Three-Dimensional Electron Microscopy of Macromolecular Assemblies

Joachim Frank

Wadsworth Center for Laboratories and Research
State of New York Department of Health
The Governor Nelson A. Rockefeller Empire State Plaza
Albany, New York
and
Department of Biomedical Sciences
State University of New York at Albany
Albany, New York

Academic Press

San Diego New York Boston London Sydney Tokyo Toronto

Chapter 5

Three-Dimensional Reconstruction

I. Introduction

The value of projection images is quite limited if one wishes to understand the architecture of an unknown structure (Fig. 5.1). This limitation is illustrated by the early controversies regarding the three-dimensional (3D) model of the ribosome, which was inferred, with different conclusions, by visual analysis of electron micrographs [see, for instance, the juxtaposition of different models in Wittmann's (1982) review]. In 1968, DeRosier and Klug published the first 3D reconstruction of a biological object, a phage tail with helical symmetry (DeRosier and Klug, 1968). Soon after that, Hoppe published an article (Hoppe, 1969) that sketches out the strategy for reconstruction of a single macromolecule lacking order and symmetry [see Hoppe *et al.* (1974) for the first 3D reconstruction of such an object from projections]. Since then, the methodologies dealing with the two types of objects have developed more or less separately, although the existence of the same mathematical thread (Crowther *et al.*, 1970) has been often emphasized.

This part of the volume is organized in the following way: first, some basic mathematical principles underlying reconstruction are laid down. Next, the different data collection schemes and reconstruction strategies are described, which answer the questions of how to maximize information, minimize radiation damage, or determine the directions of projections, while leaving open the choice of reconstruction algorithm. The main algorithmic approaches are subsequently covered: weighted back-projection, Fourier interpolation methods, and iterative algebraic methods.

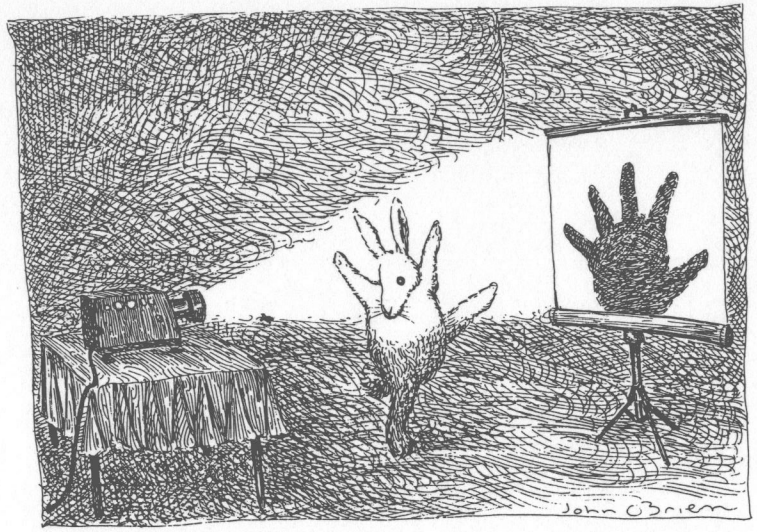


Fig. 5.1. A single projection image is plainly insufficient to infer the structure of an object. Drawing by John O'Brien; © 1991 The New Yorker Magazine.

Against this background, the random-conical reconstruction scheme is described as a scheme of data collection and processing that has gained practical importance and underlies most single particle reconstructions to date (see Bibliography at the end of the volume). In a final section, restoration and angular refinement are covered.

II. General Mathematical Principles

A. The Projection Theorem, Radon's Theorem, and Resolution

The projection theorem, which is of fundamental importance in the attempts to recover the object, is implied in the mathematical definition of a multidimensional Fourier transform. In two dimensions, let us consider the Fourier representation of a function,

$$f(x,y) = \int_{k_x}\int_{k_y} F\big(k_x(x,y_y)\big)\exp\big[-2\pi i(k_x x + k_y y)\big]dk_x dk_y. \quad (5.1)$$

Now we form a one-dimensional projection in the y-direction. The result is

$$q(x) = \int f(x,y)dy$$
$$= \int_y \left[\int_{k_x} \int_{k_y} F(k_x,k_y) \exp\left[-2\pi i(k_x x + k_y y)\right] dk_x dk_y \right] dy, \quad (5.2)$$

which immediately yields

$$g(x) = \int_{k_x} \int_{k_y} F(k_x,k_y)\delta(k_y)dk_x dk_y = \int_{k_x} F(k_x,0)dk_x, \quad (5.3)$$

where $\delta(k_y)$ is the delta function. This means that *the projection of a two-dimensional function $f(x,y)$ can be obtained as the inverse one-dimensional Fourier transform of a central section through its 2D Fourier transform $F(k_x,k_y) = F[f(x,y)]$*. The above "proof" is very simple when the projections in x and y directions are considered, making use of the properties of the Cartesian coordinate system. Of course, the same relationship holds for any choice of projection direction, see for instance the formulation by Dover *et al.* (1980). An analogous relationship holds between the projection of a three-dimensional object and the corresponding central section of its 3D Fourier transform. This suggests that reconstruction can be achieved by "filling" the 3D Fourier space with data on 2D central planes that are derived from the projections by 2D Fourier transformation (Fig. 5.2).

More rigorously, the principal possibility of reconstructing a 3D object from its projections follows from Radon's (1917) quite general theory which has as its subject "the determination of functions through their integrals over certain manifolds." The parallel projection geometry we use to describe the image formation in the transmission electron microscope [as modeled in Eq. (5.2)] is a special case of Radon's theory where the integrals are performed along parallel lines [see the integral for the 3D case, Eq. (5.43)]. According to Radon, *an object can be reconstructed uniquely from its line projections when all of its line projections are known.*

Taken literally, this theorem is rather useless because it does not address the questions of how to reconstruct the object from a limited number of experimental (noisy) projections to a finite resolution and if for a limited number of projections such a reconstruction would be unique. The effect of restricting the reconstruction problem to finite resolution can be understood by considering the projection theorem—the fact that each projection furnishes one central section of the object's Fourier transform

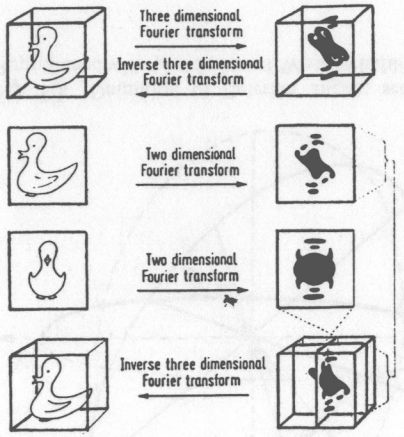


Fig. 5.2. Illustration of the projection theorem and its use in 3D reconstruction. From Lake (1971). Reproduced with permission of Academic Press Ltd.

—and taking into account the *boundedness* of the object (see Hoppe, 1969).

We make reference to Fig. 5.18 (later used to explain the related fact that adjacent projections are correlated up to a resolution that depends on the angular increment and the size of the object). A bounded object $o(\mathbf{r})$ can be described mathematically as the product of an unbounded function $\hat{o}(\mathbf{r})$ that coincides with the object inside of the object's perimeter and a shape function; i.e., a function that describes the object's shape and has the value 1 within the boundary of the object and 0 outside.

$$o(\mathbf{r}) = \hat{o}(\mathbf{r})s(r). \quad (5.4)$$

The Fourier transform of $o(\mathbf{r})$, which we seek to recover from samples supplied by projections, is

$$o(\mathbf{k}) = \hat{o}(\mathbf{k}) \odot S(\mathbf{k}), \quad (5.5)$$

i.e., every Fourier component of the unlimited object is surrounded by the *shape transform* $S(\mathbf{k}) = F\{s(\mathbf{r})\}$. For a smooth shape, the shape transform normally has a main maximum that occupies a region of Fourier space whose size is $1/D$ if D is the size of the object. This means that the Fourier transform $o(\mathbf{k})$ varies smoothly over this distance or, conversely, that measurements have only to be available on a grid in Fourier space with that spacing. As a consequence (see Fig. 5.18), roughly

$$N = \frac{\pi D}{d} \quad (5.6)$$

equispaced projections need be available to reconstruct an object with diameter D to a resolution $R = 1/d$ (Bracewell and Riddle, 1967; Crowther *et al.*, 1970). The same conclusion can be reached when one uses

a least-squares approach and formulates the reconstruction problem as the problem of finding the values of the Fourier transform on a finite polar grid from a finite number of experimental projections (Klug and Crowther, 1972).

In conclusion, we can state that, as a consequence of Radon's theorem and the boundedness of the object, *an object can be recovered to a given resolution from a finite number of projections, provided that these projections cover the angular space evenly*. For the time being, we leave this formulation general, but the problems related to gaps in angular coverage will surface throughout this chapter.

B. Projection Geometries

The purpose of this section is to define the relationship between the coordinate system of the projection and that of the molecule. Furthermore, using this formalism, we will define the two most important regular data collection geometries, *single-axis* and *conical*.

Let $\mathbf{r} = (x, y, z)^T$ be the fixed coordinate system of the molecule. By projecting the molecule along the direction $z^{(i)}$, defined by the three angles $\psi^{(i)}$, $\theta^{(i)}$, and $\phi^{(i)}$, we obtain the projection $p^{(i)}(x^{(i)}, y^{(i)})$. The transformation between the vectors in the coordinate system of the molecule and those in the coordinate system of the projection indexed i is expressed by three Eulerian rotations. In the convention used by Radermacher (1991),

$$\mathbf{r}^{(i)} = R\mathbf{r}, \tag{5.7}$$

with

$$R = R_{\psi_i} R_{\theta_i} R_{\phi_i}, \tag{5.8}$$

where

$$R_{\psi_i} = \begin{pmatrix} \cos\psi_i & \sin\psi_i & 0 \\ -\sin\psi_i & \cos\psi_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.9}$$

$$R_{\theta_i} = \begin{pmatrix} \cos\theta_i & 0 & -\sin\theta_i \\ 0 & 1 & 0 \\ \sin\theta_i & 0 & \cos\theta_i \end{pmatrix} \tag{5.10}$$

$$R_{\phi_i} = \begin{pmatrix} \cos\phi_i & \sin\phi_i & 0 \\ -\sin\phi_i & \cos\phi_i & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.11}$$

These rotations are defined as in classical mechanics and can be understood by reference to the sketch in Fig. 5.3: first the molecule is rotated by the angle ϕ_i in *positive* direction around its z axis, then by the angle θ_i in *negative* direction around its new y axis, and finally by the angle ψ_i in *positive* direction around its new z axis. It is seen that the first two angles define the direction of projection, while the third Eulerian rotation amounts to a trivial rotation of the object around an axis perpendicular to the projection. One commonly associates the angle θ_i with the concept of "tilt," although the exact tilt direction must be first defined by the size of the first "azimuthal" angle ϕ_i.

The orientations of projections accessible in a given experiment are defined by technical constraints; these constraints are tied to the degrees of freedom of the tilt stage and to the way the molecules are distributed on the specimen grid. Referring to the geometry defined by these constraints, we speak of the *data collection geometry*.

The *regular single-axis tilt geometry* (Fig. 5.4a) is generated by

$$\psi_i = 0 \quad \text{and} \quad \phi_i = 0: \tag{5.12}$$

the molecule is tilted by θ_i in equal increments around the y axis and then projected along $z^{(i)}$.

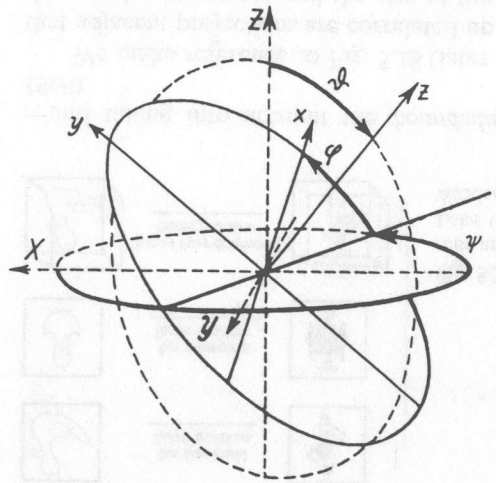


Fig. 5.3. Definition of Eulerian angles; see text for details. From Sommerfeld (1964). Reproduced with permission of Wissenschaftliche Verlagsgesellschaft Geest & Portig K. G.

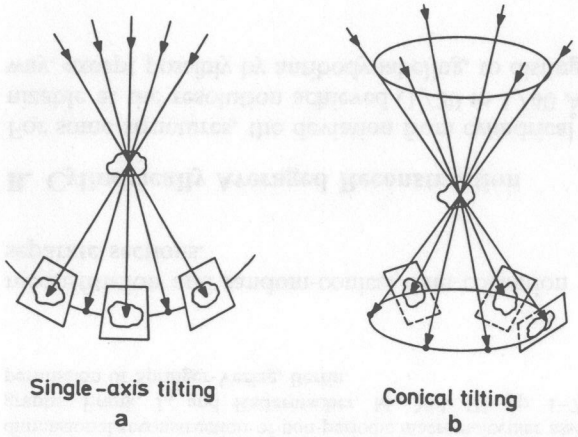


Fig. 5.4. Data collection by (a) single-axis and (b) conical tilting. From Radermacher (1980).

The *regular conical tilt geometry* (Fig. 5.4b) is generated by

$$\psi_i = 0 \quad \text{and} \quad \theta_i = \theta_0 = \text{constant}. \tag{5.13}$$

The molecule is first rotated around its z-axis by the "azimuthal angle" ϕ_i in equal increments and then tilted by θ_0 around the new y axis.

Finally, for later reference, the *random-conical geometry* is equivalent to the regular conical tilt geometry (without an explicit azimuthal rotation), except that the azimuth is randomly distributed in the azimuthal range $\{0, 2\pi\}$.

III. Rationales of Data Collection: Reconstruction Schemes

A. Introduction

In attempting to reconstruct a macromolecule from projections to a resolution of $1/30$ Å^{-1} or better, we must satisfy several mutually contradictory requirements:

(i) We need *many different* projections of the same structure to cover Fourier space as *evenly* as possible (this requirement often excludes the direct use of images of molecules showing preferred orientations, since the number of those is normally insufficient, and the angular coverage is far from even).

(ii) The total dose must not exceed 10 e$^-$/Å^2 (this excludes tomographic experiments of the type that Hoppe *et al.* (1974) introduced).

(iii) The reconstruction should be *representative* of the ensemble of macromolecules in the specimen (this excludes the use of automated tomography, by collecting all projections from a single particle while keeping the total dose low (cf. Dierksen *et al.*, 1992, 1993) unless a sizable number of such reconstructions are obtained which can be subsequently combined to form a statistically meaningful 3D average).

The three pathways that lead from a set of molecule projections to a statistically significant 3D image were summarized by Frank and Radermacher (1986) in a diagram (Fig. 5.5): following the first pathway, individual molecules are separately reconstructed from their (tomographic) tilt series, then their reconstructions are aligned in 3D and averaged (Fig. 5.5a). Following the second, molecule projections found in the micrograph are aligned, classified, and averaged by class. When a sufficient number of views are present, the molecule can be reconstructed from the class averages (Fig. 5.5b). The third possibility is to relate projections that vary widely in viewing direction to one another, so that an averaged 3D reconstruction can be directly computed (Fig. 5.5c).

(iv) Since stipulations (i) through (iii) imply that the projections have to be drawn from different "copies" (i.e., different realizations of the same structure) of the molecule, we need to establish the relative orientations of those molecules in a common frame of reference.

In other words, for a data collection and reconstruction scheme to be viable, it must be able to "index" projections reliably; i.e., it must be able to find their orientations. Of all the stipulations listed above, the last is perhaps the most difficult one to fulfill in practice. The reason that the random conical scheme, to be described below (Section III, D), has found wide popularity among several schemes proposed over the years, is that it solves the problem of finding the relative orientations of different projections unambiguously, by the use of two exposures of the same specimen field. Other methods, such as the method of angular reconstitution (van Heel, 1987b; Goncharov *et al.* 1987; Orlova and van Heel, 1994), have to find the angles a posteriori based on common lines.

In the following, I will first, for the sake for completeness, outline a method of data collection and reconstruction that draws from a single averaged projection and thus does not require orientation determination. Next, an important issue, the question of compatibility of projections, which determines the validity of all schemes that combine data from different particles, will be discussed. After that, the methods of angular

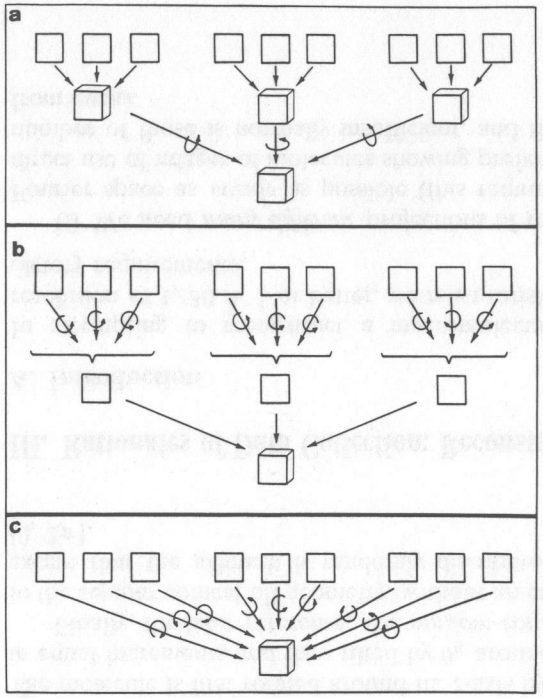


Fig. 5.5. Three principal ways of combining projection information into a statistically well-defined 3D structure. (a) Molecules are separately reconstructed from different projection sets (normally tilt series), and then the reconstructions are merged after appropriate orientation search. (b) A "naturally occurring" projection set is divided into classes of different views, an average is obtained for each class, the viewing direction is established for each average, and—if sufficient views are available—the molecule is reconstructed. (c) Projections are directly merged into a 3D reconstruction after their viewing directions have been found. From "Advanced Techniques in Biological Electron Microscopy." Three-dimensional reconstruction of non-periodic macromolecular assemblies from electron micrographs. Frank, J., and Radermacher, M., Vol. III, pp. 1–72 (1986). Reproduced with permission of Springer-Verlag, Berlin.

reconstitution and random-conical data collection will be outlined in two separate sections.

B. Cylindrically Averaged Reconstruction

For some structures, the deviation from cylindrical symmetry is not recognizable at the resolution achieved (1/20 to 1/40 $Å^{-1}$), and so there is no way, except possibly by antibody labeling, to distinguish the particle orientation (with respect to its long axis that is running parallel to the grid) from the appearance of its side views. The presentation of a cylindrically averaged reconstruction that is consistent with the observed views is the best one can do under these circumstances. The way from the projection to the 3D reconstruction is provided by the inversion of the Abel transform (Vest, 1974; Steven *et al.*, 1984):

Let us consider the two-dimensional case. The projection of a function $f(x, y)$ can be represented by the line integral

$$f_L(R, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x \cos\theta + y \sin\theta - R)dxdy, \quad (5.14)$$

which, considered as a function of the variables R and θ, is the Radon transform of $f(x, y)$. A choice of θ defines the direction of projection, and R defines the exact projection ray. Now if $f(x, y)$ is a slice of a cylindrically symmetric structure, it depends only on $r = (x^2 + y^2)^{1/2}$. In that case, Eq. (5.14) simplifies into the Abel transform:

$$f_L(R, \theta) = f_L(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r)\delta(x - R)dxdy$$

$$= 2\int_x^{\infty} \frac{f(r)rdr}{(x^2 - r^2)^{1/2}}. \quad (5.15)$$

Equation (5.15) can be inverted and solved for the unknown profile $f(r)$ by the use of the *inverse Abel transform*

$$f(r) = -\frac{1}{\pi}\int_r^{\infty} \frac{f_L(x)\,dx}{(x^2 - r^2)^{1/2}}. \quad (5.16)$$

The practical computation makes use of the fact that for a rotationally symmetric function, the Abel transform is equivalent to the Fourier transform of the Hankel transform.

This method has been used with success in the investigation of flagellar basal bodies both negatively stained (Stallmeyer *et al.*, 1989a, b) and frozen-hydrated (Sosinsky *et al.*, 1992; Francis *et al.*, 1994). Basal bodies are molecular motors effecting the rotation of flagella which are used for propulsion in water by certain bacteria. The 3D reconstructions of basal bodies from several mutants of two bacteria, *Salmonella* and *Caulobacter*, obtained by the Brandeis group over the course of the past few years, have much advanced our understanding of this fascinating "natural wheel" (Fig. 5.6). Of course, the detailed exploration of this structure will ultimately involve the more general methods discussed below, which are not based on the assumption of cylindrical symmetry. Better preparation methods and higher resolution are expected to make this improvement possible.

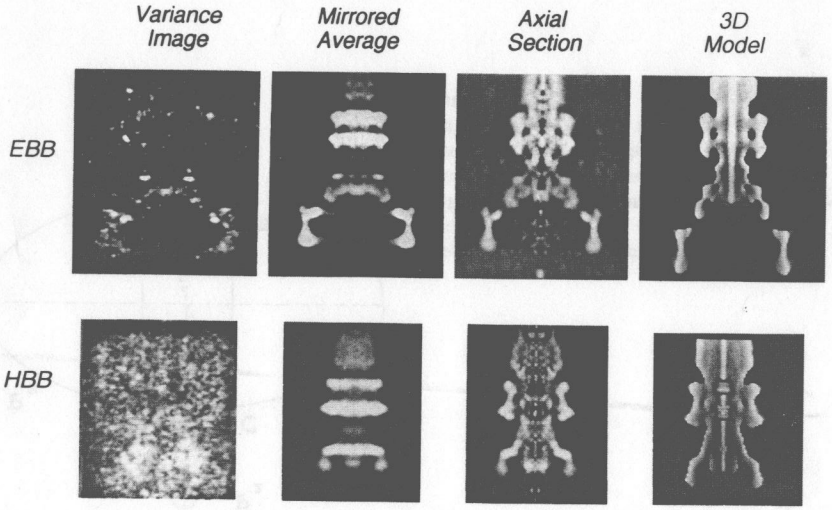


Fig. 5.6. Flagellar motor (basal body) of *Salmonella*, reconstructed from single-particle averages of frozen-hydrated specimen, assuming cylindrical symmetry. EBB, extended basal body. HBB, hook plus basal body, lacking the C ring complex and the switch protein. From Francis *et al.* (1994). Reproduced with permission of Academic Press Ltd.

C. Compatibility of Projections

When projections from different particles are combined in a three-dimensional reconstruction, the implicit assumption is that they represent different views of the same structure. If this assumption is incorrect—that is, if the structure is differently deformed in different particles—then the reconstruction will not produce a faithful 3D image of the macromolecule. Moreover, some methods of data collection and reconstruction determine the relative directions of projections by making use of mathematical relationships among them, which are fulfilled only when the structures they originate from are identical. However, macromolecules are often deformed because of an anisotropic environment: when prepared by negative staining and air-drying on a carbon grid, they are strongly flattened, down to as much as 50% of their original z dimension, in the direction normal to the plane of the grid. Even ice-embedment may not avoid a deformation entirely, because of the forces acting on a molecule at the air–water interface.

One important consideration in assessing the viability of different reconstruction schemes is therefore whether or not they mix projections from particles lying in different orientations. If they do, then some kind of

check is required to make sure that the structure is not deformed in different ways (see following). If they do not, and the molecule is deformed, then it is at least faithfully reconstructed, without resolution loss, in its unique deformed state.

It has been argued (e.g., van Heel, 1987) that by using solely 0° projections, i.e., projections perpendicular to the plane of the specimen grid, one essentially circumvents the problem of direction-dependent deformation, as this mainly affects the dimension of the particle perpendicular to the grid. Following this argument, a secondary effect could be expected to *increase* the width for each view, leading to a reconstruction that would render the macromolecule in a uniformly expanded form. As yet, this argument has not been tested. It would appear that specimens will vary widely in their behavior and that the degree of expansion may be orientation-dependent as well.

Conservation of the 3D shape of the molecule on the specimen grid can be checked by an *interconversion experiment*. Such experiments play an important role in visual model building, i.e., in attempts to build a physical model intuitively, by assigning angles to the different views, and shaping a malleable material so that the model complies with the observed views. The experiment is designed to establish an angular relationship between two particles presenting different views, A and B: on tilting, the particle appearing in view A changes its appearance into A′, and the one appearing in view B into B′. The experiment tests the hypothesis that the two particles are in fact identical but lie in different orientations. In that case, it should be possible to find a tilt angle α, around an appropriate axis, that renders A′ and B identical. Inverse tilt around that axis by an angle $-\alpha$ should also render B′ and A identical.

An example for a successful interconversion experiment is the study of Stoops *et al.* (1991) who found that the two prominent views of negatively stained α_2-macroglobulin, the "lip" and "padlock" views, interconvert for a tilt angle of 45° around the long axis of the molecule. Numerous interconversion experiments were also done, in an earlier phase of ribosome morphology research, to relate the views of ribosomes and their subunits to one another (e.g., Wabl *et al.*, 1973; Leonard and Lake, 1979) so that the 3D shape could be inferred.

D. Relating Projections to One Another Using Common Lines

Methods designed to relate different projections of a structure to one another make use of the *common lines* (Crowther *et al.*, 1970). These are lines along which, according to the projection theorem, the Fourier transforms of the projections should be identical in the absence of noise. The

common lines concept is important in several approaches to electron microscopic reconstruction and will be discussed again later (Section VIII). At present we make use of a simple model: let us represent two arbitrary projections of an object in 3D Fourier space by their associated central sections (Fig. 5.7). These intersect on a line through the origin, their *common line*. Suppose now that we do not know the relative orientation of these projections. We can then find their common line by "brute force," by comparing every (one-dimensional) section of one 2D Fourier transform with every one of the other transform. The comparison is done by cross-correlation, which is in Fourier space equivalent to the forming of the summed conjugate product. This product will assume a maximum when a match occurs. Another frequently used measure of the fidelity of match is the differential phase residual; see Section V, B, 2 in Chapter 3.

Once the common line is found, the (in-plane) rotations of the two central Fourier sections (and thereby, of the corresponding projections) are fixed. The two central sections can still move, in an unconstrained way, around the fixed common line which thereby acts as a "hinge." Obviously, a third projection and its central section, provided that it is nonplanar with either of the first two, will fix this movement and lead to a complete determination of relative angles among the three projections (apart from an ambiguity of handedness). Starting with this system of orientations, new projections are added by combining them with pairs of projections already placed. This, in essence, is the method of *angular reconstitution* (described by van Heel, 1987b; Goncharov *et al.*, 1987; Orlova and van Heel, 1994).

In practice, the common line search is performed in real space with the help of the so-called *sinogram*; this is a data table that contains in its rows the 1D projection of a 2D image (in our case, of a 2D projection)

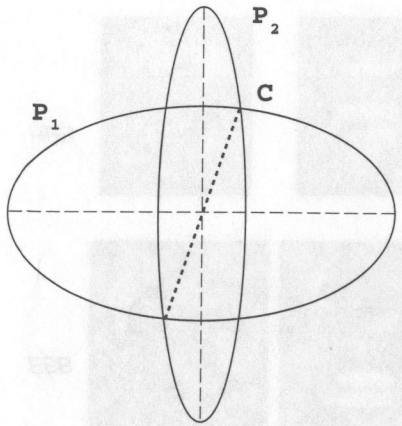


Fig. 5.7. The principle of common lines. Two projections of the same object, represented by central sections of the object's 3D Fourier transform, intersect each other along a common line through the origin. Along that line, their Fourier transforms must be identical.

exhaustively computed for all angles. (Note that the sinogram can be understood as a discrete version of the two-dimensional Radon transform.) If the 2D projections originate from the same object, then there exists an angle for which their 1D projections are identical (or closely similar): in real space, the equivalent to the common line is the common 1D projection. The angle is found by comparing or correlating the sinograms of the two 2D projections (Vainshtein and Goncharov, 1986; van Heel, 1987b; Goncharov *et al.*, 1987).

Although elegant in concept, the common line (or common 1D projection) method of orienting "raw data" projections, and thus the method of angular reconstitution as proposed originally, is normally hampered by the low signal-to-noise ratio of the data. However, as we know, the signal-to-noise ratio can be dramatically improved by averaging, either over projections of molecules presenting the same view, or over symmetry-related projections. Examples of sinograms for molecule class averages are presented in Fig. 5.8. From these examples it is clear that the determination of the common 1D projection, and hence the determination of the relative angle between two projections represented by class averages, should be quite robust. On the other hand, the averaging of molecules within classes entails a resolution loss which will be reflected by the quality of the reconstruction. Only by adding an angular refinement step, to be described in Section VIII of this chapter, can the full resolution in the data be realized.

Reconstructions utilizing this concept, mostly applied to macromolecules with symmetries, have been reported by van Heel and co-workers (Schatz, 1992; Dube *et al.*, 1994; Schatz *et al.*, 1994). Figure 5.9 presents a model of worm hemoglobin obtained by Schatz (1992), partly making use of the sinogram-based angle assignments. The first full-sized article, describing the reconstruction of the calcium release channel from 3000 cryoprojections, has just appeared as this book is being completed (Serysheva *et al.*, 1995). The results indicate (see Section III in Chapter 6 on validation), especially because they closely match the model previously obtained independently by the method of random-conical reconstruction (Radermacher *et al.*, 1994a, b), that the angular reconstitution method is a new viable approach to 3D electron microscopy of macromolecules, especially those exhibiting symmetries. Since it is based on the presence of multiple views covering the angular space as evenly as possible, the method can be seen as complementary to the random-conical method of reconstruction which is based on a different situation: the presence of a few preferred views, or even a single one. Another reconstruction of this type, applied to scanning transmission electron microscope data from

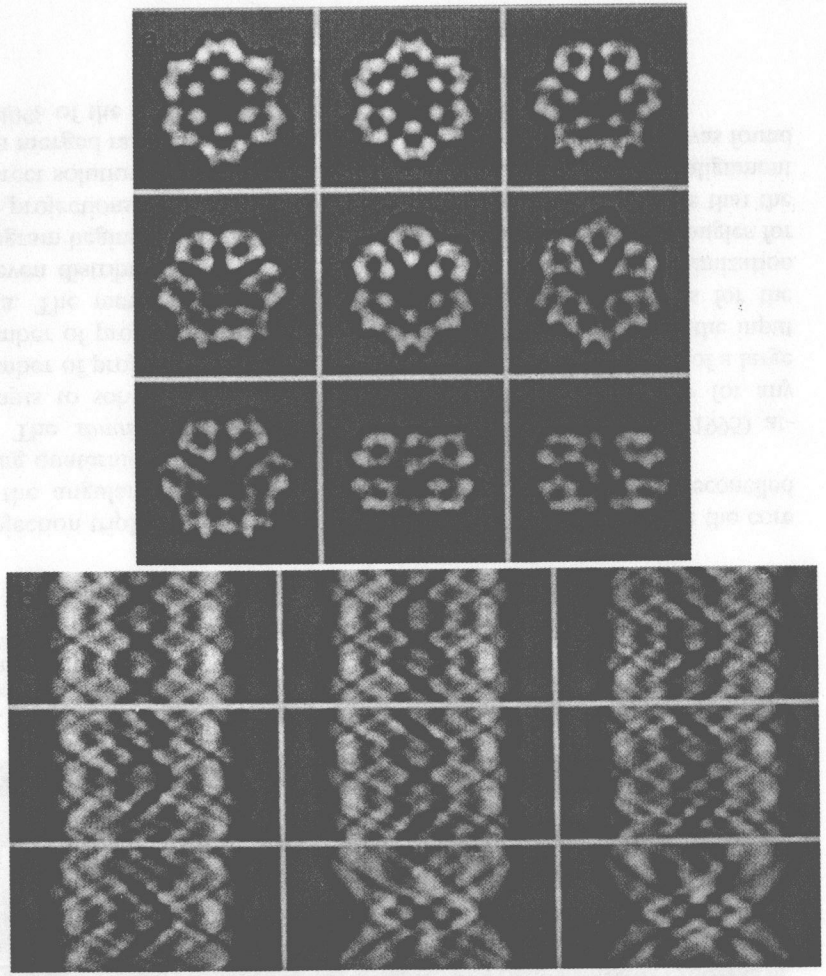


Fig. 5.8. Sinograms and their use in finding relative projection angles. (a) Class averages of worm hemoglobin showing the molecule in different views; (b) corresponding sinograms. Each horizontal line in the sinogram represents a 1D projection of the corresponding molecule image in a certain direction. The lines are ordered according to increasing angle of 1D projection, covering the full 360° range. A rotation of an image is reflected by a cyclical vertical shift of the corresponding sinogram. (c) Sinogram correlation functions (SINECORR for short). The SINECORR between sinograms of projections 1 and 2 is derived in the following way: the cross-correlation coefficient is computed between the first row of sinogram 1 and each row of sinogram 2, and the resulting values are placed into the first row of the SINECORR, and so on with the following rows of 1. The position of the maximum in the SINECORR indicates the angular relationship between the projections. Meaning of the

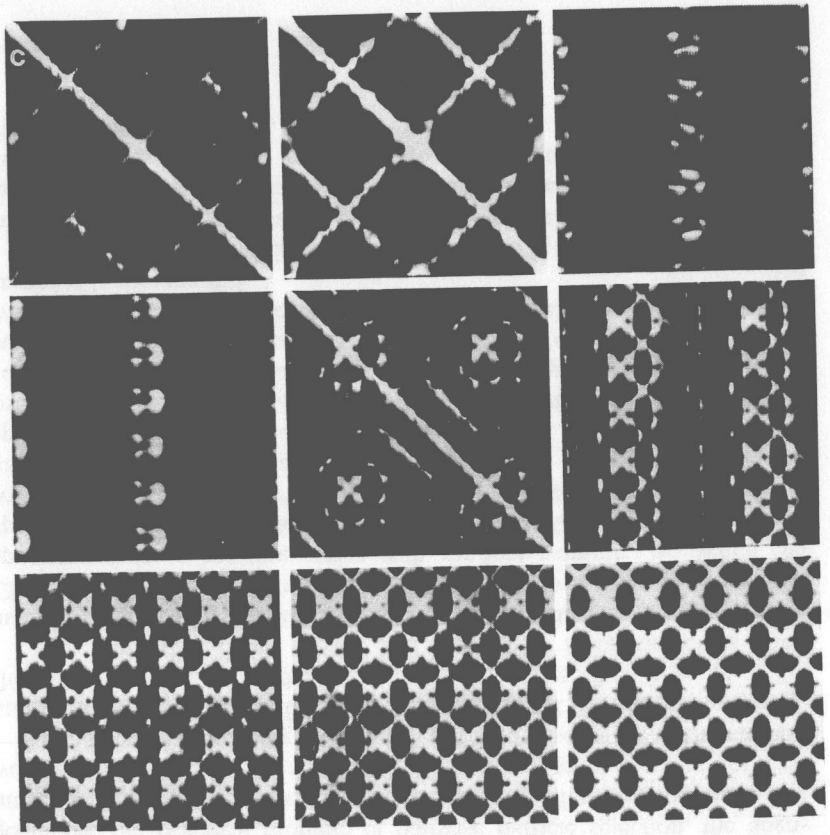


Fig. 5.8. (*continued*)

freeze–dried specimens of the signal sequence-binding protein SRP54, was presented by Czarnota *et al.* (1994).

Recognizing the noise sensitivity of van Heel's and Goncharov's method in its original form, which intended to recover relative orientations from *raw* data), Farrow and Ottensmeyer (1992; Ottensmeyer and Farrow, 1992) developed an extension of the technique. Solutions are found for many

---

panels from left to right, top to bottom: SINECORRs of 1 vs 1, 1 vs 2, 1 vs 3, 1 vs 4, 4 vs 4, 1 vs 9, 3 vs 9, vs 9, and 8 vs 8. Multiple maxima occur because of the sixfold symmetry of the molecule. From Schatz (1992). Reproduced with permission.

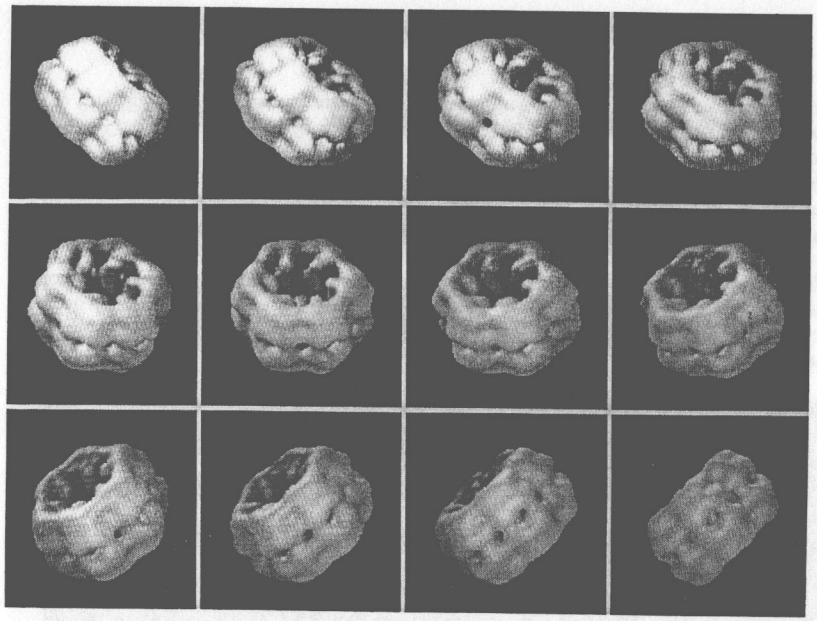


Fig. 5.9. Three-dimensional reconstruction of *Lumbricus terrestris* erythrocruorin embedded in ice from class averages shown in Fig. 5.8a. In part, the angles were assigned based on the technique of "angular reconstitution" (see also Schatz *et al.* 1994). From Schatz (1992). Reproduced with permission.

projection triplets using the common-lines triangulation that is at the core of the angular reconstitution technique, and the results are reconciled using quaternion mathematics (see Harauz, 1990).

The *simultaneous minimization technique* (Penczek *et al.*, 1995) attempts to solve the orientation search problem *simultaneously* for any number of projections greater then three. Concurrent processing of a large number of projections relaxes the requirement of high SNR for the input data. The method uses a discrepancy measure which accounts for the uneven distribution of common lines in Fourier space. The minimization program begins the search from an initial random assignment of angles for the projections. Penczek and co-workers were able to demonstrate that the correct solution (as it is known from the result of 3D projection alignment to a merged random-conical reconstruction; Frank *et al.*, 1995a) was found in 40% of the trials.

Another use of common lines is in "bootstrapping" methods, where new projections are matched to an existing reconstruction, or where the orientations of experimental projections are refined. This topic will be discussed further in Section VIII.

E. The Random-Conical Data Collection Method

The principle of this data collection scheme was first mentioned in the context of two-dimensional averaging of molecule projections (Frank *et al.*, 1978a) as an effective way for extending the single particle averaging into three dimensions. An explicit formulation and a discussion of the equivalent Fourier geometry was given by Frank and Goldfarb (1980). First attempts to implement the reconstruction technique led to a Fourier-based computer program that proved unwieldy (W. Goldfarb and J. Frank, unpublished, 1981). The first practical implementation of a reconstruction method making use of the random-conical data collection was achieved by Radermacher *et al.* (1986a, 1987a, b). For the implementation, numerous problems had to be solved, including the determination of the precise tilt geometry, the practical problem of pairwise particle selection, the alignment of tilted projections, the relative scaling of projection data, the weighting of projections in a generalized geometry, and—last, but not least—the massive bookkeeping required. A detailed description of the method and its implementation is found in Radermacher (1988). In the following, the different solutions to these problems will be described in some detail.

The method is based on the fact that single macromolecules often assume preferred orientations on the specimen grid (see Section I, E in Chapter 3). Any subset of molecules showing identical views in an untilted specimen form a rotation series with random azimuth, ϕ_i. When the specimen grid is tilted by a fixed angle, θ_0 (Fig. 5.10a, b), the above subset will appear in the micrograph as a conical projection series with random azimuths and θ_0 as cone angle (Fig. 5.10c). In the actual experiment, the specimen field is recorded twice: once tilted and once untilted (in this order). The first micrograph is used to extract the projections for the reconstruction. The purpose of the second micrograph is twofold: to (i) separate the particles according to their views ("classification") and (ii) within each subset (or class), determine the relative azimuths of all particles ("alignment").

The advantages of this scheme are evident: it allows the orientations of all molecules to be readily determined while allowing the dose to be kept to a minimum. Because of these advantages, the random conical reconstruction has come into widespread use (see the list of 3D reconstructions in Appendix 2).

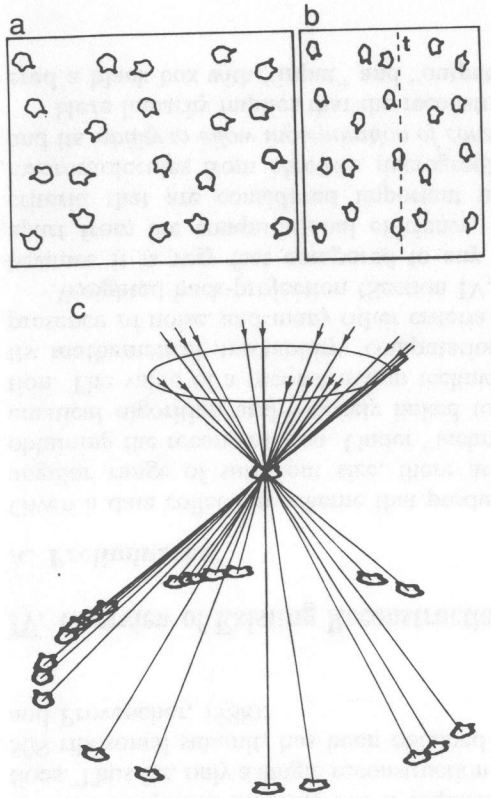


Fig. 5.10. Principle of the random-conical data collection. (a) untilted, (b) tilted field with molecule attached to the support in a preferred orientation, (c) equivalent projection geometry. From Radermacher *et al.* (1987b). Reproduced with permission of Blackwell Science Ltd., Oxford, from Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J., Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit. *J. Microsc.* **146**, 113–136.

There are some obvious limitations that restrict the resolution of the reconstruction: one is due to the fact that the observed "preferred orientation" in reality encompasses an entire orientation range (see Section VIII, D for the likely size of the angular range). Another, related limitation stems from the need to classify particles on the basis of their 0° appearance—a task which may have ambiguous results (see Section IV, K in Chapter 4). A third limitation has to do with the fact that the azimuthal angles (as well as the subsequent classification) are determined from the images of particles (at 0°) that have already been exposed to the electron beam and may have been damaged. All three limitations can be removed by the use of a refinement method according to which each projection is allowed to vary its orientation with respect to the entire data set (see Section VIII). However, the starting point is always a random conical reconstruction of the "basic" type outlined above.

The instructive drawing in Lanzavecchia *et al.* (1993) in (Fig. 5.11) shows the coverage of Fourier space afforded by the conical geometry. Since each projection is sampled on a square grid, its discrete Fourier transform is available within a square-shaped domain. The body formed by rotating an inclined square around its center resembles a yo-yo with a central cone spared out. Since the resolution of each projection is limited to a circular domain (unless anisotropic resolution-limiting effects such as astigmatism intervene, see Section II, B in Chapter 2), the coverage of the 3D Fourier transform by useful information is confined to a *sphere* contained within the perimeter of the yo-yo (not shown in Fig. 5.11).

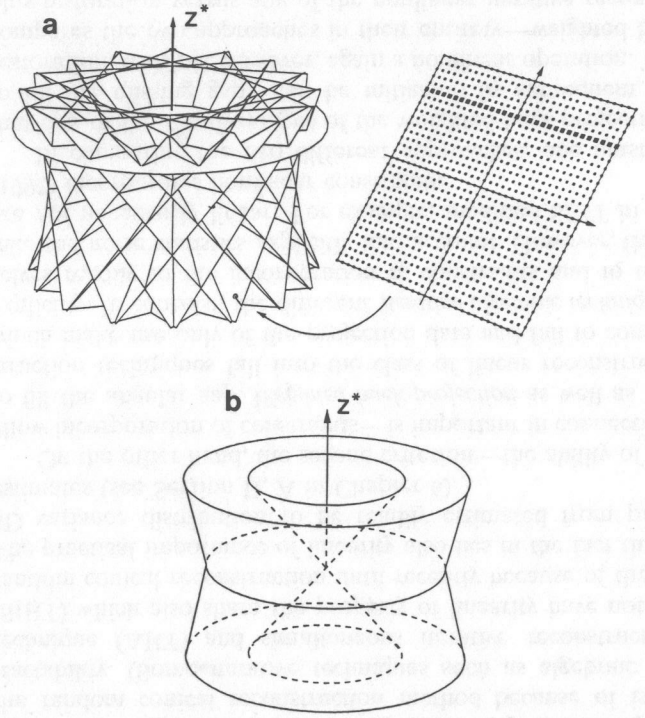


Fig. 5.11. Coverage of 3D Fourier space achieved by regular conical tilting. (a) Relationship between inclined Fourier plane representing a single projection and the 3D Fourier transform. (b) Yo-yo-shaped body (hollowed out by a double cone) covered by filling 3D Fourier space, assuming each plane contains information up to the sampling resolution. For random-conical data collection, the spacings between successive planes are irregular. Adapted from Lanzavecchia *et al.* (1993). Reproduced with permission of Blackwell Science Ltd., Oxford.

We will come back to the procedural details of the random-conical data collection and reconstruction after giving a general overview over reconstruction algorithms.

F. Reconstruction Schemes Based on Uniform Angular Coverage

For completeness, two reconstruction schemes that rely on a uniform coverage of the space of orientations should be mentioned. Both use spherical harmonics as a means to represent the object and its relationship to the input projections. The requirement of statistical uniformity and the choice of the rather involved mathematical representation have restricted the use of these schemes to model computations and few demonstrations with experimental data. The first scheme, proposed by Zvi Kam (1980), is based on a sophisticated statistical approach difficult to paraphrase here. The second scheme, introduced by Provencher and Vogel (1983; see also Provencher and Vogel, 1988; Vogel and Provencher, 1988), is designed to determine the relative orientations of the projections of a set of particles by a least squares method, but it requires approximate starting orientations. Thus far, only a single reconstruction of a nonsymmetric particle, the 50S ribosomal subunit, has been obtained with this latter method (Vogel and Provencher, 1988).

IV. Overview of Existing Reconstruction Techniques

A. Preliminaries

Given a data collection scheme that produces a set of projections over an angular range of sufficient size, there are still different *techniques* for obtaining the reconstruction. Under "technique" we understand the mathematical algorithm and—closely linked to it—its computational realization. The value of a reconstruction technique can be judged according to its mathematical tractability, computational efficiency, stability in the presence of noise, and many other criteria.

Weighted back-projection (Section IV, B) has gained wide popularity because it is very fast compared to any iterative techniques. However, apart from the computational efficiency, two—mutually contradictory—criteria that are considered important in the reconstruction of single macromolecules from electron micrographs are *linearity* of a technique and its *ability to allow incorporation of constraints*.

Here linearity implies that the reconstruction technique can be considered a black box with "input" and "output" channels and that the output

signal (the reconstruction) can be derived by linear superposition of *elementary output signals*, each of which is the response of the box to a delta-shaped input signal (projections of a point). In analogy to the point spread function, defined as the point response of an optical system, we speak of the "point spread function" of the combined system formed by the data collection and the subsequent reconstruction (see Radermacher, 1988). The linearity of the weighted back-projection technique (Section IV, B) has been important in the development and practical implementation of the random conical reconstruction method because of is mathematical tractability. (Some iterative techniques such as algebraic reconstruction technique (ART) and simultaneous iterative reconstruction technique (SIRT) which also share the property of linearity have not been used for random conical reconstruction until recently because of their slow speed.) The practical importance of linearity also lies in the fact that it allows the 3D variance distribution to be readily estimated from projection noise estimates (see Section II, A in Chapter 6).

On the other hand, the second criterion—the ability of a technique to allow incorporation of constraints—is important in connection with efforts to fill the angular gap. *Weighted back-projection* as well as Fourier reconstruction techniques fall into the class of linear reconstruction schemes, which make use only of the projection data and fail to consider the noise explicitly. In contrast, the different *iterative algebraic techniques* lend themselves readily to the incorporation of constraints and to techniques that take the noise statistics explicitly into account. However, these techniques are not necessarily linear. For example, modified SIRT in Penczek *et al.* (1992) incorporates nonlinear constraints.

In comparing the two different approaches, one must bear in mind that one of the disadvantages of the weighted back-projection—its failure to fill the missing gap—can be mitigated by subsequent application of restoration, which is, however, again a nonlinear operation. Thus when one compares the two approaches in their entirety—weighted back-projection plus restoration versus any of the nonlinear iterative reconstruction techniques—the importance of the linearity stipulation is somewhat weakened by its eventual compromise.

B. Weighted Back-Projection

Back-projection is an operation that is the inverse to projection: while the projection operation produces a 2D image of the 3D object, back-projection "smears out" a 2D image into a 3D body ("back-projection body," see Hoppe *et al.*, 1986) by translation into the direction normal to the plane of the image (Fig. 5.12). The topic of modified back-projection,

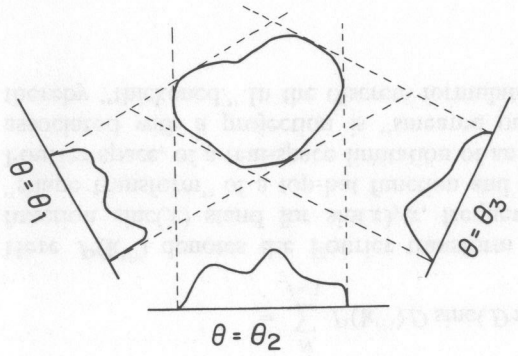


Fig. 5.12. Illustration of the back-projection method of 3D reconstruction. The density distribution across a projection is "smeared out" in the original direction of projection, forming a "back-projection body". Summation of these back-projection bodies generated for all projection yields an approximation to the object. For reasons that become clear from an analysis of the problem in Fourier space, the resulting reconstruction is predominated by low-spatial frequency terms. This problem is solved by Fourier weighting of the projections prior to the back-projection step. From Frank *et al.* (1985). Reproduced with permission of van Nostrand–Reinhold, New York.

as it is applied to the reconstruction of single particles, has been systematically presented by Radermacher (1988, 1991, 1992), and some of this work will be paraphrased here.

Let us consider a set of N projection into arbitrary angles. As a notational convention, we keep track of the different 2D coordinate systems of the projections by a superscript; thus, $p_i(\mathbf{r}^{(i)})$ is the ith projection, $\mathbf{r}^{(i)} = \{x^{(i)}, y^{(i)}\}^T$ are the coordinates in the ith projection plane, and $z^{(i)}$ is the coordinate perpendicular to that.

With this convention, the back-projection body belonging to the ith projection is

$$b_i(\mathbf{r}^{(i)}, z^{(i)}) = p_i(\mathbf{r}'^{(i)})t(z^{(i)}), \tag{5.17}$$

where $t(z)$ is a "top hat" function:

$$t(z) = \begin{bmatrix} 1 & \text{for } -D/2 \leqq z \leqq D/2 \\ 0 & \text{elsewhere} \end{bmatrix}. \tag{5.18}$$

Thus b_i is the result of translating the projection by D (a distance that should be chosen larger than the—anticipated—object diameter). As

more and more such back-projection bodies for different angles θ are added together, a crude reconstruction of the object is obtained:

$$\bar{\rho}(\mathbf{r}) = \sum_{i=1}^{N} b_i(\mathbf{r}^{(i)}, z^{(i)}), \tag{5.19}$$

with $\mathbf{r} = \{x, y, z\}$ being the coordinate system of the object.

The reason why such a reconstruction is crude is found by an analysis of the back-projection summation in Fourier space: it essentially corresponds to a simple filling of Fourier space by adding the central sections associated with the projections. It is immediately seen (Fig. 5.13) that the density of sampling points *decreases* with *increasing* spatial frequency, so that low spatial frequencies are overemphasized. As a result, the 3D image formed by back-projection appears like a blurred version of the object. Intuitively, it is clear that multiplication with a suitable radius-dependent weight might restore the correct balance in Fourier space. *Weighted back-*

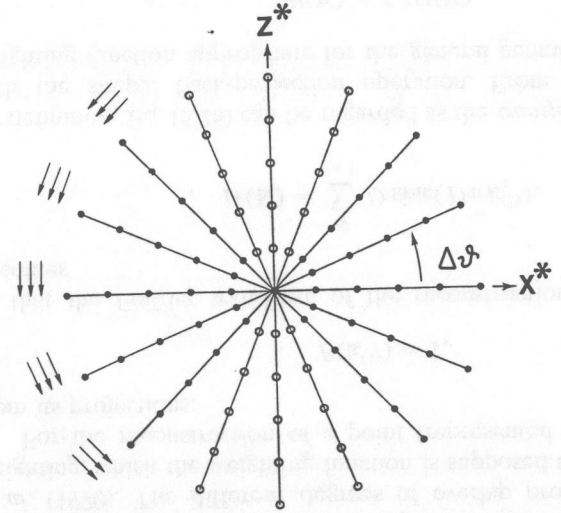


Fig. 5.13. Density of sampling points in Fourier space obtained by projections decreases with increasing spatial frequency. Although this is shown here for single-axis tilting, the same is obviously true for all other data collection geometries. From "Advanced Techniques in Biological Electron Microscopy." Three-dimensional reconstruction of non-periodic macromolecular assemblies from electron micrographs. Frank, J., and Radermacher, M., Vol. III, pp. 1–72 (1986). Reproduced with permission of Springer-Verlag, Berlin.

projection makes use of a weighting function $W_3(\mathbf{k})$ tailored to the angular distribution of projections:

$$\rho(\mathbf{r}) = F^{-1}\{W_3(\mathbf{k})F\{\bar{p}(\mathbf{r})\}\} \tag{5.20}$$

or, equivalently, by application of a *two-dimensional* weighting function to the projections,

$$p'(\mathbf{r}) = F^{-1}\{W_2(\mathbf{k})F\{p(\mathbf{r})\}\}. \tag{5.21}$$

The weighted back-projection algorithm makes it possible to design weighting functions for arbitrary projection geometries, and, specifically, to deal with the random azimuths encountered in the random-conical data collection (Radermacher *et al.*, 1986a, 1987a, b). In the following, we must distinguish between the coordinates affixed to the object (denoted by uppercase letters) and those affixed to the individual projections (lowercase letters). Similarly, we need to distinguish between the 2D Fourier coordinates $\mathbf{k}^{(i)} = \{k_x^{(i)}, k_x^{(i)}\}$ of the ith projection and the 3D Fourier coordinates of the object, $\mathbf{K} = \{K_x, K_y, K_z\}$. X, Y, Z are object coordinates, with z being the coordinate perpendicular to the specimen plane. $\mathbf{R} = \{X^{(i)}, y^{(i)}, z^{(i)}\}$ are the coordinates of the ith projection body.

The transformations from the coordinate system of the object to that of the ith projection (azimuth ϕ_i, tilt angle θ_i) is defined as follows:

$$\mathbf{r}^{(i)} = R_y(\theta_i)R_z(\phi_i)\mathbf{R}, \tag{5.22}$$

where R_y, R_z represent rotations around the y axis and z axis, respectively.

The weighting function for arbitrary geometry is now derived by comparing the Fourier transform of the reconstruction resulting from back-projection, $F\{\bar{p}(\mathbf{R})\}$, with the Fourier transform of the original object (Radermacher, 1991):

$$\bar{F}\{\bar{p}(\mathbf{R})\} = \sum_{i=1}^{N} F\{b_i(\mathbf{r}^{(i)}, z^{(i)})\} \tag{5.23}$$

$$= \sum_{i=1}^{N} P(\mathbf{k}^{(i)})D\,\mathrm{sinc}(D\pi k_z^{(i)}). \tag{5.24}$$

Here $P(\mathbf{k}^{(i)})$ denotes the Fourier transform of the ith projection. The function $\mathrm{sinc}(x)$ stand for $\sin(x)/x$, frequently used because it is the "shape transform" of a top-hat function and thus describes the effect, in Fourier space, of a real-space limitation of an object. Each central section associated with a projection is "smeared out" in the k_x-direction and thereby "thickened." In the discrete formulation, each Fourier coefficient

of such a central section is spread out and modulated in the direction normal to the section plane. This is in direct analogy to the continuous "spikes" associated with the reflections in the transform of a thin crystal (Amos *et al.*, 1982). However, in contrast to the sparse sampling of the Fourier transform by the reciprocal lattice, in the case of the crystal, we now have a sampling that is at least in principle continuous.

It is at once clear that the degree of overlap between adjacent "thick central sections" is dependent on the spatial frequency radius. The exact radius beyond which there is no overlap is the resolution limit of Crowther *et al.* (1970). The different degrees of overlap produce an imbalanced weighting, which the weighting function is supposed to overcome.

For the reconstruction of a point (represented by a delta function) from its projections,

$$P(\mathbf{k}^{(i)}) = 1, \tag{5.25}$$

so that the Fourier transform of the reconstruction by back-projection becomes

$$H(\mathbf{K}) = \sum_{i=1}^{N} D\,\mathrm{sinc}(D\pi k_z^{(i)}). \tag{5.26}$$

By definition, Eq. (5.26) can be regarded as the *transfer function* associated with the simple back-projection operation. From this expression the weighting function appropriate for the general geometry can be found as

$$W(\mathbf{K}) = 1/H(\mathbf{K}). \tag{5.27}$$

From this *weighting function for general geometries*, any weighting functions for special geometries can be derived (Radermacher, 1991). Specifically, the case of *single axis tilting* and *regular angular intervals* yields the well-known "r*-weighting" (in our nomenclature, $|\mathbf{K}|$-weighting) of Gilbert (1972).

In constructing the weighting function $W(\mathbf{K})$ according to Eq. (5.27), the regions where $H(\mathbf{K}) \to 0$ require special attention as they lead to singularities. In practice, to avoid artifacts, Radermacher *et al.* (1987a) found it sufficient to impose a threshold on the weighting function,

$$W(\mathbf{K}) \leqq 1.66; \tag{5.28}$$

i.e., to replace $W(\mathbf{K})$ by 1.66 wherever $1/H(\mathbf{K})$ exceeds that limit. In principle, a more accurate treatment can be conceived that would take the spectral behavior of the noise explicitly into account.

Note that $W(\mathbf{K})$ as derived here is a 3D function, and can be used directly in 3D Fourier space to obtain the corrected reconstruction. In practice, its central sections $w(\mathbf{k}^{(i)})$ are frequently used instead and applied to the 2D Fourier transforms of the projections. It should be noted that both ways of weighting are mathematically, but not necessarily numerically, equivalent.

Radermacher (1992) discussed reconstruction artifacts caused by approximation of Eq. (5.27), and has recommended replacing the sinc function by a Gaussian function in Eq. (5.24), tantamount to the assumption of a "soft" Gaussian-shaped object limitation.

C. Fourier Methods

Fourier approaches to reconstruction [see, for instance, Radermacher's (1992a) brief overview] utilize the projection theorem directly and regard the Fourier components of the projections as samples of the 3D transform to be determined. In most case, the positions of these samples do not coincide with the regular three-dimensional Fourier grid. This situation leads to a complicated interpolation problem, which can be stated as follows: given a number of measurements in Fourier space at arbitrary points not lying on the sampling grid, what set of Fourier components on the sampling grid are consistent with these measurements? The key to this problem lies in the fact that the object is of finite dimensions (Hoppe, 1969); because of this, the arbitrary measurements are related to those on the grid by the Whittaker–Shannon interpolation (Hoppe, 1969; Crowther *et al.*, 1970; Radermacher, 1992a; Lanzavecchia *et al.*, 1993; Lanzavecchia and Bellon; 1994).

Following Radermacher's account, we consider the unknown object bounded by a rectangular box with side lengths a, b, and c. In that case, its 3D Fourier transform is completely determined when all samples on the regular 3D sampling grid with grid size $(1/a, 1/b, 1/c)$ are known. We index them as $F_{hkl} = F(h/a, k/b, l/c)$. For an arbitrary position (denoted by the coordinate triple x^*, y^*, z^*) *not* lying on this grid, the Whittaker–Shannon theorem yields the relationship

$$F(x^*, y^*, z^*) = \sum_h \sum_k \sum_l F_{hkl} \frac{\sin \pi(ax^* - h)}{\pi(ax^* - h)} \times \frac{\sin \pi(by^* - k)}{\pi(by^* - k)} \frac{\sin \pi(cz^* - l)}{\pi(cz^* - l)}. \quad (5.29)$$

[The three terms whose product from the coefficient of F_{hkl} are again sinc functions.] This relationship tells us how to compute the value of the Fourier transform at an arbitrary point from those on the regular grid, but the problem we wish to solve is exactly the opposite: how to compute the values of the Fourier transform on every point h, k, l of the grid from a given set of measurements at arbitrary positions $\{x_j^*, y_j^*, z_j^*; j = 1 \ldots J\}$, as furnished by the projections. By writing Eq. (5.29) for each of these measurements, we create a system of M equations for $H * K * L$ unknown Fourier coefficients on the regular grid. The matrix C representing the equation system has the general element

$$C_{jhkl} = \frac{\sin \pi(ax_j^* - h) \sin \pi(by_j^* - k) \sin \pi(cz_j^* - l)}{\pi(ax_j^* - h)\pi(by_j^* - k)\pi(cz_j^* - l)}. \quad (5.30)$$

To solve this problem, we must solve the resulting equation system as follows:

$$F_{hkl} = \sum_j F(x_j^*, y_j^*, z_j^*) C_{jhkl}^{-1}, \quad (5.31)$$

where C_{jhkl}^{-1} are the elements of the matrix that is the inverse to C. It is obvious that this approach is infeasible because of the large number of terms. Basically, this intractability is the result of the fact that at any point that does not coincide with a regular Fourier grid point, the Fourier transform receives contributions from sinc functions centered on *every single grid point*.

Remedies designed to make the Fourier approach numerically feasible have been discussed by Lanzavecchia *et al.* (1993) and Lanzavecchia and Bellon (1994). These authors use the so-called "moving window" method to curb the number of sinc functions contributing in the interpolation, and thus obtain an overall computational speed that is considerable faster than the efficient weighted back-projection technique. A demonstration with experimental data—albeit with evenly distributed projections—indicated that the results of the two techniques were virtually identical. Other remedies are to truncate the sinc functions or to use different (e.g., triangular) interpolation functions.

D. Iterative Algebraic Reconstruction Methods

In the discrete representation, the relationship between the object and the set of projections can be formulated by a set of algebraic equations (Crowther *et al.*, 1970; Gordon *et al.*, 1970). For the parallel projection

geometry, the jth sample of projection i is obtained by summing the object ρ along parallel rays (indexed j) defined by the projection direction θ_i. The object is a continuous density function, represented by samples ρ_k on a regular grid, along with some rule (interpolation rule) for how to obtain the values on points not falling on the grid from those lying on the grid. Consequently, the discrete points of the object contribute to the projection rays according to weights $w_{jk}^{(i)}$ that reflect the angle of projection and the particular interpolation rule:

$$p_j^{(i)} = \sum w_{jk}^{(i)} \rho_k . \tag{5.32}$$

With a sufficient number of projections at different angles, Eq. (5.32) could be formally solved by matrix inversion, as pointed out by Crowther *et al.* (1970), but the number of unknowns is too large to make this approach feasible. Least-square, pseudoinverse methods (see Carazo, 1992) involve the inversion of a matrix whose dimensionally is given by the number of independent measurements, still a large number but somewhat closer to being manageable for the 2D case (see Zhang, 1992). There is, however, an entire class of reconstruction algorithms based on an approach to estimate the solution to Eq. (5.32) iteratively. The principle of these algorithms is that they start from an original estimate $\rho_k^{(0)}$ and compute its projection $\hat{p}_j^{(i)}$ following Eq. (5.32). The discrepancy between the actually observed projection $p_j^{(i)}$ and the "trial projection" $\hat{p}_j^{(i)}$ can now be used to modify each sample of the estimate $\rho_k^{(0)}$, giving a new estimate $\rho_k^{(1)}$, etc. In the algebraic reconstruction technique (ART) [proposed by Gordon *et al.*, 1970, but essentially identical with Karczemarz' (1937) algorithm for approximating the solutions of linear equations], the discrepancy is subtracted from the object estimate along the projection rays in each step, so that perfect agreement is achieved for the particular projection direction considered. In the simultaneous iterative reconstruction technique (SIRT) (proposed by Gilbert (1972)), the discrepancies of all projections are simultaneously corrected. For an exhaustive description of these and other iterative techniques, the reader is referred to Herman (1980).

Iterative methods have the advantage over the other approaches to 3D reconstruction that they are quite flexible, allowing constraints and statistical considerations to be introduced into the reconstruction process (e.g., Penczek *et al.*, 1992). They have the disadvantage of much larger computational expense than the weighted back-projection method. The use of nonlinear constraints (e.g., prescribed value range) introduces another disadvantage: the reconstruction process is no longer linear, making its characterization by a point spread function (see Section IV, B) or the 3D variance estimation by projection variance back-projection (Section II, A in Chapter 6) impossible to achieve.

V. The Random-Conical Reconstruction Scheme in Practice

A. Overview

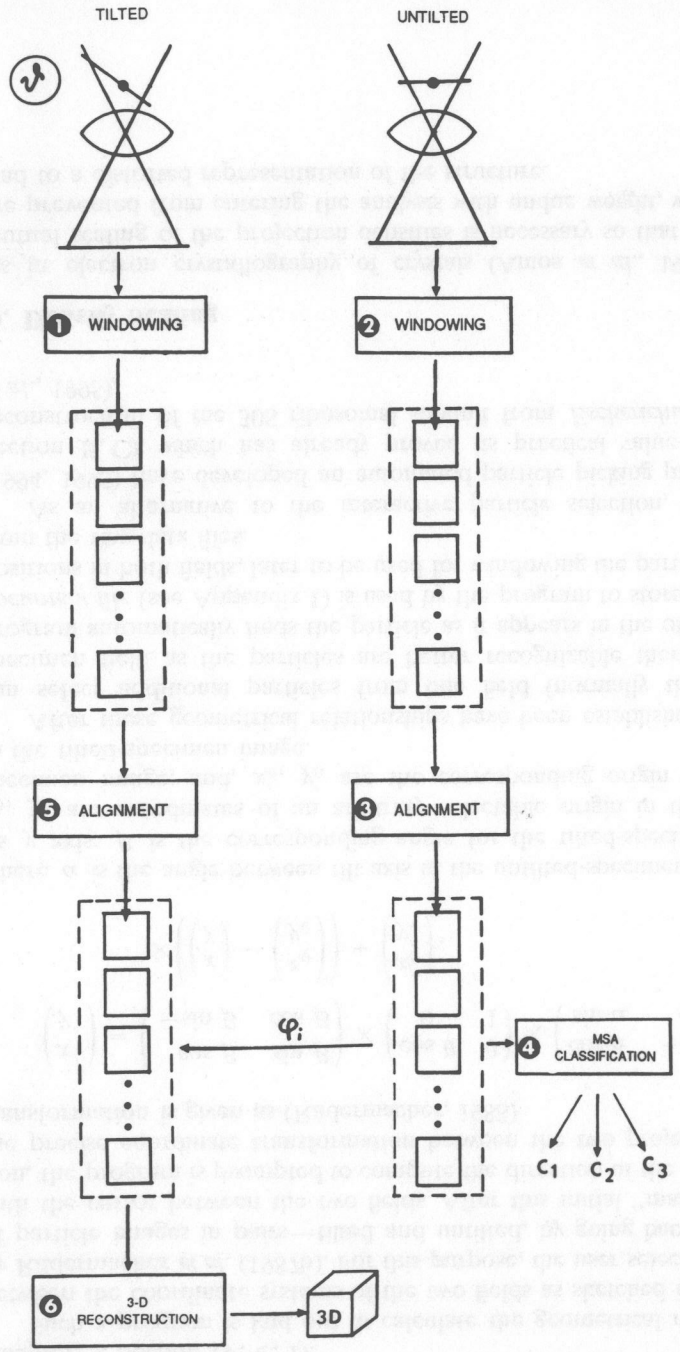
The concept of the random-conical data collection was introduced above (Section III, E). In the following section, all steps of the reconstruction scheme that makes use of this data collection method will be outlined. In this we will follow the detailed account given by Radermacher (1988), but with the addition of the multivariate statistical analysis (MSA)/classification step and some modifications that reflect changes in the procedures as they have developed since 1988.

To start with an overview of the procedure (Fig. 5.14), the particles are first selected ("windowed") simultaneously from the tilted and the untilted specimen field (steps 1 and 2, respectively), yielding two sets of images. Next, the untilted set is subjected to alignment (step 3), producing a set of "aligned" images. These are then analyzed using multivariate statistical analysis and classification (step 4), resulting in the rejection of certain particles and in the division of the data set into different classes according to particle view. From here on, the different classes are processed separately. For each class, the following steps are followed:

The tilted-particle images are sorted according to the azimuthal angle found in the alignment procedure. Once in correct order, they are aligned with respect to one another or with a common reference (step 5). After this, they may be used to obtain a 3D reconstruction (step 6). Thus, in the end, as many reconstructions are obtained as there are classes, the only requirement being that the class has to be large enough for the reconstruction to be meaningful.

B. Optical Diffraction Screening

The tilted-specimen micrograph covers a specimen field with a typical defocus range of 1.4 μm (at 50,000 × magnification and 50° tilt) perpendicular to the direction of the tilt axis. Because of the properties of the contrast transfer function (Section II, C in Chapter 2), useful imaging conditions require the defocus to be in the *range of underfocus* [i.e., $\Delta z < 0$ in Eq. (2.5) of Chapter 2]. In addition, the entire range must be restricted, to prevent a blurring of the reconstruction on account of the defocus variation that is equivalent to the effect of energy spread (Section I, C, 2 in Chapter 2). How much the range must be restricted, to a practical "defocus corridor" parallel to the tilt axis, depends on the resolution expected and can be inferred by reference to the transfer function charac-



teristics (Chapter 2, Section II, C, 4) and the sharp falloff in the "energy spread" envelope produced by the defocus spread.

When the micrograph shows a field with carbon film, the useful defocus range can be found by optical diffraction analysis. Details of this screening procedure have been described by Radermacher (1988). The selection aperture must be small enough so that the focus variation across the aperture is kept in limits. By probing different parts of the micrograph, the tilt axis direction is readily found as the direction in which the optical diffraction pattern remains unchanged. Perpendicular to that direction, the diffraction pattern changes most dramatically (see Fig. 5.15), and following this direction of steepest defocus change, the useful range must be established by comparison with the transfer function characteristics.

Electron microscopes equipped for spot scanning allow an automatic compensation for the defocus change perpendicular to the tilt axis (Zemlin, 1989b; Downing, 1992). For micrographs or data collected electronically from such microscopes, particles from the entire specimen field can be used for processing (Typke *et al.*, 1992).

Another way of including all data in the processing irrespective of their defocus is by keeping track of a spatial variable, in the course of selecting and storing the individual particle windows, that gives the particle position within the micrograph field in the direction perpendicular to the tilt axis. This variable can later be interpreted in terms of the effective local defocus, and used to make compensations or assign appropriate Fourier weighting functions in the reconstruction (see also Section IX).

C. Interactive Tilted/Untilted Particle Selection

Selection of particles from a pair of tilted- and untilted-specimen micrographs is a tedious task. A computer program for simultaneous interactive selection was first described in the review by Radermacher (1988). The two fields are displayed side by side on the screen of the workstation (Fig. 5.16) with a size reduction of 1 : 3 to 1 : 4. The size reduction makes it possible to select all particles from a micrograph field at once. It has the beneficial side effect that the particles stand out with enhanced contrast, since size

Fig. 5.14. Schematic diagram of the data flow in the random-conical reconstruction. Simultaneously, particles are selected from the micrograph (1) of the tilted specimen and that (2) of the untilted specimen. Those from the untilted field are aligned (3), resulting in azimuthal angles ϕ_i, and classified (4), resulting in the separation into classes C_1–C_3. Separately, for each of these classes, tilted-specimen projections are now aligned (5) and passed to the 3D reconstruction program (6). From Radermacher *et al.* (1987b). Reproduced with permission of Blackwell Science Ltd., Oxford, from Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J., Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit. *J. Microsc.* **146**, 113–136.

reduction with concurrent band limitation enhances the signal-to-noise ratio (SNR), provided (as is the case here) that the signal spectrum falls off faster then the noise spectrum with increasing spatial frequency (see Chapter, 3, Section IV, C, 1).

Such a program is laid out to calculate the geometrical relationships between the coordinate systems of the two fields as sketched in the paper by Radermacher *et al.* (1987b). For this purpose, the user selects a number of particle images in pairs—tilted and untilted, by going back and forth with the cursor between the two fields. After this initial "manual" selection, the program is prompted to compute the direction of the tilt axis and the precise coordinate transformation between the two projections. This transformation is given as (Radermacher, 1988)

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \times \begin{pmatrix} \cos\theta & 0 \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \times \left(\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \right) + \begin{pmatrix} x'_0 \\ y'_0 \end{pmatrix}, \tag{5.33}$$

where α is the angle between tilt axis in the untilted-specimen image and its y axis; β is the corresponding angle for the tilted-specimen image; x_0, y_0 are coordinates of an arbitrary selectable origin in the untilted-specimen image; and, x'_0, y'_0 are the corresponding origin coordinates in the tilted-specimen image.

After these geometrical relationships have been established, the user can select additional particles from one field (normally the untilted-specimen field as the particles are better recognizable there), and the program automatically finds the particle as it appears in the other field. A *document file* (see Appendix 1) is used by the program to store all particle positions in both fields, later to be used for windowing the particles images from the raw data files.

As an alternative to the interactive particle selection, Lata *et al.* (1994, 1995) have developed an automated particle picking program (see Section II, C), which has already proved its practical value in the 3D reconstruction of the 30S ribosomal subunit from *Escherichia coli* (Lata *et al.*, 1995).

D. Density Scaling

As in electron crystallography of crystals (Amos *et al.*, 1982) correct mutual scaling of the projection densities is necessary so that projections are prevented from entering the analysis with undue weight, which would lead to a distorted representation of the structure.

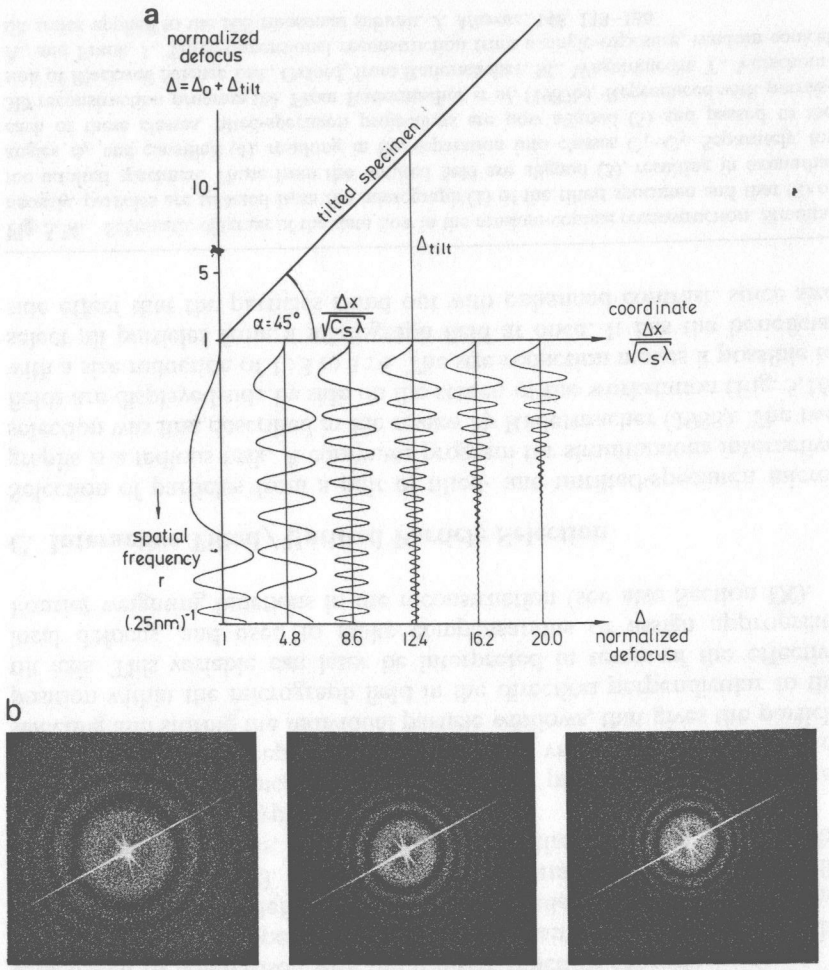
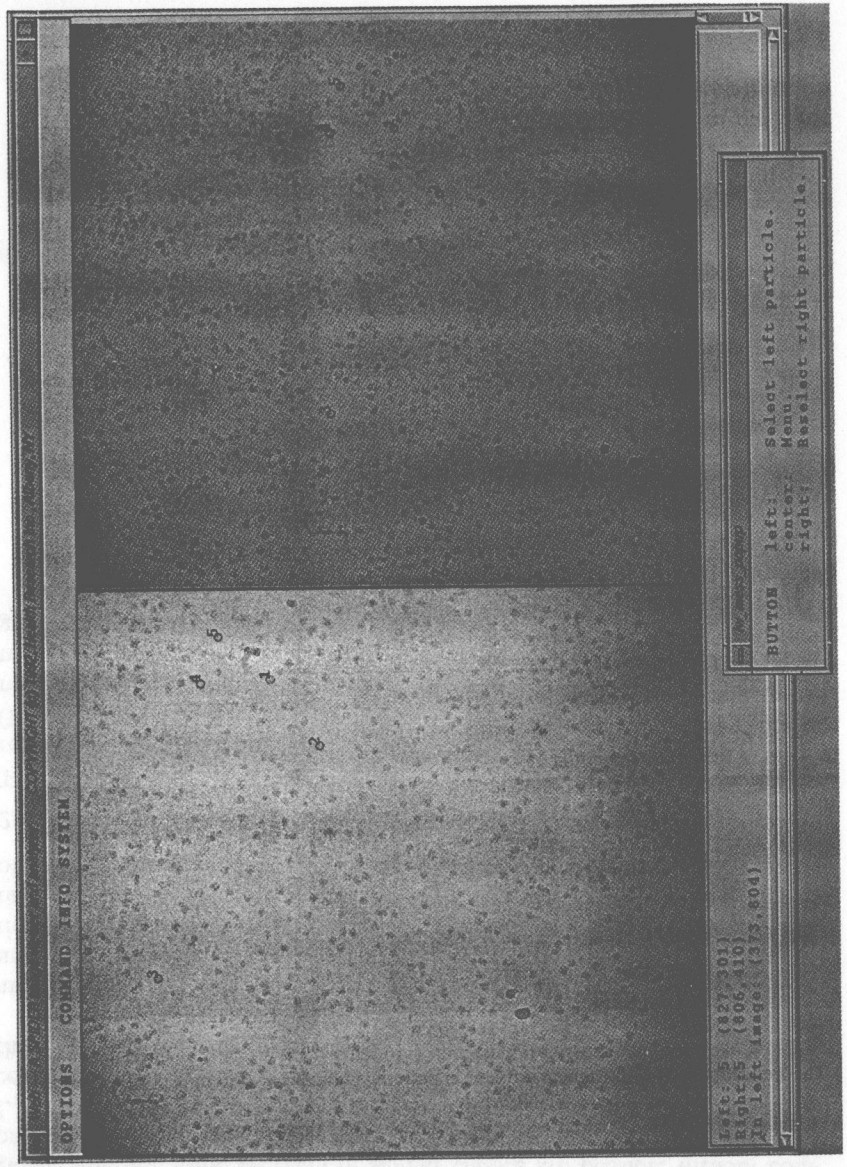


Fig. 5.15. Change of contrast transfer function across the micrograph of a tilted specimen. (a) Theoretical behavior; (b) optical diffraction patterns obtained at different positions of the area-selecting aperture of the optical diffractometer along a line perpendicular to the tilt axis. From Zemlin (1989b). Dynamic focussing for recording images from tilted samples in small-spot scanning with a transmission electron microscope. *J. Electron Microsc. Tech.*



Radermacher *et al.* (1987b) normalized the individual images representing tilted particles as follows: next to the particle (but staying away from the heavy stain accumulation if the preparation is with negative staining), the average density of a small reference field, $\langle D \rangle$, is calculated. The density values measured in the particle window are then rescaled according to the formula

$$D_i' = \frac{D_i - \langle D \rangle}{\langle D \rangle}. \tag{5.34}$$

Boisset *et al.* (1993) introduced another scaling procedure that makes explicit use of the statistical distribution of the background noise: it is assumed that the background noise surrounding the particle has the same statistical distribution throughout. Using a large portion of one of the tilted-specimen micrographs, a reference histogram is calculated. Subsequently the density histogram of the area around each particle is compared with the reference histogram, and the parameters a and b of a linear density transformation

$$D_i' = aD_i + b \tag{5.35}$$

are estimated. This transformation is then applied to the entire particle image, with the result that in the end all images entering the reconstruction have identical noise statistics. In fact, the match in statistics extends beyond the first and second moment.

E. Processing of Untilted-Particle Images

1. Alignment and Classification

Following the random-conical scheme, the untilted-specimen projections are first aligned and then classified. The alignment furnishes the azimuthal angle ϕ_i of the particle, which is needed to place the corresponding tilted-specimen projections into the conical geometry. The classification results in a division into L subsets of particles which are presumed to have

Fig. 5.16. Interactive particle selection using WEB. Two micrographs of the same field (left, untilted: right, tilted by 36°) are displayed side by side. Equivalent particle images are identified and numbered in both micrographs. Shown here is the beginning phase of the program where each particle has to be tracked down in both micrographs. After the initial phase, the parameters of the underlying coordinate transformation are known, and the program is able to identify the companion in the second micrograph for any particle selected in the first micrograph.

different orientations and ideally should be processed separately to give L different reconstructions. In practice, however, one chooses the classification cutoff level rather low, so that many classes are initially generated. By analyzing and comparing the class averages, e.g., by using the differential phase residual criterion (Section V, B, 2 in Chapter 3), it is possible to gauge whether some mergers of similar classes are possible without compromising resolution. This procedure leads to some small number of $L_1 < L$ "superclasses" which are fairly homogeneous on the one hand but contain sufficient numbers of particles on the other hand to proceed with three-dimensional reconstruction.

A substantial number of particles that fall in none of the superclasses are left out at this stage, their main fault being that their view is underrepresented, with a number that is insufficient for a 3D reconstruction. These particles have to "wait" till a later stage of the project when they can be merged with a well-defined reconstruction based on a merger of the superclass reconstructions, see Section VI below.

2. *Number of Particles Needed: Angular Histogram*

The number of projections required for computing a "self-standing" reconstruction is determined by the statistics of the angular coverage (Radermacher *et al.*, 1987b; see Fig. 5.17). If we require a self-standing reconstruction to be mathematically supported according to the conical-reconstruction resolution formula (Section V, I below), there should be no gap in the azimuthal distribution larger than

$$\Delta\theta_{\min} = 360/N = 360 \times d/(2\pi D \sin\theta_0), \tag{5.36}$$

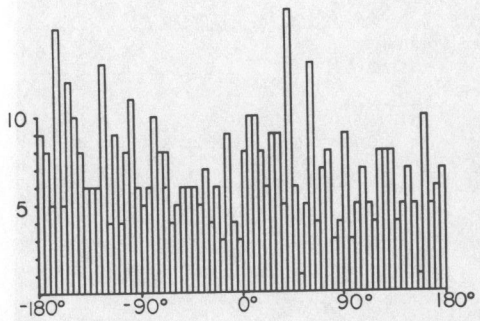


Fig. 5.17. Distribution of azimuthal angles of 50S ribosomal particles extracted from five pairs of micrographs. The angles were determined by alignment of particles showing the crown view as they appear in the untilted-specimen micrographs. From Radermacher *et al.* (1987b). Reproduced with permission of Blackwell Science Ltd., Oxford, from Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J., Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit. *J. Microsc.* **146**, 113–136.

where N is the number of equidistant projections in a regular conical series, $d = 1/R$ is the resolution distance aimed for, D is the diameter of the object, and θ_0 is the cone angle. To reconstruct the 50S ribosomal subunit of *E. coli* (D = 200 Å) to a resolution of $R = 1/30$ Å^{-1} (d = 30 Å) from a 50° tilted specimen (i.e., $\theta_0 = 40°$), the minimum angular increment works out to be $\Delta\theta_{\min} = 13°$. Because of the statistical fluctuations in a random coverage of the azimuthal range, the required number of projections is much larger then the minimum of $N = 360/13 \approx 28$. In practice, a set of 250 projections proved sufficient to cover the 360° range with the largest gap being 5° (Radermacher *et al.*, 1987b). Although there is no iron-clad rule, molecules in that size range (200 to 300 Å) and to that resolution (1/30 Å^{-1}) appear to require a minimum of 200 projections.

F. Processing of Tilted-Particle Images

1. *Alignment*

Reference to the Fourier description (Fig. 5.18) explains why an alignment of neighboring tilted-particle projections by cross-correlation is possible: because each Fourier component is surrounded by a "circle of influence," as a result of the boundedness of the object. It is the very reason that reconstruction from a finite number of projections is feasible.

There are four methods of alignment that have been tried at one stage or the other: (i) sequential alignment "along the circle" with cosine stretching (Radermacher *et al.*, 1987b), (ii) alignment to the corresponding 0° projection with cosine stretching (Radermacher, 1988; Caraza *et al.*, 1988), (iii) alignment to a perspectively distorted disk (Radermacher, 1988), and (iv) alignment to a disk or "blob" (Penczek *et al.*, 1992).

The reference to cosine stretching requires an explanation. According to Guckenberger's (1982) theory, the projection to be aligned to the untilted-particle projection (method (ii) above) must first be stretched by the factor $1/\cos(\theta_t)$ in the direction perpendicular to the tilt axis, where θ_t is the tilt angle. When this philosophy is applied to the alignment of two adjacent tilted-particle projections (method (i) above), both must be stretched by that factor prior to alignment. The cosine stretching procedure appears to work well for oblate objects, i.e., those objects that are more extended in the x and y directions than in the z-direction, as for instance the 50S ribosomal subunit in the negatively stained double layer preparation (Radermacher *et al.* , 1987a, b; Radermacher, 1988) The reason that it works well for such specimens is that the common signal includes the surrounding carbon film which is flat. For globular objects, such as the 70S ribosome embedded in ice (Frank *et al.*, 1991, 1995a, b;

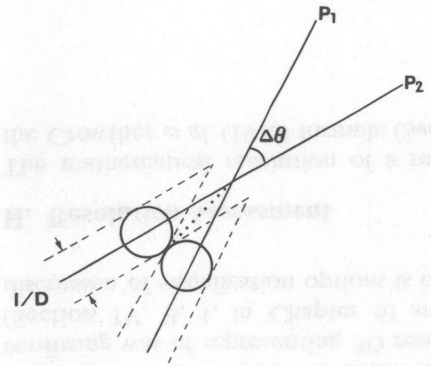


Fig. 5.18. Statistical dependence of Fourier components belonging to different projections. We consider the Fourier transform along central sections P_1, P_2 representing two projections of an object with diameter D. Each Fourier component is surrounded by a "circle of influence" with diameter $1/D$. Thus the central section is accompanied on both sides by a margin of influence, whose boundaries are indicated by the dashed lines. The diagram can be used to answer two interrelated questions: (1) to what resolution are two projections separated by $\Delta\theta$ correlated? (2) what is the minimum number of projections with equispaced orientations that are required to reconstruct the object to a resolution R without loss of information? From "Advanced Techniques in Biological Electron Microscopy." Three-dimensional reconstruction of non-periodic macromolecular assemblies from electron micrographs. Frank, J., and Radermacher, M., Vol. III, pp. 1–72 (1986). Reproduced with permission of Springer-Verlag, Berlin.

Penczek *et al.*, 1992, 1994), alignment to a nonstretched disk (blob) whose diameter is equal to the diameter of the particle leads to better results.

Sequential alignment of neighboring projections is errorprone because errors can accumulate in the course of several hundred alignments, as may be easily checked by computing the closure error (Radermacher *et al.*, 1987b). Radermacher and co-workers developed a fail-safe variant of the sequential method in which neighboring projections are first averaged in groups spanning an angle of 10°; these averaged tilt projections are aligned "along the circle", and, in the end, the individual projections are aligned to their corresponding group average. However, for cryospecimens with their decreased SNR, sequential alignment has largely been abandoned in favor of the alignment of each tilted-particle projection with the 0° projection or, as pointed out before, with a blob.

It may seem that the availability of 3D refinement methods (see Section VIII) has relaxed the requirement for precise tilted-particle projection alignment at the stage of the first reconstruction; on the other hand, it must be realized that the accuracy of the refinement is determined

by the quality of the reference reconstruction which is in turn critically dependent on the success of the initial alignment.

2. *Screening*

Screening of tilted-particle projections is a way of circumventing the degeneracy of 0° classification (Section IV, K in Chapter 4). In the first study of a molecule whose structure is unknown, it is advisable to verify that the 0° projection does not hide a mixture of two or more molecule orientations. The most important stipulation to apply is the continuity of shape in a series of angularly ordered projections. As members of a conical tilt series, projections must occur in a strict one-dimensional similarity order: for instance, the ordering of five closely spaced projections along the cone in the sequence A–B–C–D–E implies that the cross-correlation coefficients (denoted by symbol ⊛) are ranked in the following way:

$$A \circledast B > A \circledast C > A \circledast D > A \circledast E$$

If we look at projections forming an entire conical series, then the associated similarity pathway is a closed loop (see Section IV, H). In any short segment of the loop, we find a cross-correlation ranking of the type stated above as a local property. If all projections of the conical series are closely spaced, then the shape variation from one neighbor to the next is quasi-continuous. In practice this means that a gallery of tilted-particle projections presented in the sequence in which they are arranged on the cone should show smooth transitions, and the last one of the series should be similar to the first. The most sensitive test of shape continuity is a presentation of the entire projection series as a "movie": any discontinuity is immediately spotted by eye.

The correct ordering on a similarity pathway can also be monitored by multivariate statistical analysis. In the absence of noise, the factor map should show the projections ordered on a closed loop (Frank and van Heel, 1982b; van Heel, 1984a). In fact, the ordering is on a closed loop in a high-dimensional space R^J, and this closed loop appears in many different *projected versions in the factor maps. In the presence of noise, it proves difficult to visualize the similarity pathway (and thereby spot any outliers that lie off the path). One way out of this problem is to form "local" averages over short angular intervals, and apply MSA to these very robust manifestations of the projections (Frank *et al.*, 1986). A closed loop becomes then indeed visible. It can be used, in principle, to screen the original projections according to their distance, in factor space, from the averaged pathway. This method has not been pursued, however, and has been partly replaced by the 3D projection matching (Section VIII, B) and 3D Radon transform (Section VIII, C) methods.

Flip/flop ambiguities are normally resolved by classification of the 0° degree views, on the basis of the mirroring of the projected molecule shape (van Heel and Frank, 1981). However, a peculiar problem emerges when the shape of the molecule in projection is symmetric. In that case, there is no distinction between flip and flop projections—unless induced by one-sidedness of staining. Lambert *et al.* (1994a) had to deal with this problem when they processed images of the barrel-shaped chiton hemocyanin which sits on the grid with one of its (unequal) round faces. As these authors realized, the cylindric shape is unique in that it allows flip and flop orientations of the molecule to be sorted by MSA of the tilted-particle images. Perfect cylinders give uniformly rise to double-elliptic barrel projections, irrespective of the tilt direction. Any asymmetry in the z distribution of mass (in this case, the existence of a "crown" on one side of the molecule) leads to a difference in appearance between molecules tilted to the one side from those tilted to the other. This difference was clearly picked up by correspondence analysis, and the two different populations could be separated on this basis.

G. Reconstruction

After the screening step to verify that the tilted-particle projections follow one another in a reasonable sequence, the particle set is ready for 3D reconstruction. Both weighted back-projection (Radermacher *et al.*, 1987b) and a modification of SIRT (Penczek *et al.*, 1992) are being used. For the weighted back-projection, the weighting step is either performed individually on each projection (Radermacher *et al.*, 1987b) (implying the steps FT $\rightarrow$ Weighting $\rightarrow$ FT^{-1}) or, summarily, by applying a 3D weighting function on the volume obtained by simple back-projection (Hegerl *et al.*, 1991).

The reconstruction exists as a set of slices (see Fig. 5.19). Each slice is represented by a 2D image that gives the density distribution on a given z level, which is a multiple of the sampling step. The representation as a gallery of successive slices (the equivalent of a cartoon if the third dimension were the time) is both the most comprehensive and the most confusing way of representing 3D results. Instead, surface representations (Section IV, B, 1, in Chapter 6) are now in common use. A general discussion of visualization options is contained in Section V, 4.

H. Resolution Assessment

The mathematical resolution of a reconstruction is determined through the Crowther *et al.* (1970) formula (Section II, A). Corresponding formulae

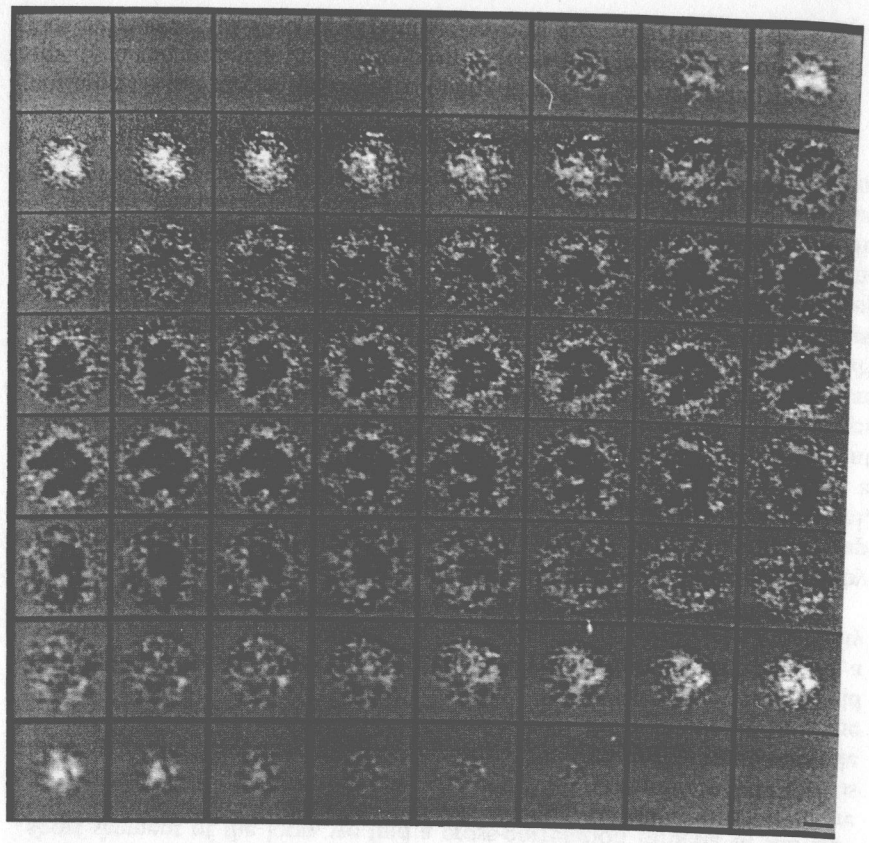


Fig. 5.19. The 50S ribosomal subunit, reconstructed from cryoimages, represented as a series of slices. From Radermacher (1994). Reproduced with permission of Elsevier Science, Amsterdam.

have been given by Radermacher (1991) for the conical projection geometry. Here we mention the result for an even number of projections:

$$d = 2\pi(D/N)\sin\theta_0 \qquad (5.37)$$

where $d = 1/R$ is the "resolution distance," i.e., the inverse of the resolution, D the object diameter, and θ_0 is the tilt of the specimen grid. Furthermore, for such a data collection geometry, the resolution is direc-

tion-dependent, and the above formula gives only the resolution in the directions perpendicular to the direction of the electron beam. In directions that form an angle oblique to those directions, the resolution is deteriorated. In the beam direction, the effect of the missing cone is strongest, and the resolution falls off by a factor of 1.58 (for $\theta_0 = 45°$) or 1.23 (for $\theta_0 = 60°$).

The account given thus far relates to the theoretical resolution expected from the data collection geometry. However, whether this resolution is actually realized is quite another matter. The reasons that it is normally not realized are manifold: the structure of the biological particle may not be defined to that level of resolution because of conformational variability; in stained preparations, the stain fluctuations and finite graininess limit the definition of small specimen features; and there are a number of electron optical effects (partial coherence, charging, specimen movement) that limit the transfer of information from the specimen to the image. Another important limitation is due to errors in the assignment of projection angles (see Section VIII, D).

For these reasons, the significant resolution of a reconstruction needs to be independently assessed. (Here the term "significant resolution" denotes the resolution up to which object-related features are represented in the 3D image.) The procedure is quite similar to the assessment of 2D resolution (Section V, B in Chapter 3), by computing two reconstructions from randomly drawn subsets of the projection set, and comparing these in Fourier space using differential phase residual (DPR) or Fourier ring correlation (FRC) criteria. However, in the 3D case, the summation in the defining formulas [Chapter 3, Eqs. (3.64) and (3.65)] now has to go over shells $|\mathbf{k}|$ = constant. This extension of the differential resolution criterion to three dimensions was first implemented (under the name of "Fourier shell correlation") by Harauz and van Heel (1986a) for the case of the FRC. Thus far, a resolution assessment analogous to the spectral signal-to-noise ratio (SSNR) has not been developed for the 3D reconstruction, but some considerations along these lines have been developed by Liu (1993).

Resolution assessment based on the comparison of experimental projections with projections "predicted" from the reconstruction is useful, but cannot replace the full 3D comparison mentioned above. One of the consistency tests for a random-conical reconstruction is the ability to predict the 0° degree projection from the reconstruction, a projection that does not enter the reconstruction procedure yet is available for comparison from the analysis of the 0° data. Mismatch of these two projections is an indication that something has gone wrong in the reconstruction; however, on the other hand, an excellent match, up to a resolution R, is *not* a

guarantee that such a resolution is realized in all directions. This is easy to see by invoking the projection theorem as applied to the conical projection geometry (see Fig. 5.10): provided that the azimuths ϕ_i are correct, the individual tilted planes representing tilted projections furnish correct data for the 0° plane, irrespective of their angle of tilt. It is therefore possible to have excellent resolution in directions defined by the equatorial plane, as evidenced by the comparison between the 0° projections, while the resolution might be severely restricted in all other directions. The reason that this might happen is that the actual tilt angles of the particles differ from the nominal tilt angle assumed (see Penczek *et al.*, 1994).

Another resolution test, used initially for assessing the results of 3D reconstruction, employs the pairwise comparison of selected slices, again using the two-dimensional DPR and FRC (e.g., Radermacher *et al.*, 1987b, Verschoor *et al.*, 1989; Boisset *et al.*, 1990b). Although such a test, in contrast to the 0° projection check mentioned above, *is* indeed sensitive to the accuracy of assignment of tilt angles to the particles, it still fails to give an overall assessment of resolution including all spatial directions, as only the Fourier shell measures can give.

Because of these shortcomings of 2D resolution tests in giving a fair assessment of 3D resolution, the use of DPR and FRC computed over shells is now common practice (e.g., Akey and Radermacher, 1993; Radermacher *et al.*, 1994b, Serysheva *et al.*, 1995). (Unfortunately, the two measures in use give widely different results for the same data set, with FRC being as a rule more optimistic than DPR; see Section V, B in Chapter 3. It is therefore a good practice to quote both in the publication of a 3D map.) However, there is as yet no convention on how to describe the direction dependence of the experimental resolution. Thus, the resolution sometimes relates to an average over the part of the shell within the measured region of 3D Fourier space; sometimes (e.g., Boisset *et al.*, 1993, 1995; Penczek *et al.*, 1994) it relates to the entire Fourier space without exclusion of the missing cone. It is clear that, as a rule, the latter figure gives a more pessimistic estimate than the former.

VI. Merging of Reconstructions

A. The Rationale of Merging

Each reconstruction shows the molecule in an orientation that is determined by the orientation of the molecule on the specimen grid (and by a trivial "in-plane" rotation angle whose choice is arbitrary but which also figures eventually in the determination of relative orientations). The goal

of filling the angular gap requires that several reconstructions with different orientations be combined, to form a "merged" reconstruction as the final result. The procedure to obtain the merged reconstruction is easily summarized in three steps (i) 3D orientation search, (ii) expression of the different projection sets in a common coordinate system, and (iii) reconstruction from the full projection set.

The premise of the merging is that the reconstructions based on molecules showing different 0° views represent the same molecule, without deformation. Only in that case will the different projection sets be consistent with a common object model. The validity of this assumption will be investigated in the next section. In each case, the merging must be justified, by applying a similarity measure: until such verification is achieved, the assumption that the particles reconstructed from different 0°-view sets have identical structure and conformation remains unproven. This extra scrutiny is required because molecules are potentially more variable in the single particle form than when ordered in crystals.

B. Preparation-Induced Deformations

Molecules prepared by negative staining and air-drying show evidence of flattening. Quantitative data on the degree of flattening are scarce; however, as an increasing number of comparisons between reconstructions of molecules negatively stained and embedded in ice become available, some good estimations can now be made. Another source of information is the comparison of molecules that have been reconstructed in two different orientations related to each other by a 90° rotation. In interpreting such data, it is important to take incomplete staining into account. Without using the sandwiching technique, some portions of the particle, pointing away from the grid, may "stick out" of the stain layer and thus be rendered invisible. On the other hand, the sandwiching may be responsible for an increase in flattening.

Boisset *et al.* (1990b) obtained two reconstructions of negatively stained, sandwiched *Androctonus australis* hemocyanin, drawing either from particles lying in the top or those in the side view, and reported a factor of 0.6 when comparing a particle dimension perpendicular to the specimen grid with the same dimension parallel to the grid. Since the flattening is accompanied by an increase in lateral dimensions, one can assume that the flattening is less severe than this factor might indicate. Another estimate, a factor of 0.65, comes from a comparison between two reconstructions of an *A. australis* hemocyanin–Fab complex, one obtained with negative staining (Boisset *et al.* 1993b) and the other using vitreous ice (Boisset *et al.*, 1994b; 1995).

Cejka *et al.* (1992) obtained a somewhat larger factor (12 nm/18 nm = 0.67) for the height (i.e., the dimension along the cylindric axis) of the negatively stained (unsandwiched) *Ophelia bicornis* hemoglobin. They were able to show that molecules embedded in aurothioglucose and frozen-hydrated have essentially the same height (18.9 nm) when reconstructed in their top view as molecules negatively stained presenting the side view.

The 50S ribosomal subunit, reconstructed in its crown view from negatively stained, sandwiched (Radermacher *et al.*, 1987a, b), and ice-embedded preparations (Radermacher *et al.*, 1992), appears to be flattened according to the ratio 0.7:1. A factor of 0.6:1 holds for the calcium release channel, as can be inferred from a comparison (Radermacher *et al.*, 1994b) of the side views of the cryo-reconstruction (Radermacher *et al.*, 1994a, b) with the reconstruction from the negatively stained specimen (Wagenknecht *et al.*, 1989a).

In summary, then, it is possible to say that as a rule, molecules prepared by negative staining are flattened to 60–70% of their original dimension. All evidence suggests that the flattening is normally avoided when ice or aurothioglucose embedment is used. In addition, the high degree of preservation of 2D bacteriorhodopsin crystals in glucose-embedded preparations (Henderson and Unwin, 1975; Henderson *et al.*, 1990) would suggest that single molecules embedded in glucose might also retain their shape, although to date this has been neither proved nor disproved by a 3D study. The important lesson for data merging is that *3D reconstructions from negatively stained molecules cannot be merged unless they are based on the same view of the molecule, i.e., on images showing the molecule facing the support grid in the same orientation.*

C. Three-Dimensional Orientation Search

1. Orientation Search Using Volumes

The reconstructed volumes have to be aligned both translationally and with respect to their orientation. This is achieved by a computational search in which the different parameters of shift and 3D orientation (five in all) are varied and the cross-correlation coefficient is used as similarity measure (Knauer *et al.*, 1983; Carazo and Frank, 1988; Carazo *et al.* 1989; Penczek *et al.*, 1992). The search range and the computational effort can be kept small if the approximate matching orientation can be estimated. Often it is clear from other evidence (e.g., interconversion experiments, see Section III, C; symmetries; or knowledge of architectural building principles, as in the case of an oligomeric molecule) how particular views

are related to one another in angular space. Examples are the top and side views of *A. australis* hemocyanin, which are related by a 90° rotation of the molecule around a particular axis of pseudosymmetry (see Appendix in Boisset *et al.*, 1988).

A major difficulty in finding the relative orientation of different reconstructions is presented by the missing cone. In Fourier space, the correlation between two volumes in any orientation is given by the sum over the terms $F_1(\mathbf{k})F_2^*(\mathbf{k})$ for all possible $\mathbf{k}$. Here $F_1(\mathbf{k})$ denotes the Fourier transform of the first volume in its original orientation, and $F_2^*(\mathbf{k})$ the complex conjugate of the Fourier transform of the second volume after it has been subjected to a "probing" rotation. Since the missing cones of data sets originating from different molecule views lie in different orientations, the orientation search is biased by a "vignetting effect." This effect results from the cones either intersecting each other to different extents, depending on the angles, or sweeping through regions of the companion transform that carry important parts of the structural information. Real-space methods of orientation determination search generally fail to deal with this complication and may therefore be inaccurate. Fourier methods, in contrast, allow precise control over the terms included in the correlation sum and are therefore usually preferable.

Figure 5.20 shows the outcome of an orientational search for the 70S *Escherichia coli* ribosome. In the matching orientations, the correlation coefficients are in the range between 0.8 and 0.86, justifying the assumption that we are indeed dealing with the same structure. The angles obtained when three or more structures are compared can be checked for *closure*. Penczek *et al.* (1992) applied this principle in the case of the 70S ribosome of *E. coli* where three reconstructions S_2, S_2, S_3 were available: the combination of the rotations found in the orientation search (S_1, S_2), (S_2, S_3) should give a result close to the rotation resulting from the search (S_1, S_3). [Note that each of these rotations is expressed in terms of three Eulerian angles (see Section II, B), so that the check actually involves the multiplication of two matrices]. More generally, in the situation where N reconstructions S_i ($i = 1 \ldots N$) are compared, then any "closed" string of pairwide orientation determinations (i.e., a string that contains one volume twice, in two different comparisons),

$$(S_{i_1}, S_{i_2}), \ldots, (S_{i_N}, S_{i_1}),$$

should result in rotations that, when combined, amount to no rotation at all. In the case of the 70S ribosome, the resulting closure error for three reconstructions was found to be in the range of 2° (Penczek *et al.*, 1992), which signifies excellent consistency.

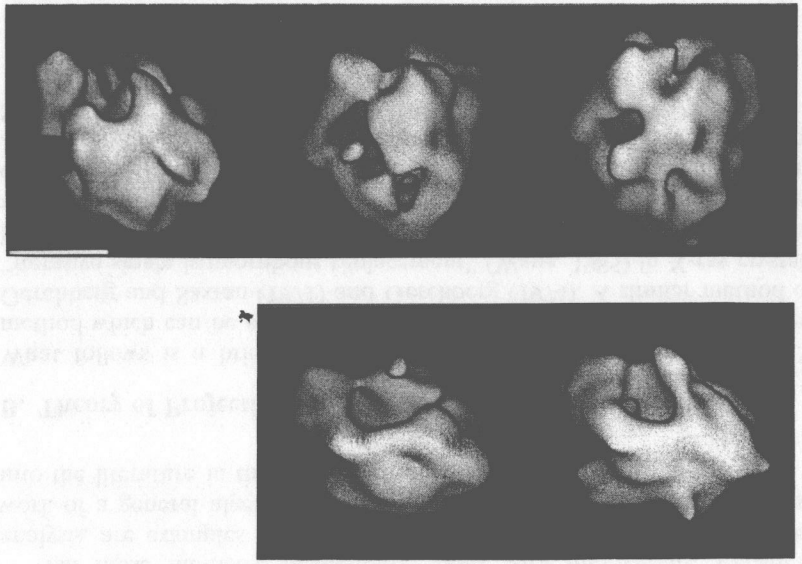


Fig. 5.20. Result of orientation search among three independent reconstructions of the ribosome (DPR-resolution; 1/47 Å^{-1}). Top row, the three reconstructions before orientational alignment; bottom row, reconstructions 2 and 3 after alignment with reconstruction 1. From Penczek *et al.* (1992). Reproduced with permission of Elsevier Science, Amsterdam.

2. *Orientation Search Using Sets of Projections (OSSP)*

Instead of the reconstructions, the projection sets themselves can be used for the orientation search (Frank *et al.*, 1992; Penczek *et al.*, 1994). In Fourier space, each random-conical projection set is represented by a set of central sections, tangential to a cone, whose mutual orientations are fixed (Fig. 5.21). It is obvious that the real-space search between volumes can be replaced by a Fourier space search involving the two sets of central sections. Instead of a single common line, the comparison involves N common lines (N being the number of projections in the two sets combined) *simultaneously*, with concomitant increase in signal-to-noise ratio. The method of comparison uses a discrepancy measure $1\text{-}\rho_{12}$, where ρ_{12} is the cross-correlation coefficient, computed over Fourier coefficients along the common line. Penczek *et al.* (1994) described the geometry underlying this search: the angle between the cones is allowed to vary over the full range. For any given angle, it is necessary to locate the "common line" intersections between all central sections of one cone with all central sections of the other.

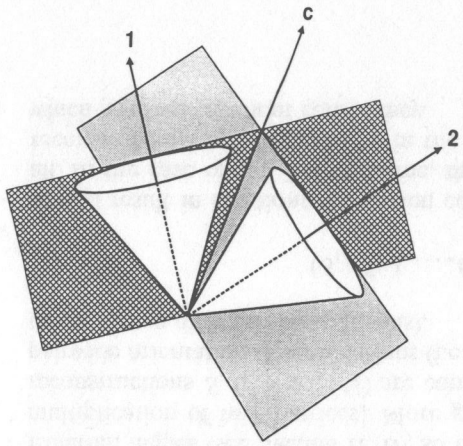


Fig. 5.21. Principle of the method of orientation search using sets of projections (OSSP). The two cones around axes 1 and 2 belong to two random-conical data collection geometries with different orientations of the molecule. Two given projections in the two geometries are represented by central Fourier sections tangential to the cones. These central sections intersect each other along the common line C. The OSSP method simultaneously considers every common line generated by the intersection of every pair of central sections. From Penczek *et al.* (1994). Reproduced with permission of Elsevier Science, Amsterdam.

While the location of the common line is constructed in Fourier space, the actual computation of the discrepancy measure is performed in real space, exploiting the fact that the one-dimensional common line found for any particular pairing of central sections is the Fourier transform of a one-dimensional projection. The OSSP method has two advantages over the orientation search between reconstruction volumes: it makes the computation of reconstructions prior to merging unnecessary, and it offers a rational way of dealing with the angular gap.

D. Reconstruction from the Full Projection Set

Once the relative orientations of the projection sets are known, either from an orientation search of the reconstruction volumes, or from an orientation search of the projection sets (as outlined in the previous section), the Eulerian angles of all projections of the combined projection sets can be formulated in a common coordinate system. It is then straightforward to compute the final, merged reconstruction.

It is important to realize that because of the properties of the general weighting functions, it is not possible to merge reconstructions by simply adding them. Instead, one must first go back to the projections, apply the appropriate coordinate transformations so that all projection angles relate to a single coordinate system, and then perform the reconstruction from the entire set. The same is true when one uses iterative reconstruction techniques, where the solution is also tied to the specific geometry of a projection set.

VII. Three-Dimensional Restoration

A. Introduction

3D restoration (as distinct from restoration of the contrast transfer function; see Section II, H in Chapter 2 and Section IX in this chapter) is the term we will use for techniques designed to overcome the angular limitation of a reconstruction which leads to resolution anisotropy and an elongation of the molecule in the direction of the missing data. *Maximum entropy methods* present one approach to restoration (Barth *et al.*, 1989; Farrow and Ottensmeyer, 1989; Lawrence *et al.*, 1989). These methods are known to perform well for objects composed of isolated peaks, e.g., stars in astronomical applications, but less well for other objects; see Trussell (1980). Another approach, based on a set theoretical formulation of the restoration and the enforcement of mathematical constraints, is known under the name of *projection onto convex sets* (*POCS*). POCS was developed by Youla and Webb (1982) and Sezan and Stark (1982) and introduced into electron microscopy by Carazo and Carrascosa (1987a, b). An overview chapter by Carazo (1992) addresses the general question of fidelity of 3D reconstructions and covers POCS as well as related restoration methods.

As was earlier mentioned, iterative reconstruction techniques allow nonlinear constraints to be incorporated quite naturally. In this case they actually perform a reconstruction-cum-restoration, which can also be understood in terms of the theory of POCS (Penczek, unpublished work, 1993).

All these methods, incidentally, along with multivariate statistical analysis, are examples of a development that treats image sets in framework of a general algebra of images. Hawkes (1993) has given a glimpse into the literature in this rapidly expanding field.

B. Theory of Projection onto Convex Sets

What follows is a brief introduction into the philosophy of the POCS method which can be seen as a generalization of a method introduced by Gerchberg and Saxton (1971) and Gerchberg (1974). A similar method of "iterative single isomorphous replacement" (Wang, 1985) in X-ray crystallography is also known under the name of *solvent flattening*. Still another, related method of "constrained thickness" in reconstructions of one-dimensional membrane profiles was proposed earlier on by Stroud and Agard (1979). An excellent primer for the POCS method was given by Sezan (1992).

Similarly as in multivariate statistical analysis of *images* (Chapter 4), which takes place in the space of all functions with finite 2D support, we now consider the space of all functions with finite 3D support. In the new space (Hilbert space), every conceivable bounded 3D structure is represented by a (vector end-) point. Constraints can be represented by sets. For instance, one conceivable set might be the set of all structures that have zero density outside a given radius R. The idea behind restoration by POCS is that the enforcement of known constraints that were not used in the reconstruction method itself will yield an improved version of the structure. This version will lie in the intersection of all constraint sets and thus closer to the true solution than any version outside of it. In Fourier space, the angular gap will tend to be filled. The only problem to solve is how to find a pathway from the approximate solution, reconstructed, for instance, by back-projection or any other conventional technique, to one of the solutions lying in the intersection of the constraint sets.

Among all sets representing constraints, those that are both *closed* and *convex* proved of particular interest. Youla and Webb (1982) showed that for such sets the intersection can be reached by an iterative method of consecutive *projections. A *projection from a function $f(\mathbf{r})$ onto a set C in Hilbert space is defined as an operation that determines a function $g(\mathbf{r})$ in C with the following property: "of all functions in C, $g(\mathbf{r})$ is the one closest to $f(\mathbf{r})$," where "closeness" is defined by the size of a distance; for instance by the generalized Euclidean distance in Hilbert space:

$$E = ||g(\mathbf{r}) - f(\mathbf{r})|| = \sum_{j=1}^{J} |g(\mathbf{r}_j) - f(\mathbf{r}_j)|^2. \tag{5.38}$$

Note that, by implication, repeated applications of *projection onto the same set lead to the same result.] In symbolic notation, if $\mathbf{P}_i$ denotes the operation of *projection onto set C_i, so that $f' = \mathbf{P}_i f$ is the function obtained by *projecting f onto C_i, the iterative restoration proceeds as follows:

$$\begin{aligned} f^{(1)} &= \mathbf{P}_1\mathbf{P}_2 \ldots \mathbf{P}_n f^{(0)} \\ f^{(2)} &= \mathbf{P}_1\mathbf{P}_2 \ldots \mathbf{P}_n f^{(1)} \end{aligned} \tag{5.39}$$

etc. As the geometric analogy shows (Fig. 5.22), by virtue of the convex property of the sets, each iteration brings the function (represented by a point in this diagram) closer to the intersection of all constraint sets.

Carazo and Carrascosa (1987a, b) already discussed closed, convex constraint sets of potential interest in electron microscopy: spatial boundedness (as defined by a binary mask), agreement with the experimental

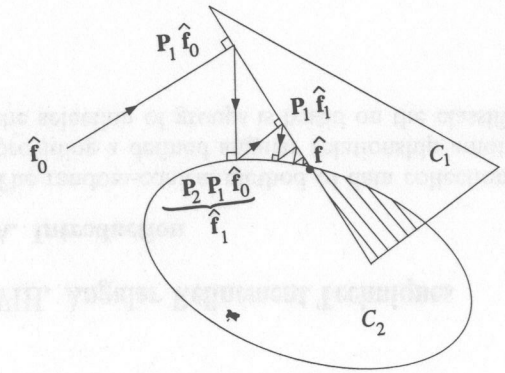


Fig. 5.22. Principle of restoration using the method of *projection onto convex sets. C_1 and C_2 are convex sets in the space R^N, representing constraints, and P_1, P_2 are associated *projected operators. Each element f_i is a 3D structure. We seek to find a pathway from a given blurred structure f_0 to the intersection set (shaded). Any structure in that set fulfills both constraints and is thus closer to the true solution than the initial structure f_0. From Sezan (1992). Reproduced with permission of Elsevier Science, Amsterdam.

measurements in the measured region of Fourier space, value boundedness, and energy boundedness. Thus far, in practice (see following), only the first two on this list have gained much importance, essentially comprising the two components of Gerchberg's (1974) method.[20] Numerous other constraints of potential importance [see, for instance, Sezan (1992)] still await exploration.

C. Projection onto Convex Sets in Practice

In practice, the numerical computation in the various steps of POCS has to alternate between real space and Fourier space for each cycle. Both support (mask) and value constraints are implemented as operations in real space, while the "replace" constraint takes place in Fourier space. For the typical size of a 3D array representing a macromolecule (between $64 \times 64 \times 64$ and $100 \times 100 \times 100$), the 3D Fourier transformations in both directions constitute the largest fraction of the computational effort.

[20] Gerchberg's (1974) method, not that of Gerberg and Saxton (1972), is a true precursor of POCS since it provides for replacement of *both* modulus and phases.

The *support*-associated *projector is of the following form:

$$\mathbf{P}_s f = \begin{cases} f(i); & i \in M \\ 0 & \text{otherwise,} \end{cases} \tag{5.40}$$

where M is the set of indices defining the "pass" regions of the mask. In practice, a mask is represented by a binary-valued array with "1" representing "pass" and "0" representing "stop." The mask array is simply interrogated, as the discrete argument range of the function f is being scanned in the computer, and only those values of $f(i)$ are retained for which the mask M indicates "pass." The support constraint is quite powerful if the mask is close to the actual boundary, and an important question is how to find a good estimate for the mask in the *absence* of information on the true boundary (which represents the normal situation). We will come back to this question later after the other constraints have been introduced.

The *value constraint* is effected by the *projector (Carazo, 1992)

$$\mathbf{P}_v f = \begin{cases} a & f < a \\ f & a \leqq f \leqq b \\ b & f > b \end{cases} \tag{5.41}$$

The *measurement constraint* is supposed to enforce the consistency of the solution with the known projections. This is rather difficult to achieve in practice because the projection data in Fourier space are distributed on a polar grid while the numerical Fourier transform is sampled on a Cartesian grid. Each POCS *projection would entail a complicated Fourier–sinc interpolation. Instead, the measurement constraint is normally used in a weaker form, as a "global replace" operation: within the range of the measurements (i.e., in the case of the random-conical data collection, within the cone complement that is covered with projections; see Fig. 5.11), all Fourier coefficients are replaced by the coefficients of the solution found by weighted back-projection.

This kind of implementation is somewhat problematic, however, because it reinforces a solution that is eventually not consistent with the true solution because it incorporates a weighting that is designed to make up for the lack of data in the missing region. A much better "replace" operation is implicit in the iterative schemes in which agreement is enforced only between projection data and reprojections. In Fourier space, these enforcements are tantamount to a "replace" that is restricted to the Fourier components for which data are actually supplied.

The use of the global replace operation also fails to realize an intriguing potential of POCS: the possibility of achieving anisotropic

superresolution, beyond the limit given by Crowther *et al.* (1970). Intuitively, the enforcement of "local replace" (i.e., only along central sections covered with projection data) along with the other constraints will fill the very small "missing wedges" between successive central sections on which measurements are available much more rapidly, and out to a much higher resolution, than the large missing wedge or cone associated with the data collection geometry. The resolution factor might be as large as two—theoretical calculations are still pending.

What could be the use of anisotropic superresolution? An example is the tomographic study of the mitochondrion (Mannella *et al.*, 1994), so far hampered by the extremely large ratio between size (several microns) and the size of the smallest detail we wish to study (5 nm). The mitochondrion is a large structure that encompasses, and is partially formed by, a convoluted membrane. We wish to obtain the spatial resolution *in any direction* that allows us to describe the spatial arrangements of the different portions of the membrane: Do they touch? Are compartments formed? What is the geometry of the diffusion-limiting channels? The fact that the important regions where membranes touch or form channels occur in different angular directions makes it highly likely in this application to pick up the relevant information. The subject of tomography is outside the scope of this book, but similar problems where even anisotropic resolution improvement may be a bonus could well be envisioned in the case of macromolecules.

Examples for the application of POCS to experimental data are found in the work by Akey and Radermacher (1993) and Radermacher *et al.* (1992b, 1994b). In all these cases, only the measurement and the finite support constraints were used. In the first case, the nuclear pore complex was initially reconstructed from data obtained with merely 34° and 42° tilt and thus had an unusually large missing-cone volume. Radermacher *et al.* (1994b) observed that POCS applied to a reconstruction from a negatively stained specimen led to a substantial contraction in z direction (see Fig. 5.23 in this Chapter and Fig. 7.6 in Chapter 7) while the ice reconstruction was relatively unaffected.

VIII. Angular Refinement Techniques

A. Introduction

The random-conical method of data collection was developed as a way of providing a defined angular relationship among a set of molecules. Since the selection of groups is based on the classification of molecule views in

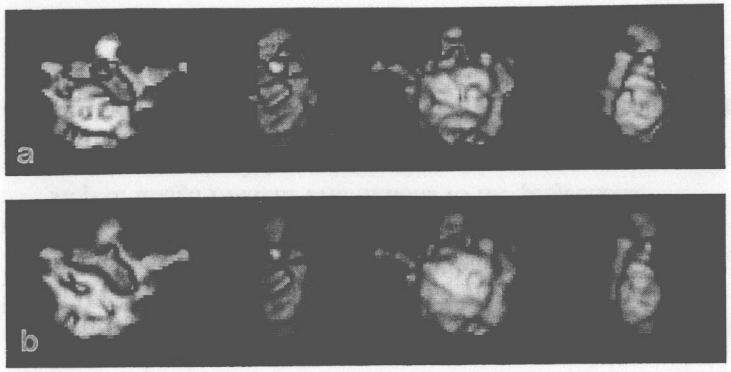


Fig. 5.23. Example for application of POCS. (a) Four views of the 50S ribosomal subunit, reconstructed from images of negatively stained specimen using random-conical data collection and weighted back-projection; (b) the structure in (a) after application of POCS using only "replace" and "boundedness" *projectors. The particle is seen to flatten in z direction, as a result of some filling of the missing cone in the low spatial frequency region. Reproduced with permission of M. Radermacher (unpublished).

the 0° micrograph, there exists an uncertainty in the actual size of the θ angle which defines the inclination of the central section associated with the projection. This angular uncertainty can easily reach $\pm 10°$ and reduce the resolution of the reconstruction substantially (see the estimates given in Section VIII, D).

At this point it should be reiterated (see Section V, H) that the cross-resolution comparison between the 0° average of a molecule set in a certain view and the 0° projection of the reconstruction, valuable as it is as a check for internal consistency (Section V, I), nevertheless fails to provide an adequate estimate of over-all resolution. It has been pointed out (Penczek *et al.*, 1994) that the 0° projection is insensitive to incorrect assignments of θ, since its associated central section in Fourier space is built up from one-dimensional lines, each of which is an intersection between the 0°-central section and the 50°-central section.

Angular refinement, by giving each projection a chance "to find a better home" in terms of its orientation and phase origin, improves the resolution of the reconstruction substantially. For the 70S ribosome, an improvement from 1/47.5 to 1/40 $Å^{-1}$ was reported (Penczek *et al.*, 1994). An angular refinement technique essentially based on the same principle, of using an existing lower-resolution reconstruction as a template, has been used recently in the processing of virus particles (Cheng *et al.*, 1994). The two schemes resemble earlier schemes proposed by

van Heel (1984) and Harauz and Ottensmeyer (1984a). In fact, all four schemes, although differing in the choice of computational schemes and the degree of formalization, can be understood as variants of the same approach.

B. Three-Dimensional Projection Matching Method

In the 3D projection matching scheme, reference projections (in the following termed "reprojections") are computed from the (low-resolution) template structure such that they cover the entire angular space evenly (Fig. 5.24). As template structure, an existing 3D volume is used, which might have been obtained by merging several random-conical projection sets. In the application by Penczek *et al.* used to demonstrate the technique—the 70S ribosome from *E. coli*—5266 such reference projections were obtained. A given experimental projection is cross-correlated with all reference projections. The angle giving the largest CCF peak is the desired projection angle. The cross-correlation function between reference and current projection gives, at the same time, the shift and the azimuthal orientation of the particle under consideration. Using the new parameters for each experimental projection, a new reconstruction is computed, which normally has improved resolution. This refined reconstruction can now be used as new template, etc. Usually, the angles no longer change by significant amounts after two or three iterations of this scheme (see Section VIII, D for a discussion of this point in the light of experimental results). In the demonstration by Penczek *et al.* (1994), additional 0° or low-tilt data could be used, following this approach, to improve the resolution from 1/40 to 1/29 $Å^{-1}$. [0° Data here means "the collection of particle images in an untilted specimen field."] The end results of the refinement, and the evenness of the distribution can be checked by plotting a chart that shows the 3D angular distribution of projections (Fig. 5.25). Before refinement, each random-conical set of projections will be mapped into a circle of data points on such a chart. It is seen that the angular corrections in the refinement have wiped out all traces of these circular paths.

The 3D projection matching technique as described above (Penczek *et al.*, 1994; Cheng *et al.*, 1994) is closely related to Radermacher's 3D Radon transform method (Radermacher, 1994; see below). [This relationship can be understood in terms of the so-called X-ray transform versus Radon transform (see Natterer, 1986).]. It also has close similarities with projection matching techniques that were formulated some time ago, when exhaustive search techniques with the computer were quite time-

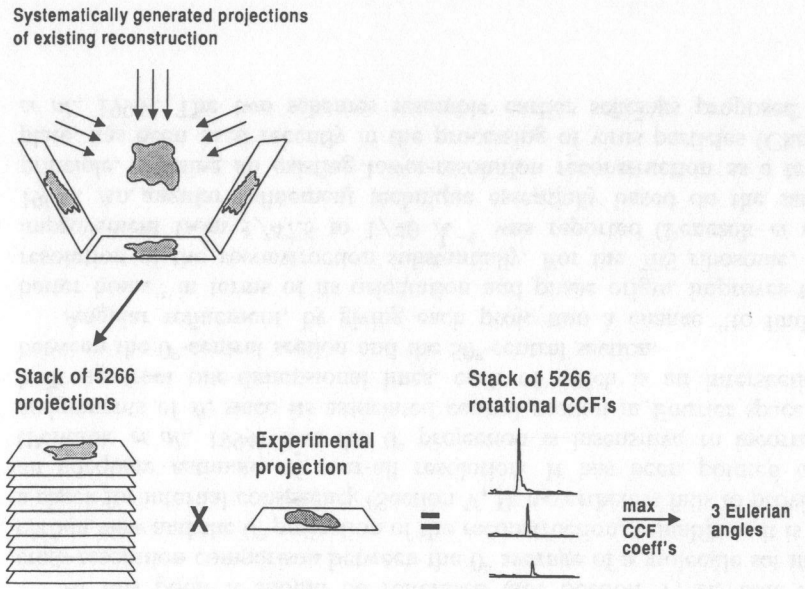


Fig. 5.24. Principle of 3D angular alignment or refinement. From an existing reconstruction (top left), a large number of projections are obtained, covering the 3D orientation space as evenly as possible. In the case illustrated, 5266 projections are computed. 3D angular alignment of new data: a given experimental projection *that is not part of the projection set from which the reconstruction was obtained* is now cross-correlated with all trial projections. The direction for which maximum correlation is obtained is then assigned to the new projection. In this way, an entire new data set can be merged with the existing data so that a new reconstruction is obtained. Angular refinement of existing data: each projection that is part of the experimental projection set is matched in the same way with the computed projections to find a better angle than originally assigned. However, in this case, the search range does not have to extend over the full space of orientations, because large deviations are unlikely. The best choice of search range can be gauged by histograms of angular "movements"; see Section VIII, D and particularly Fig. 5.27. From Penczek *et al.* (1994). Reproduced with permission of Elsevier Science, Amsterdam.

consuming and therefore still impractical to use. These relationships will be briefly summarized in the following:

(i) van Heel (1984b) proposed a method to obtain the unknown orientations of a projection set using the following sequence:

step 0: assign random orientations to the projections;

step 1: compute 3D reconstruction;

step 2: project 3D reconstruction in all directions in space to match experimental projections. The parameters for which best matches are

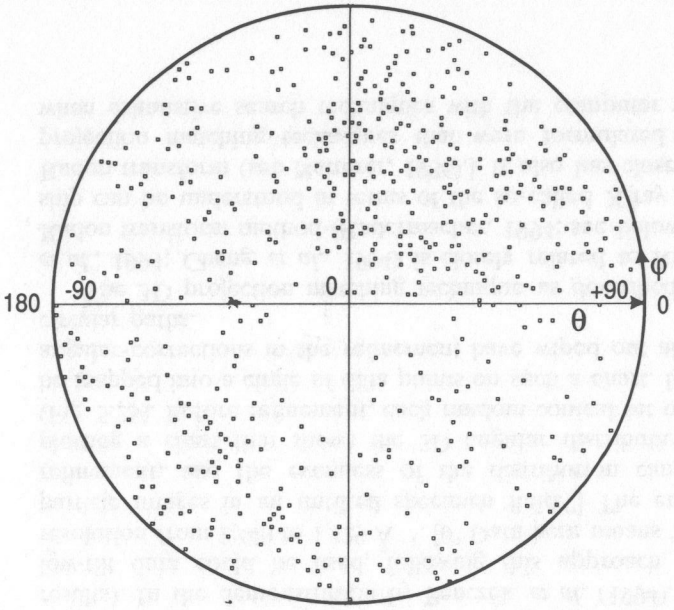


Fig. 5.25. Distribution of directions of 567 projections after refinement through 3D projection alignment. A single random-conical projection set prior to refinement would be represented by a circular pattern of dots on this map. From Penczek *et al.* (1994). Reproduced with permission of Elsevier Science, Amsterdam.

obtained yield new orientations for the projections; compute differences between current model projections and experimental projections; if summed squared differences are larger than a predefined value then GO TO step 1, otherwise STOP.

This procedure thus contains the main ingredients of the algorithm of Penczek *et al.*, except for the choice of starting point (step 0), which makes it difficult for the algorithm to find a satisfactory solution except in quite fortuitous cases where the random assignment of angles happens to come close to the actual values. In all other cases, the initial reconstruction will not likely resemble a low-resolution version of the true structure, and will fail to steer the orientation assignments in the correct directions.

(ii) The *Harauz and Ottensmeyer* (1984a, b) approach differs from all approaches discussed thus far in that it uses a computer-generated model of the predicted structure rather than an experimental reconstruction as an initial 3D template and a combination of visual and computational

analyses to obtain an optimum fit for each projection. Because of the small size of the object (nucleosome cores whose phosphorus signal was obtained by energy filtering) and the type of specimen preparation used (air-drying), the result was greeted with considerable scepticism. There is also a principal question whether the use of an imposed 3D reference (as opposed to an experimental 3D reference) might not bias the result. We recall the reports about reference-induced averages in the two-dimensional case (Radermacher *et al.*, 1986b; Boekema *et al.*, 1986), and our previous discussion of this question as in introduction to reference-free alignment schemes (see Chapter 3, Section III, E, 1).

As van Heel (1984b), Harauz and Ottensmeyer (1984a, b) also make use of the summed squared difference, rather than the cross-correlation, to compare an experimental projection with the reprojections. It was earlier pointed out (Chapter 3, Section III, C, 1) that there is no practical difference between the summed squared difference (or Euclidean distance) and the cross-correlation as measures of "goodness of fit" when comparing 2D images that are rotated and translated with respect to one another. This is so because the variance terms in the expression of the Euclidean distance are translation- and rotation-invariant. In contrast, the two measures *do* behave differently when employed to compare projections of two structures, since the variance of a projection may strongly depend on the projection direction.

(iii) *Alignment of correlation-averaged projections.* Saxton *et al.* (1984) developed a somewhat related alignment technique as part of a strategy to reconstruct a crystal from projections that have been obtained by correlation averaging. New projections are added incrementally to a data set by aligning it to a "pseudoprojection" generated from the existing layer lines. A refinement pass was designed in which each projection, again, has the opportunity "to find a new home": each projection is matched with the data set which has been modified by exclusion of that projection.

(iv) *Multiresolution approach.* Dengler (1989) discussed projection matching in the general framework of a multiresolution approach to reconstruction. He addressed the important problem of error propagation, which can be solved only by, in the words of the author, "iterative control strategy from coarse to fine." Some of Dengler's ideas, relating to the modeling of a space variant displacement vector field, have yet to be implemented and tested.

(v) *Method of "inactive" *projection in factor space* (Carazo *et al.*, 1989). When a projection set is analyzed by correspondence analysis or by a similar method of multivariate statistical analysis, closely matched projections come to lie in close proximity to one another in factor space. When an existing model structure is projected successively along a closed

angular pathway, and the resulting projection set is analyzed by correspondence analysis, the data points representing the projections can be observed to fall on a closed loop (see the demonstration of this principle by Frank *et al.*, 1986). In principle, this offers the possibility of determining orientations of experimental projections: these projections are "inactively" (i.e., without participating in the factor analysis) *projected into the factor space spanned by the model projection set. Assignment of angles then is on the basis of proximity to active data points with known angles. A similar strategy could be used for refinement (although the computational effort would be substantial): by making each projection of a data set in turn "inactive."

(vi) *Matching of projections with a theoretical model.* For completeness, it should be mentioned that projection matching plays a role in attempts to fit a theoretical model to experimental projections. The approach is initially similar to that of Harauz and Ottensmeyer (1984a, b), in that a theoretical model is used, but differs from the latter in the important fact that no attempt is made to obtain an experimental 3D reconstruction, mainly because of insufficient angular coverage. For example, De Haas and van Bruggen (1994) investigated the orientations of the four hexamers of the tarantula hemocyanin by cross-correlating their averaged projections with a model projected into a large range of orientations. van Heel and Dube (1994) refined the parameters of the architectural model of *Limulus polyphemus* hemocyanin using a similar matching method. Boisset *et al.* (1990a) were able to explain the appearance of stained molecules of *Scutigera coleoptrata* hemocyanin on a single carbon film by calculating projections of a stain-exclusion model generated in the computer (see Fig. 3.4 in Chapter 3).

C. Three-Dimensional Radon Transform Method

Although its principle is quite similar to the projection matching approaches described above, the Radon transform method (Radermacher, 1994) is distinguished from the former by its formal elegance and the fact that it brings out certain relationships that may not be evident in the other approaches. Radermacher's method takes advantage of the relationship between the 3D Radon transform and the 2D Radon transform. The 3D Radon transform is defined for a 3D function $f(\mathbf{r})$ as

$$\hat{f}(p, \xi) = \int f(\mathbf{r}) \delta(p - \xi \mathbf{r})\, d\mathbf{r} \tag{5.42}$$

where $\mathbf{r} = (x, y, z)^T$ and $\delta(p - \xi^T\mathbf{r})$ represent a *plane* defined by the direction of the (normal) unit vector ξ. The 2D Radon transform is defined for a 2D function $g(\mathbf{r})$, in analogy to Eq. (5.42), as

$$\hat{g}(p, \xi) = \int g(\mathbf{r})\delta(p - \xi^T\mathbf{r})\, d\mathbf{r} \qquad (5.43)$$

where $\mathbf{r} = (x, y)^T$ and $\delta(p - \xi^T\mathbf{r})$ now represents a *line* defined by the direction of the (normal) unit vector ξ. The discrete 2D Radon transform is also known under the name *sinogram* (e.g., van Heel, 1987b). Radermacher (1994) shows that the determination of the unknown orientation of a projection is solved by cross-correlating its discrete 2D Radon transform with the discrete 3D Radon transform of the existing model (Fig. 5.26). Translational alignment (equivalent to "phasing" in Fourier space) can be done simultaneously.

As we recall from Section III, D, the use of the cross-correlation function between 2D sinograms was proposed by van Heel (1987b) under the name *angular reconstitution*, as a means of determining the relative orientation between two or more *raw*, *experimental* projections. Applied to such data, that method generally fails because of the low SNR values normally encountered in electron micrographs of stained or frozen-hydrated specimens, unless high point symmetries are present. However, the method has proved viable when applied to the matching of an *experimental* to an *averaged* projection, or the matching of one averaged projection to another; see van Heel *et al.*, 1994; Orlova and van Heel, 1994; Serysheva *et al.*, 1995).

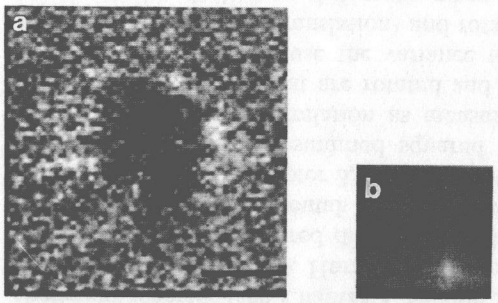


Fig. 5.26. Demonstration of angle search with the 3D Radon transform, using the cryo-reconstruction of the 50S ribosomal subunit of *Escherichia coli* (see Fig. 5.19). (a) Computed projection of the reconstruction into the direction given by the angles $\{\psi = 0°, \theta = 45°, \phi = 30°\}$, with noise added (SNR = 0.88). (b) The $\psi = 0$ plane of the 3D cross-correlation between the Radon transform of the projection in (a) and the 3D Radon transform of the 50S-subunit reconstruction. The peak is found centered at $\{\theta = 45°, \phi = 30°\}$. From Radermacher (1994). Reproduced with permission of Elsevier Science, Amsterdam.

D. The Size of Angular Deviations

Little is known about the actual size of the angular deviations of a molecule within its class,[21] but inferences can be drawn from the results of the angular refinements. Some data (unpublished) are available from the ribosome study of Penczek *et al.* (1994). As described in Section VIII, B the refinement by projection matching is done in several passes. For each pass, one can make a histogram of angular adjustments of the particles. These are given in terms of three Eulerian angles (ψ, θ, ϕ), of which ψ and ϕ have to do with the rotations of the particle around axes not of interest here, and only θ gives the information about its "tilt." In the first pass (Fig. 5.27a), one half of the particles "moved" by less than 10°, one-third moved by angles between 10° and 35°, and the rest moved by larger angles. In the second pass (Fig. 5.27b), a full 80% of the particles moved by less than 5° (50% are even within 1%; not shown in this figure), indicating approximate stabilization of the solution. One can therefore take the angular adjustments in the first pass as estimates for the actual angular deviations.

The size of these deviations—a full half of the particles are tilted by more than 10° away from the orientation of their class—is at first glance surprising. It certainly explains (along with ψ- and ϕ-deviations not depicted in Fig. 5.27) the great gain in resolution that was achieved by correcting the angles. The deviations are the result of both misalignment and misclassification of the extremely noisy data. What it means is that the unrefined reconstruction is essentially a superposition of a high-quality reconstruction (where the resolution limitation is due to factors unrelated to angular deviations, namely electron-optical limitations, conformational changes, etc.) and a blurred reconstruction, with the former based on projections whose θ angles are closely matching and the latter based on projections whose θ angles fall into a wide range. With such a concept, it is now possible to understand the power of the angular refinement method: basically, a high-resolution reconstruction is already "hidden" in the unrefined reconstruction and it furnishes weak but essentially accurate reference information in the course of the angular refinement passes (see Fig. 5.27).

[21] This obviously depends on the definition of the class obtained by MSA and classification. However, the classification in the example used is rather typical for particles in the 200 Å size range.

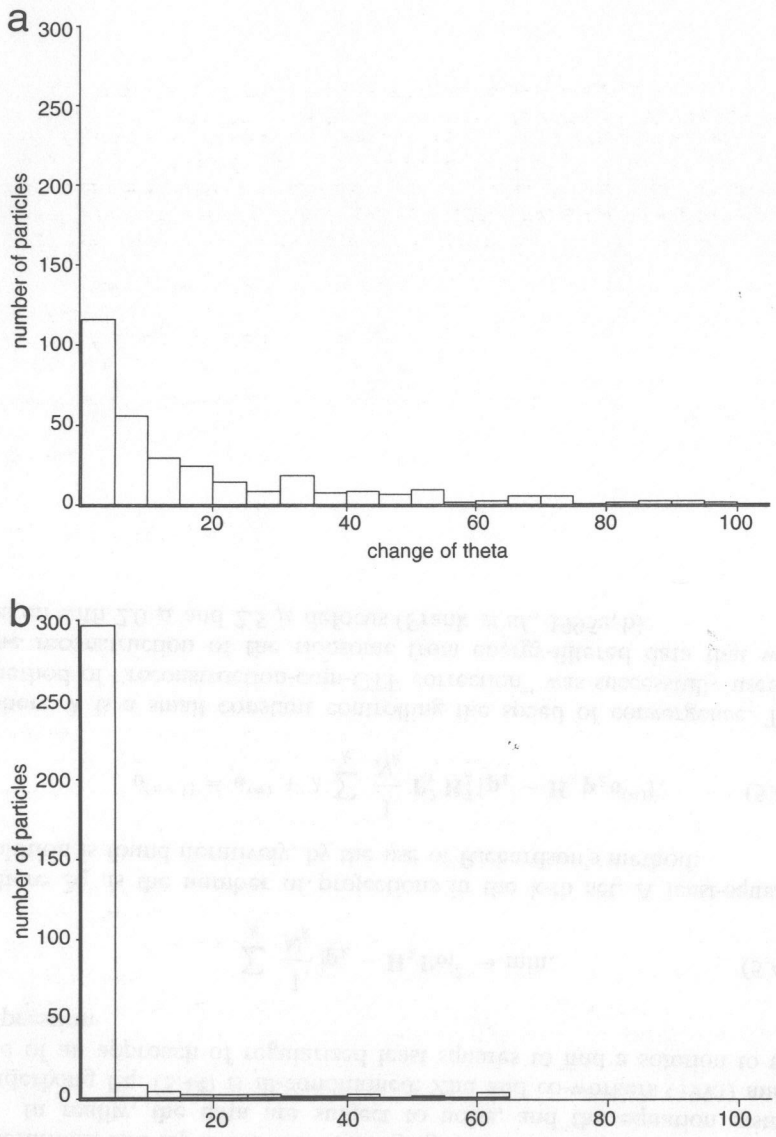


Fig. 5.27. Histograms showing the change of the theta-angle during angular refinement. (a) First refinement pass, (b) second refinement pass. For explanation, see text. (P. Penczek, R. Grassucci, and J. Frank, 1994, unpublished data).

Histograms similar to those shown in Fig. 5.27 were already obtained, in model computations, by Harauz and Ottensmeyer (1984a), who gave the projections intentionally incorrect angular assignments. Since these authors used a model structure, they could study the actual angular improvements as a function of the number of iterations. What is interesting in the results of Harauz and Ottensmeyer is that angular error limits of $\pm 10°$ are rapidly compensated, to a residual error below 2°, while limits of $\pm 20°$ lead to residual errors in the range of 8°. Since the algorithm driving the correction of angles in this work differs somewhat from that employed by Penczek *et al.* (1994), the behavior of angular correction is not strictly comparable. However, it is likely that there is again a limit of rms angular deviation below which the angular refinement is well behaved and very efficient, but above which only small improvements might be achievable.

IX. Transfer Function Correction

Without correction for the effects of the contrast transfer function, the reconstruction will have exaggerated features in the size range passed by the CTF spatial frequency band. Most importantly, the definition of the whole particle against the background is affected. Procedures for CTF correction were already discussed in Chapter 2. These can either be applied to the raw data (i.e., the projections) or to the 3D volume (or volumes). The most effective correction is obtained by combining data sets obtained with two or more different defocus settings.

In deciding whether to apply correction *before* or *after* the 3D reconstruction, one has to consider the following pros and cons: when applied to averages, such as the 3D reconstructions, all procedures mentioned in Section II, H of Chapter 2 are very well behaved numerically because of the high SNR of the data. The opposite is true when these procedures are applied to raw data. On the other hand, correction after reconstruction runs into the difficulty that the projections from tilted-specimen micrographs have different defocus values (see Frank and Penczek, 1995). In order to proceed in this way, one has to sort the raw data according to the distance from the tilt axis and perform separate reconstructions for each defocus strip.

For processing data from untilted specimens taken at different defocus settings, Zhu and co-workers (1995) developed a method of reconstruction that implicitly corrects for the CTF. The is done by including the CTF into the mathematical model describing the relationship between the three-dimensional model and the observed projections.

In algebraic form, this relationship can be formulated as

$$\mathbf{p}_k = \mathbf{H}_k \mathbf{P}\mathbf{o}, \tag{5.44}$$

where $\mathbf{p}_k$ is a matrix containing the projection data for the k-th defocus setting, $\mathbf{o}$ is a vector representing the elements of the 3D object in lexicographic order, $\mathbf{P}$ is a non-square matrix describing the projection operations, and $\mathbf{H}_k$ is the CTF belonging to the k-th defocus.

In reality, the data are subject to noise, and the equation system underlying Eq. (5.44) is ill-conditioned. Zhu and co-workers (1995) make use of an approach of regularized least squares to find a solution to the expression

$$\sum_k \frac{1}{N_k} |\mathbf{p}_k - \mathbf{H}_k \mathbf{P}\mathbf{o}|^2 \rightarrow \min, \tag{5.45}$$

where N_k is the number of projections in the k-th set. A least-squares solution is found iteratively, by the use of Richardson's method:

$$\mathbf{o}^{(n+1)} = \mathbf{o}^{(n)} + \lambda \sum_k \frac{1}{N_k} \mathbf{P}_k^T \mathbf{H}_k^T \{\mathbf{p}_k - \mathbf{H}_k \mathbf{p}_k \mathbf{o}^{(n)}\}, \tag{5.46}$$

where λ is a small constant controlling the speed of convergence. This method of "reconstruction-cum-CTF correction" was successfully used in the reconstruction of the ribosome from energy-filtered data that were taken with 2.0 μ and 2.5 μ defocus (Frank *et al.*, 1995a, b).