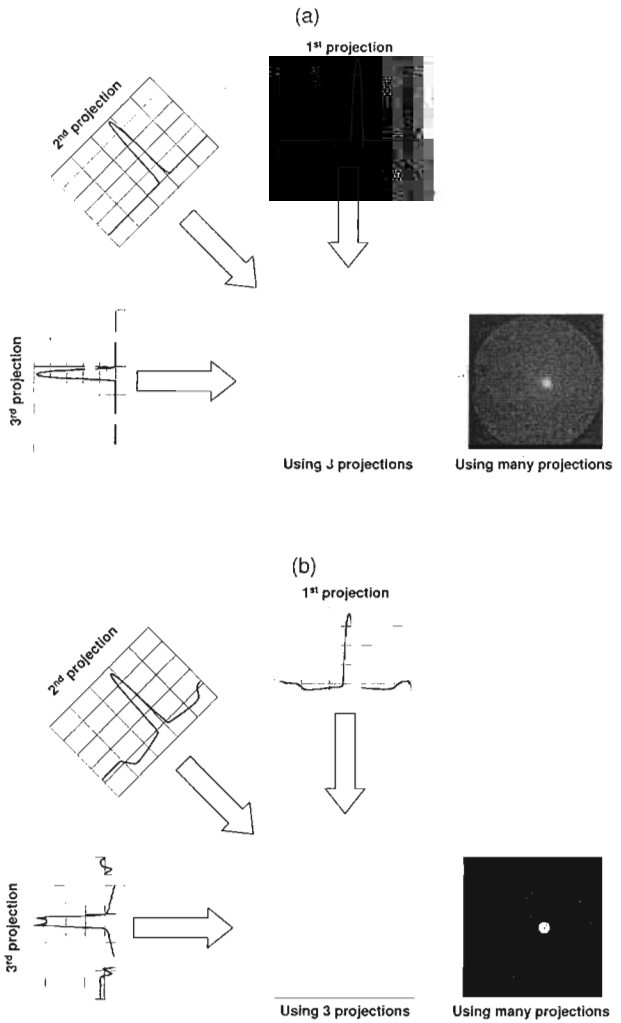


FIGURE 6. Schematic representation of simple back-projection (a) and weighted back-projection (b) for data collected using single-axis tilt geometry. In all cases, a small sphere has been reconstructed and a slice of the reconstruction perpendicular to the tilt axis and intersecting the sphere center is shown. Three plots showing one line (perpendicular to the tilt axis) of three different projections are shown with (b) and without (a) weighting based on the H_m described in the text.

A heuristic way of arriving at the WBP method is the following. Consider a point object at the origin (this is mathematically described by the Dirac delta distribution $\delta(\mathbf{r})$ whose FT has value 1 everywhere) enclosed in a sphere of radius a. Projections of such a 3D point object will be 2D point objects, and when those are back-projected into the enclosing sphere they give rise to line objects of length $2a$ going through the origin. There is one such line for each projection; see Fig. 6a. We can work out mathematically the FT H of this back-projected object. Note that H depends on a, as well as on the geometry of data collection. Restricting H to the central section in Fourier space that is parallel to the plane of the mth projection in real space (see Fig. 3) results in a function $H_m(X_m, Y_m)$ where X_m and Y_m are the coordinates within the central section in question. $H_m(X_m, Y_m)$ can also be worked out mathematically (more details on how this may be done can be found in Chapter 8 of this volume). For the point object of our study, the value of the FT in this central section is the constant 1: this implies that we should be dividing the values of the FT of the back-projection by $H_m(X_m, Y_m)$. This is impossible to do if $H_m(X_m, Y_m) = 0$ and is dangerous to do if $H_m(X_m, Y_m)$ is very small or negative, and so we introduce a (small) positive number *threshold* and replace $H_m(X_m, Y_m)$ by *threshold* whenever $H_m(X_m, Y_m) < threshold$.

Based on these ideas, we can describe the WBP algorithm as follows:

- *Weighting*: for $1 \le m \le M$ (M is the number of 2D projections), calculate p_m^c as the inverse FT of $P_m(X^m, Y^m)/H_m(X^m, Y^m)$, where P_m is the FT of the mth projection and H_m is defined as in the paragraph above.
- *Back-projection*: sum the 'smeared' versions (into the 3D reconstruction volume) of the p_m^c to form $\hat{f}$.

Note that the reconstruction depends on the choices of a and threshold.

In comparing WBP with FBP, we note the following. In the derivation of the weighting in WBP, we have taken into consideration the geometrical arrangement of the projections; in contrast, the corresponding convolving functions in FBP ignore this geometry. However, the weighting in WBP is calculated based on the assumption that we wish to reconstruct a point object at the origin and it is, strictly speaking, not valid for reconstructing point objects away from the origin. In fact, the appearance of a reconstructed point object will be dependent on its location in space. This variation can be reduced by increasing the size of the radius a. The convolving function in FBP is derived without any assumption on the location of the object to be reconstructed, hence the appearance of the reconstruction of a point object is independent of its location. (This last statement is not strictly true: there is also a lesser variation in the appearance of a reconstructed point object due to the fact that the original projections are sampled and the convolved/weighted projections are also calculated at the sample points, resulting in having to use interpolation during the back-

projection. This property is shared by WBP and FBP; its effects were nicely illustrated by Rowland (1979).)

4.2. Block ART

In this section, we discuss a particular variant of ART (introduced in Section 2) that we have used in the experiments reported below. The algorithm depends on the choice of the basis functions b_j in the 3D series expansion (which corresponds to equation (5) that was used for the 2D case). In our work, we have been using two kinds of basis functions (voxels and blobs, to be defined below), but the mathematical nature of the algorithm is the same in both cases, as will be now elaborated.

The basis functions are defined by a grid G that is a finite set $\{\mathbf{g}_1, \cdots, \mathbf{g}_J\}$ of points in 3D space and a basic basis function b that is just a function of three variables. For $1 \le j \le J$, the individual basis functions b_j are defined by

$$b_j(\mathbf{r}) = b(\mathbf{r} - \mathbf{g}_j), \tag{11}$$

i.e. the jth individual basis function is obtained from the basic basis function by shifting its center to the jth grid point. The set of grid points consist of that subset of a regular lattice (e.g. a simple cubic lattice or a body-centered cubic lattice) that lie within a sphere large enough to include the structure to be reconstructed. As in the case of WBP discussed in the previous section, we assume that there are M 2D projections. We further assume that in each 2D projection there are N measurements. Thus the total number of measurements is $I = MN$. Just as in Section 2, we use $r_{i,j}$ to denote what the ith measurement would be if the structure consisted of only the jth basis function and $\mathbf{r}_i$ (for $1 \le i \le I$) to denote the J-dimensional vector whose jth component is $r_{i,j}$. The basic difference between the variant of ART that is discussed in Section 2 and the block ART that is presented here is that in the latter the algorithm proceeds by taking into consideration not only one measurement at a time, but rather a block of measurements that come from a particular 2D projection. This results in the replacement of equations (7) by

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda \sum_{i=m_k N+1}^{(m_k+1)N} \frac{y_i - \mathbf{r}_i \cdot \mathbf{x}^{(k)}}{\mathbf{r}_i \cdot \mathbf{r}_i} \mathbf{r}_i, \tag{12}$$

where m_k is the index of the 2D projection used in the kth iterative step. Just as in the 2D case, we choose $\mathbf{x}^{(0)}$ to be the vector of all zeros. The theory of such block ART algorithms was introduced by Eggermont *et al.* (1981). Their limiting convergence properties have been carefully studied in that and consequent publications (see Censor and Zenios, 1997); however, such limiting convergence results are not necessarily relevant in practical applications in which (for reasons of computational time and cost) we wish to

stop the iterative process as early as possible. One of the reasons for this early termination is that many times the solution in the limit is something like a least squares solution, which are known to fit the noise as well as the signal, and so the corresponding reconstructed volumes are too noisy (since they tend to minimize the squared error between the noisy experimental projections and the projections taken from the reconstructed volume).

The order in which the projections are picked is important. Equation blocks are determined by all those pixels belonging to the same projection image. We order the projections in such a way that each one is as nearly orthogonal as possible to the previous two projections (a projection is orthogonal to another projection if their corresponding projection directions are orthogonal).

In order to turn equation (12) into a reconstruction algorithm, we need to decide how to choose the parameter λ and at which value of k to stop the iterative process.

These choices depend on many things (see, for example, Marabini *et al.*, 1997, 1998). Generally speaking, for the same number of iterations, the value of λ should be smaller if the data are noisier. Typically, in the single-particle reconstruction problem (in which the number M of 2D projections tends to be quite large), it is sufficient to cycle through the measurement data only once (i.e. to stop the iterative process at $k = M$). In the tomographic reconstruction problem (in which M tends to be much smaller), one needs to cycle through the data several times.

4.2.1. Block ART with Voxels

A basic voxel basis function depends on a variable that we will call here VEL (for voxel edge length): it is defined to have the value 1 at points strictly inside the cube that is centered at the origin and that has edges of length VEL parallel to the coordinate axes, and to have the value 0 at points strictly outside this cube. (While this does not make any difference in any possible application, for completeness we can define the basic voxel basis function to have the value 1/2 at the faces, the value 1/4 at the edges, and the value 1/8 at the corners of the cube.) The associated grid G is defined as the set of all points of the form VEL$\mathbf{R}$, where the coordinates of the vector $\mathbf{R}$ are integers, that lie within a sphere that is large enough to include the structure to be reconstructed. A consequence of this definition is that any reconstruction produced by block ART with voxels will have a constant value within cubes of edge length VEL.

4.2.2 Block ART with Blobs

A basic blob basis function is a generalization of a well-known class of functions in digital signal processing called Kaiser–Bessel (see Lewitt, 1990); it is spherically symmetric (i.e. its value at a point depends only on

the distance r of that point from the origin), has non-zero values only in a sphere (of radius a) around the origin, and smoothly decreases from a positive value at the origin to zero at the surface of that sphere. The general form of a blob is:

$$b(m,\alpha,a;r)=\begin{cases}\dfrac{I_m\left(\alpha\sqrt{1-\left(\dfrac{r}{a}\right)^2}\right)}{I_m(\alpha)}\left(\sqrt{1-\left(\dfrac{r}{a}\right)^2}\right)^m, & \text{if } 0\le r\le a,\\ 0, & \text{if } r>a,\end{cases} \tag{13}$$

where I_m denotes the modified Bessel function of order m and α is a parameter controlling the blob shape. The three parameters m (a non-negative integer), a and α (non-negative real numbers) control the smoothness and shape of a blob and influence the results yielded by reconstruction and visualization algorithms; therefore, the appropriate selection of them is highly important. In our work, we set m equal to 2, which makes the blobs have continuous first derivatives everywhere.

The choice of the grid G is also important. It was shown by Petersen and Middleton (1962) that the body-centered cubic (*bcc*) grids provide the most efficient sampling of $\mathbb{R}^3$. The *bcc* grids are defined by

$$B_\Delta=\{\Delta\mathbf{k}\,|\,\mathbf{k}\in\mathbb{Z}^3 \text{ and } k_1\equiv k_2\equiv k_3 (mod\ 2)\}, \tag{14}$$

where $\mathbf{k}$ is a 3D vector, whose components (k_1, k_2 and k_3) belong to the set of integers denoted by $\mathbb{Z}$, and Δ is a positive real number (the sampling distance). In order to visualize this grid, we can use a small portion of it and take advantage of its periodic repetition (see Fig. 7). For reconstruction purposes, Matej and Lewitt (1995) demonstrated that whenever a linear combination of blobs is employed to obtain a reconstruction, the *bcc* grids provide desirable sets of locations for the center of the blobs.

Having decided that we use $m = 2$ and the *bcc* grid, there are three parameters to be chosen: Δ, a and α. Clearly, to be able to approximate

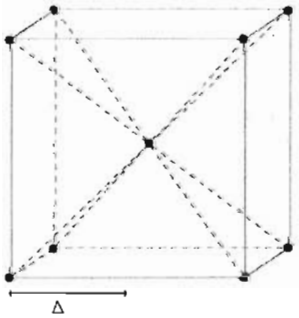


FIGURE 7. Points in the body-centered cubic grid in a $2 \times 2 \times 2$ portion of space (assuming $\Delta = 1$). The rest of the points can be obtained by filling in space by the most natural repetition of the indicated $2 \times 2 \times 2$ portion.

arbitrary structures, the value of Δ should be small. However, in a fixed volume of space, the number of grid points (and consequently the computational cost of a reconstruction algorithm) is proportional to $1/\Delta^3$ and so practical considerations do not allow us to choose Δ to be very small. The cost of reconstruction is also proportional to a^2. The computational cost does not depend on α, and so this parameter may be chosen purely based on the quality of the resulting reconstructions. Matej and Lewitt (1995) proposed a method for the selection of the parameters with the aim of ensuring that ART will produce 'good' reconstructions. In particular, they postulate that a linear combination of blobs with $x_j = 1$, for $1 \leq j \leq J$, should approximate a constant-valued function. They show that for this we should select Δ, a and α such that

$$\alpha = \sqrt{2\pi^2\left(\frac{a}{\Delta}\right)^2 - 6.987932^2} \tag{15}$$

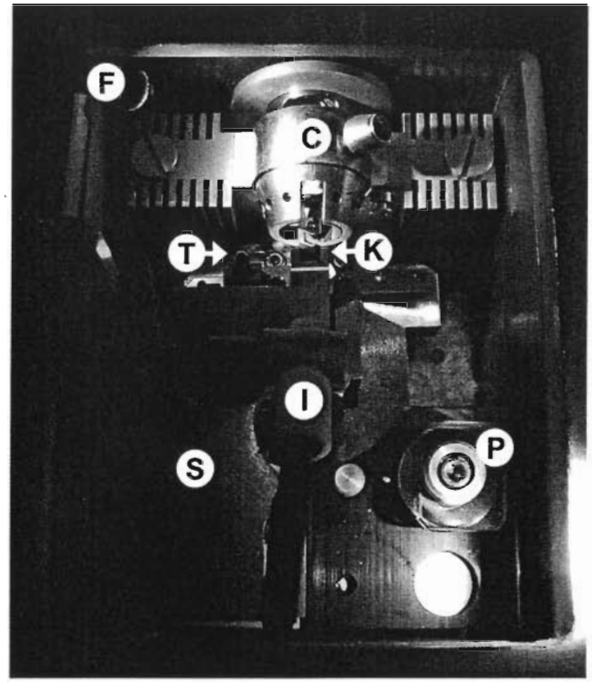
is satisfied.

Garduño and Herman (2004) reported that the parameters yielded by the methodology suggested by Matej and Lewitt (1995) produced in some cases non-convex reconstructions from data obtained from convex structures, causing a significant inaccuracy in the visualization of the resulting surfaces. To correct this problem, Garduño and Herman (2004) proposed an additional criterion for the selection of the parameters Δ, a and α: a should be chosen as small as possible, consistent with both satisfying equation (15) and achieving the result that if two blobs at nearest grid points in the grid B_Δ (those separated by $\sqrt{3}\Delta$) are given coefficients 1 with all other blobs given coefficients 0, then the surface obtained by thresholding at $t = 0.5$ should enclose a convex set. The selection of the latter criterion is not arbitrary but is based on the fact that the location of the two nearest blobs will determine the best resolution that can be achieved by an approximation using blobs. Furthermore, given a choice of Δ, such a criterion provides us with a unique pair a and α that satisfies equation (15). This is the methodology used in this chapter for determining the blob parameters.

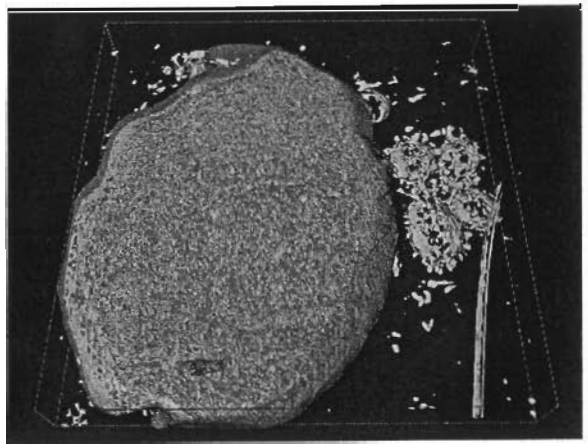
5. PRACTICAL ILLUSTRATION

5.1. General Considerations

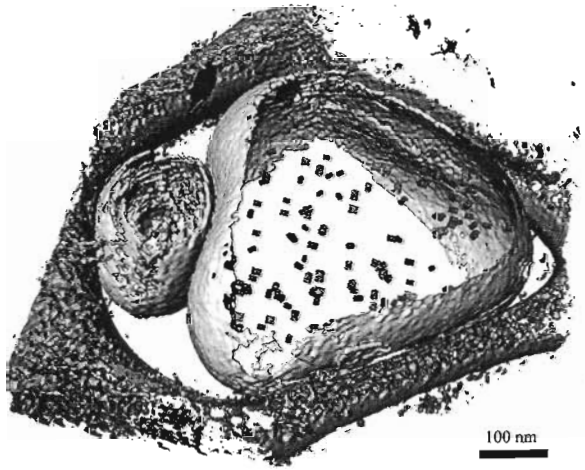
In this section, we present practical cases of how different reconstruction methods treat the 'imperfect data' problem that is typical in 3D electron tomography. As a general remark, it is important to stress that 'the best reconstruction method' does not exist. Each technique makes its own assumptions about the possible set of solutions, the noise or the type and extent of *a priori* knowledge. Depending on the specific problem at hand, these assumptions may or may not be the most appropriate ones.



Chapter 2, FIGURE 1. Cryochamber of an UCT ultramicrotome with an EM-FCS cryo kit (Leica, Vienna, Austria). The chamber is cooled to −160°C and is filled with cold, dry nitrogen vapor. One of the two auxiliary fiber-optic illuminators is seen, labeled (F). The specimen is held in the chuck (C). The knife stage is shown positioned for trimming, with the diamond trimming tool (T) adjacent to the specimen. For microtomy, the sectioning knife (K) is moved in front of the specimen, and a brass shelf (S) is pushed forward so that it is close to the knife edge. The shelf, fabricated in-house, provides a surface for collecting sections. Both diamond knives (T and K) are from Diatome (Biel, Switzerland). The head of the ionizer unit (I; Static Line II, Haug, Biel, Switzerland) is mounted ~3 cm from the knife edge. The glass screw press tool is mounted at (P).



Chapter 12, FIGURE 14. Eigenvector-based segmentation presented in Fig. 13, visualized in three dimensions. (From Frangakis *et al.* (2002), reproduced with permission of Elsevier).



Chapter 14, FIGURE 6. Segmented and denoised 3D image of a vesicle containing two types of macromolecules, visualized in different colors. In the isosurface visualization, two vesicles (in beige) surrounded by the carbon film (in gray) are shown. In the larger vesicle, the two different macromolecular complexes were located and positively identified, and were then replaced with their low-pass-filtered X-ray structure. Proteasomes are shown in orange and thermosomes in blue. (From Frangakis *et al.* (2002), reproduced wth permission of the National Academy of Sciences).

There are two important points directly related to 3D electron microscopy applications that need to be addressed here. One is the characterization of the image degradations introduced by the electron microscope as an image-forming system, the other is the noise model to be used in the reconstruction process.

Not surprisingly, the electron microscope cannot be considered a perfect imaging system. It is well known that it introduces some degradations into the images it renders. Erickson and Klug (1971) presented an analysis on how these instrumental degradations can be modeled, within certain limitations, in terms of a single-system optical transfer function (OTF), providing a parametric model for it (in the field of electron microscopy, the OTF is usually called the contrast transfer function or CTF). The corresponding point-spread function is referred to as the 2D TEM-PSF. In general, the 2D TEM-PSF, being the inverse FT of the instrument CTF, does not have a finite support, although in practice it can be approximated by a finite-support blurring function.

With respect to the contents of this chapter, knowledge of the CTF is important because it tells us that the relationship between the 3D object and the 2D image obtained from it is not that of a simple projection, but rather a projection convolved by the 2D TEM-PSF (defined in the previous paragraph). This fact should be kept in mind when modeling the image formation.

In a first approximation to the reconstruction problem, one might think of neglecting this instrumental degradation. However, this is not an acceptable choice in the quest for high-resolution structural information. Additionally, the fact that the geometry of data collection is known only in an approximate way introduces imperfections in our initial data set as well.

Regarding the noise model, it is clear that part of the noise is due to electron counting, and thus follows a Poisson process. It is also clear that there are other contributions coming from the specimen embedding and supporting material, from the photographic emulsion and processing, the photographic grain, and inaccuracies in the scanning devices. Other sources of 'noise' are more difficult to consider, such as the inaccuracies in the determination of the geometry of data collection or inaccuracies in the determination and restoration for the CTF. It follows from all these considerations that realistic noise models are not trivial, and therefore should always be used with some caution, being aware of their limitations.

5.2. Test Data Set Description and Testing Approach

In this section, we present a practical study of two widely used reconstruction methods: WBP and block ART. Only block ART with blobs has been studied, based on its well-established superiority to block ART with voxels (Matej and Lewitt, 1996; Matej *et al.*, 1994). The two methods have been tested on three simulated data sets, two typical of high-resolution

studies of single particles (one typical of the conical tilt geometry as depicted in Fig. 5, and the other typical of a geometry with even angular distribution) and the third typical of electron tomography. Each data set consists of a number of independently generated collections of projection images (each collection in the data set corresponds to one instance of what is denoted by g in Section 1). The performance of the methods is measured and compared using a well-defined FOM in a statistical manner. This work has been done using the X-Window-based Microscopy Image Processing Package (XMIPP) that is freely available from the authors' web site (http://xmipp.cnb.uam.es/).

The first test data set consisted of 30 collections of 10,000 projection images which were obtained using random projection directions of the thermosome from *Thermoplasma acidophilum* (Ditzel *et al.*, 1998). The sampling interval, in both the images and the volume, was 0.266 nm. The dimensions of the images and of the volumes were 96×96 and $96 \times 96 \times 96$, respectively. The simulation was intended to represent a practical case of high-resolution (below 1 nm) single-particle reconstruction.

The second test data set consisted of 10 collections of 10,000 projection images each that were obtained randomly within a conical tilt data collection geometry with a tilt angle of 60° of the thermosome from *T. acidophilum* (Ditzel *et al.*, 1998).

The third test data set consisted of 10 collections of 70 regularly distributed projection images following the single-axis tilt data collection geometry with a maximum tilt angle of ±60° of the mitochondria-like phantom described by Fernández *et al.* (2002) and Bilbao-Castro *et al.* (2004). Each of the images was of dimensions 256×256 pixels, and the reconstructed volume was $256 \times 256 \times 256$ voxels. The rationale here was to simulate the case of a tomographic study of a cell organelle.

Note that these data sets represent very different mathematical situations when they are translated into a system of conditions defining the reconstruction in equation (6). In the first two cases, we have a very large collection of images and the task is the reconstruction of a relatively small volume, while it is exactly the opposite that happens in the third case, since here we have relatively few large images and the task is to reconstruct a large volume. To be precise, in the first two cases, the total number of pixels in a collection g of projection images is $10{,}000 \times 96 \times 96$, while the number of voxels in the volume is $96 \times 96 \times 96$; in the third case, the total number of pixels in g is $70 \times 256 \times 256$ and the number of voxels is $256 \times 256 \times 256$. As a consequence, the system (equation 6) is overdetermined in the first two cases by a factor of ~100, and is underdetermined in the third case. Additionally, as the size of the images become larger and larger in tomographic applications with a reduced number of views, the tomographic reconstruction problem will become more and more underdetermined, while in the single particle cases, as the number of images coming from different specimens become larger, the reconstruction problem becomes more and more overdetermined.

Another difference is the coverage of the Fourier space that can be achieved using g. This is reasonably uniform in the first case, has a missing cone in the second case (see Fig. 5) and has a missing wedge in the third case (see Fig. 4). WBP and ART each have their ways of handling such data: WBP takes the geometry of projection directions into consideration in designing the weighting function, while the system (equation 6) that ART attempts to solve is explicitly based on the projection directions.

Regarding the generation of the projection images of the thermosome, each of them was calculated from the set of atomic coordinates contained in Protein Data Base (PDB) entry 1A6D by replacing each non-hydrogen atom in the structure with a blob (equation 13) with appropriately chosen parameters, which allows a perfect analytical description of the projections. Projections were then sampled at intervals of 0.266 nm. The generation of the mitochondrion-like phantom was performed as in Fernández *et al.* (2002). This phantom consists of hollow cylinders representing the membranes and a set of solid cylinders simulating the cristae. The cristae are embedded in a region of intermediate density, resembling the mitochondrial inner matter.

The electron microscope-induced projection blurring was taken into account using a family of CTFs whose average is plotted in Fig. 8. The defocus value was randomly changed between views within the interval [–0.315, –0.285] μm for the single particle case and the interval [–1.575, –1.475] μm for the tomographic reconstruction. In the experiment with random conical tilt, the tilt angle was fixed to 60°. Restoration of the CTF was performed such that it tended to resemble a practical experimental case: only a phase flipping was performed based on a CTF estimate provided by all the projection images.

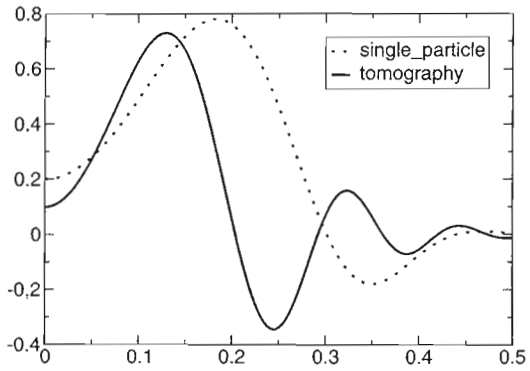


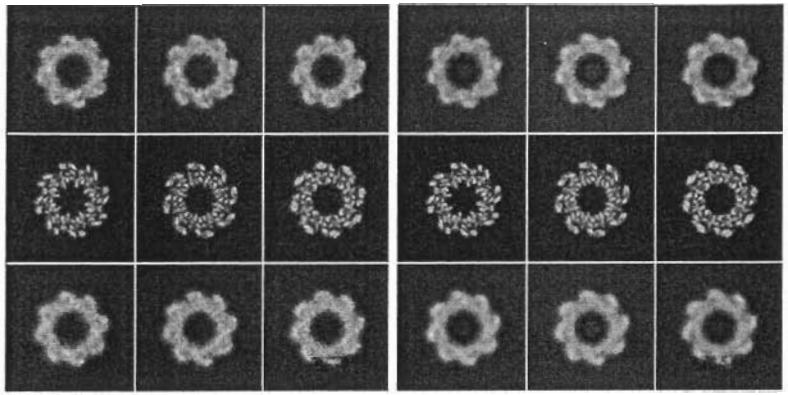
FIGURE 8. Plot of the average CTFs used in the single particle (dotted line) and the tomographic (solid line) experiments, respectively. The horizontal axis represents normalized frequencies so that the maximum frequency given by the Nyquist theorem is mapped to 0.5.

As for inaccuracies related to the determination of the actual geometry of projection directions, Gaussian noise was added to all the parameters determining it, so that views were not perfectly centered, but they had a noise with 2 pixel edge lengths standard deviation in both x and y, and 2° in each of the three Euler angles. Gaussian noise was added to each pixel value so that the signal-to-noise ratio was 1/3 in the case of single particles and 2/3 in the case of tomography.

The testing of the different algorithms should be based on a statistical comparison of their relative ability to perform well-defined and relatively simple tasks coded in the form of FOMs. However, since all algorithms have free parameters that have to be adjusted, prior to their comparisons it is necessary to perform a parameter optimization approach in which, for all reconstruction algorithms, their free parameters are adjusted to optimal performance. The parameter optimization and the algorithm evaluation should be performed on the sames FOMs. This methodology has been presented in a number of previous works, such as Furuie *et al.* (1994), Matej *et al.* (1994) and Sorzano *et al.* (2001). For this chapter, we have selected, as an illustrative example, a FOM based on the Euclidean distance between the reconstructed volume and the original volume: $scL2 = 1 - \frac{1}{V}\sum_{v=1}^{V}\left(\frac{p_v - r_v}{2}\right)^2$ (where p_v and r_v are the values assigned to the vth voxel of the phantom and the reconstruction, respectively, and V is the total number of voxels within a sphere of radius half the reconstructed volume edge). Clearly, this FOM measures only a global characteristic of reconstruction accuracy.

The parameters to be optimized for WBP, as described in Section 4.1, are a and *threshold*. In fact, based on past experience, we have in all cases fixed a so that it is half the length of the edge of the cubic volume within which we perform the reconstruction. For block ART, we need to select the relaxation parameter, λ, which controls the magnitude of the update of the reconstructed volume during each iteration; see equation (12). Note that the number of iterations for block ART is normally also a free parameter. However, we set it to M (the number of projection images) for the overdetermined case following the results and discussions presented in Marabini *et al.* (1998), and to $10M$ for the underdetermined case (see Marabini *et al.*, 1997). With these choices for the number of iterations, the selected values of λ were 0.3124, 0.0625 and 0.5, respectively, for the three test data sets.

In Fig. 9a, we present the results corresponding to the first data set. In this overdetermined case, the performances of the two reconstruction algorithms according to this FOM called *scL2* were similar, rendering FOM values of 0.62 ± 0.03 for block ART and 0.60 ± 0.03 for WBP. The first value is the mean of the FOM over the 30 collections of 10,000 projection images each, and the second value is the standard deviation. Bearing in mind that the standard deviation of the mean of 30 samples is $(1/\sqrt{30})$ times the standard deviation of the original distribution, we see that the difference



(a) Radomly distributed images (b) Conically distributed images

FIGURE 9. Three slices perpendicular to the Z-axis corresponding to, from top to bottom, the block ART reconstruction, the phantom volume and the WBP reconstruction, respectively.

between the means is over three times the standard deviations of the means, implying that the observed difference has statistical significance, albeit not a large one. (The phrase 'statistical significance' is used here in the sense that is standard in hypothesis testing. Its meaning is not that the difference between the means of the outcomes of the experiments using the two methods is large, or even that it is necessarily important from the point of view of the application. Rather it is a measure of our confidence by which we can reject the null hypothesis that the methods are equally good, from the point of view of the FOM, in favour of the alternative hypothesis that the one for which the mean is larger is the better one. This confidence depends not only on the difference between the means, but also on the variability of the experiments, as indicated by the standard deviations. Thus statistical significance is indicated even by a very small difference in means, as long as the standard deviations are correspondingly small.)

Regarding the case in which there was a missing cone of data, the results are presented in Fig. 9b. This time, the global FOMs (0.53 ± 0.01 for block ART and 0.49 ± 0.02 for WBP) are less similar: bearing in mind again that the means are averages of several (this time 10) samples, the difference between the mean FOMs for the two reconstruction methods is even more significant than in the previous case, with block ART presenting more details than back-projection. This type of result was expected, since we found in our previous work (Marabini *et al.*, 1997) that WBP is more sensitive to a missing angular range than ART with blobs.

As for the third case of a tomographic reconstruction problem, it is an underdetermined situation with a missing angular range. The results

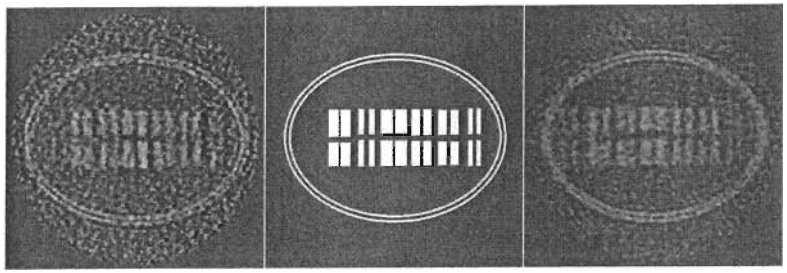


FIGURE 10. One slice perpendicular to the *Z*-axis corresponding, from left to right, to the ART reconstruction, the phantom volume and the WBP reconstruction, respectively.

obtained are presented in Fig. 10. The FOM specified above provides values of 0.9958 ± 0.0001 for the block ART reconstruction and of 0.9954 ± 0.0002 for WBP that are again significantly different. Two observable disadvantages of WBP when compared with block ART are: a blurring along the *z*-axis (notice the two extra cylinders that appear on the left in the WBP reconstruction that have been 'smeared' from another plane) and the loss of the resolution in the 'bilayer' that surrounds the specimen as it goes farther away from the center (a behavior that is in agreement with the theory, since as we move away from the center we increasingly violate the assumed nature of the point-spread function of simple back-projection, and so the correction becomes less reliable).

6. DISCUSSION

In this work, we have presented two different families of reconstruction algorithms (transform methods and series expansion methods) used in the field of 3D electron microscopy, including a more detailed analysis of two common algorithms: WBP and block ART. There are several lessons to be learned from this exercise.

The first one is the realization that no algorithm can be treated as a black box, and that special attention has to be given to the actual value of the parameters that characterize each algorithm in the context of a particular application. In this study, we have presented an approach in which the value of these parameters is selected according to the maximization of a well-defined FOM. Clearly, there is ample opportunity to define other FOMs and, consequently, to 'tune' the algorithms for a better recovery of certain types of information, such as 'feature separability', 'contrast' etc., which is a common practice in the field of computed tomography (CT) in medicine. This review does not pretend to explore this issue in depth, as it

deserves a work of its own, but simply presents this situation explicitly since many times it is hidden in the common use of the methods.

The second lesson is the importance of limits, other than the ones imposed by the reconstruction algorithms, on our capacity to achieve better reconstructions and, in this way, higher resolution biological information from our 3D reconstructions. Of utmost importance are the errors due to our imperfect knowledge of the geometry of data collection and, also, due to the inaccuracies introduced by the instrument's CTF. This consideration indicates that a very interesting way to overcome these limitations could be a combination, into a single algorithm, of the functionalities of reconstruction techniques and geometry estimation methods (see Scheres *et al.* (2005) for a way to advance in this direction).

Finally, the third reflection is that under the general title of '3D EM Reconstruction Methods' there are quite different types of situations. A clear example is the realization that typically single-particle reconstruction problems are greatly overdetermined in terms of the basic characteristics of the systems of equations that arise from them, while tomographic reconstruction problems fall at the opposite extreme. Furthermore, the two types of problems are evolving in such a way that the former one becomes even more overdetermined, while the latter one becomes more and more underdetermined. Being so different in their basic mathematical characteristics, it is clear that the relative performance of the methods in these two areas could also be very different: what performs 'best' in the one case may be inferior in the other.

ACKNOWLEDGEMENTS

The authors are thankful to Dr J. J. Fernandez for kindly providing the mitochondria phantom. We acknowledge partial support from the 'Comunidad Autónoma de Madrid' through grants CAM-07B-0032-2002 and GR/SAL/0342/2004, the 'Comisión Interministerial de Ciencia y Tecnología' through grants BIO2001-1237, BIO2001-4253-E, BIO2001-4339-E, BIO2002-10855- E, BFU2004-00217/BMC and TIC990361, the European Union through grants QLK2-2000-00634, QLRI-2000-31237, QLRT-2000-0136 and QLRI-2001-00015, the NIH through grant HL70472, and the 'Fundación de Investigación Sanitaria' through network FIS G03/185 and grant PI040683.

REFERENCES

Bilbao-Castro, J. R., Sorzano, C. O. S., García, I. and Fernández, J. J. (2004). Phan3D: design of biological phantoms in 3D electron microscopy. *Bioinformatics* **20**:3286–3288.

Censor, Y. and Zenios, S. A. (1997). *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, New York.

Chalcroft, J. P. and Davey, C. L. (1984). A simply constructed extreme-tilt holder for the Philips eucentric stage. *J. Microsc.* **134**:41–48.

Crowther, R. A., DeRosier, D. J. and Klug, F. R. S. (1970). The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proc. R. Soc. A* **317**:310–340.

DeRosier, D. J. and Moore, P. B. (1970). Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *J. Mol. Biol.* **52**:355–369.

Ditzel, L., Lowe, J., Stock, D., Stetter, K. O., Huber, H., Huber, R. and Steinbacher, S. (1998). Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**:125–138.

Eggermont, P. P. B., Herman, G. T. and Lent, A. (1981). Iterative algorithms for large partitioned linear systems with applications to image reconstruction. *Linear Algebra Appl.* **40**:37–67.

Erickson, H. P. and Klug, A. (1971). Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs. *Philos. Trans. R. Soc. B* **261**:105–118.

Fernández, J., Lawrence, A., Roca, J., García, I., Ellisman, M. and Carazo, J. (2002). High performance electron tomography of complex biological specimens. *J. Struct. Biol.* **138**:6–20.

Furuie, S. S., Herman, G. T., Narayan, T. K., Kinahan, P. E., Karp, J. S., Lewitt, R. M. and Matej, S. (1994). A methodology for testing for statistically significant differences between fully 3D PET reconstruction algorithms. *Phys. Med. Biol.* **39**:341–354.

Garduño, E. and Herman, G. T. (2004). Optimization of basis functions for both reconstruction and visualization. *Discrete Appl. Math.* **139**:95–112.

Herman, G. T. (1980). *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, New York.

Herman, G. T. and Meyer, L. B. (1993). Algebraic reconstruction techniques can be made computationally efficient. *IEEE Trans. Medical Imaging* **12**:600–609.

Lewitt, R. M. (1990). Multidimensional digital image representations using generalized Kaiser–Bessel window functions. *J. Optic. Soc. Am. A* **7**:1834–1846.

Lewitt, R. M. (1992). Alternatives to voxels for image representation in iterative reconstruction algorithms. *Phys. Med. Biol.* **37**:705–716.

Marabini, R., Herman, G. T. and Carazo, J. M. (1998). 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). *Ultramicroscopy* **72**:53–65.

Marabini, R., Rietzel, E., Schröder, R., Herman, G. T. and Carazo, J. M. (1997). Three dimensional reconstruction from reduced sets of very noisy images acquired following a single-axis tilt schema: application of a new three-dimensional reconstruction algorithm and objective comparison with weighted backprojection. *J. Struct. Biol.* **120**:363–371.

Matej, S., Herman, G. T., Narayan, T. K., Furuie, S. S., Lewitt, R. M. and Kinahan, P. E. (1994). Evaluation of task-oriented performance of several fully 3D PET reconstruction algorithms. *Phys. Med. Biol.* **39**:355–367.

Matej, S. and Lewitt, R. M. (1995). Efficient 3D grids for image reconstruction using spherically-symmetric volume elements. *IEEE Trans. Nucl. Sci.* **42**:1361–1370.

Matej, S. and Lewitt, R. M. (1996). Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Trans. Med. Imaging* **15**:68–78.

Natterer, F. and Wübbeling, F. (2001). *Mathematical Methods in Image Reconstruction*. SIAM, Philadelphia.

Petersen, D. P. and Middleton, D. (1962). Sampling and reconstruction of wavenumber limited functions in *n*-dimensional Euclidean spaces. *Inform. Control* **5**:279–323.

Radermacher, M. (1980). Dreidimensionale Rekonstruktion bei kegelförmiger Kippung im Electronenmikroskop. PhD Thesis, Technische Universität München.

Radermacher, M. (1988). Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J. Electron Microsc. Tech.* **9**:359–394.

Radermacher, M., Wagenknecht, T., Verschoor, A. and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal unit. *J. Microsc.* **146**:113–136.

Radon, J. (1917). Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. *Math. Phys. Klasse* **69**:262–277.

Rowland, S. (1979). Computer implementation of image reconstruction formulas. In *Image Reconstruction from Projections* (G.Herman, ed.). Springer-Verlag, Berlin, pp. 29–79.

Scheres, S. H. W., Valle, M., Núñez, R., Sorzano, C. O. S., Marabini, R., Herman, G. T. and Carazo, J. M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348**:139–149.

Sorzano, C. O. S., Marabini, R., Boisset, N., Rietzel, E., Schröder, R., Herman, G. T. and Carazo, J. M. (2001). The effect of overabundant projection directions on 3D reconstruction algorithms. *J. Struct. Biol.* **133**:108–118.