

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

$$k = \mathcal{F}^{-1}[\mathcal{F}J_1 \cdot \mathcal{F}^*J_2]. \quad (2.5.5.27)$$

The probability density of samples for images has the form

$$p(J_1 J_2 \dots J_n) = \frac{1}{(\sigma\sqrt{2\pi})^n} \times \exp\left[\frac{-1}{2\sigma^2} \sum^n \int [J_k(\mathbf{x} + \mathbf{x}_k) - J(\mathbf{x})]^2 dx\right]. \quad (2.5.5.28)$$

Here  $J$  is the tentative image (as such, a certain ‘best’ image can first be selected, while at the repeated cycle an average image is obtained),  $J_k(\mathbf{x})$  is the image investigated,  $\sigma$  is the standard deviation of the normal distribution of noises and  $x_k$  the relative shift of the image. This function is called a likelihood function; it has maxima relative to the parameters  $J(x)$ ,  $x_k$ ,  $\sigma$ . The average image and dispersion are

$$J(\mathbf{x}) = (1/n) \sum^n [J_k(\mathbf{x} - \mathbf{x}_k)], \quad \sigma^2 = (1/n) \sum^n [J_k(\mathbf{x} - \mathbf{x}_k) - J(\mathbf{x})]^2. \quad (2.5.5.29)$$

This method is called the maximum-likelihood method (Cramér, 1954; Kosykh *et al.*, 1983).

It is convenient to carry out the image alignment, in turn, with respect to translational and angular coordinates. If we start with an angular alignment we first use autocorrelation functions or power spectra, which have the maximum and the symmetry centre at the origin of the coordinates. The angular correlation maximum

$$f(\theta') = \int f_k(\theta - \theta') f_c(\theta) d\theta \quad (2.5.5.30)$$

gives the mutual angle of rotation of two images.

Then we carry out the translational alignment of rotationally aligned images using the translational correlation function (2.5.5.26) (Langer *et al.*, 1970).

In the iteration alignment method, the images are first translationally aligned and then an angular shift is determined in image space in polar coordinates with the centre at the point of the best translational alignment. After the angular alignment the whole procedure may be repeated (Steinkilberg & Schramm, 1980).

The average image obtained may have false high-frequency components. They can be excluded by multiplying its Fourier components by some function and suppressing high-space frequencies, for instance by an ‘artificial temperature factor’  $\exp\{-B|\mathbf{u}|^2\}$ .

For a set of similar images the Fourier filtration method can also be used (Ottensmeyer *et al.*, 1977). To do this, one should prepare from these images an artificial ‘two-dimensional crystal’, *i.e.* place them in the same orientation at the points of the two-dimensional lattice with periods  $a$ ,  $b$ .

$$J = \sum_{k=1}^n J_k(\mathbf{x} - \mathbf{t}_p); \quad \mathbf{t} = p_1 \mathbf{a} + p_2 \mathbf{b}. \quad (2.5.5.31)$$

The processing is then performed according to (2.5.5.18), (2.5.5.19); as a result one obtains  $\langle I(xy) \rangle$  with reduced background. Some translational and angular errors in the arrangement of the images at the artificial lattice points act as an artificial

temperature factor. The method can be realized by computing or by optical diffraction.

 2.5.6. Three-dimensional reconstruction<sup>5</sup>

By B. K. VAINSHEIN AND P. A. PENCZEK

## 2.5.6.1. The object and its projection

In electron microscopy (EM) and single-particle reconstruction, three-dimensional (3D) reconstruction methods are used for studying biological structures; that is, symmetric or asymmetric associations of biomacromolecules (muscles, spherical and rod-like viruses, bacteriophages, individual proteins and ribosomes) (Frank, 2006). The electron microscope is used to obtain parallel-beam two-dimensional (2D) projections  $\varphi_2(\mathbf{x}, \boldsymbol{\tau})$  (ray transform) of frozen hydrated 3D macromolecules  $\varphi_3(\mathbf{r})$  suspended in random orientations (Fig. 2.5.6.1). The function  $\varphi_2(\mathbf{x}_\tau)$  is the 2D projection of the 3D molecular electron distribution  $\varphi_3(\mathbf{r})$ . One can also consider one-dimensional (1D) projections  $\varphi_1(s, \boldsymbol{\tau})$  of multidimensional distributions; the set of these projections is called a Radon transform. For 2D distributions, ray and Radon transforms differ only in the notation. For ( $d > 2$ )-dimensional distributions the two are different: in a ray transform the integrals are calculated over straight lines and yield ( $d - 1$ )-dimensional projections, while in Radon transforms the integrals are calculated over ( $d - 1$ )-dimensional hyperplanes and yield 1D projections. In electron microscopy, Radon transforms are not directly measurable, but can be formed computationally and used in intermediate steps of the 3D reconstruction or in alignment procedures (Radermacher, 1994).

Within the linear weak-phase-object approximation of the image-formation process in the microscope [see equation (2.5.2.42) in Section 2.5.2 of this chapter], 2D projections represent line integrals of the potential of the particle under examination convoluted with the point-spread function of the microscope,  $s$ , so, using (2.5.2.43),

$$I(xy) = 1 + 2\sigma s(xy) * \varphi(xy) = 1 + 2\sigma s(xy) * \int \varphi_3(\mathbf{r}) dz. \quad (2.5.6.1)$$

Since

$$\varphi_2(\mathbf{x}_\tau) = \int \varphi_3(\mathbf{r}) d\tau, \quad \boldsymbol{\tau} \perp \mathbf{x}, \quad (2.5.6.2)$$

we have

$$I(xy) = 1 + 2\sigma s(\mathbf{x}_\tau) * \varphi_2(\mathbf{x}_\tau), \quad \boldsymbol{\tau} \perp \mathbf{x}. \quad (2.5.6.3)$$

If we omit constant terms, we obtain

$$I(xy) = s(\mathbf{x}_\tau) * \varphi_2(\mathbf{x}_\tau) = s(\mathbf{x}_\tau) * \int \varphi_3(\mathbf{r}) d\tau, \quad \boldsymbol{\tau} \perp \mathbf{x}. \quad (2.5.6.4)$$

In this section, we will assume that all images were collected using the same defocus setting, so the point-spread function  $s$  is constant and does not depend on the projection direction  $\boldsymbol{\tau}$ . Thus, we will concern ourselves with the inversion of the projection problem

$$\varphi_2(\mathbf{x}_\tau) = \int \varphi_3(\mathbf{r}) d\tau, \quad \boldsymbol{\tau} \perp \mathbf{x}. \quad (2.5.6.5)$$

<sup>5</sup> The original version of Section 2.5.6, written by the late B. K. Vainshtein, is here updated and expanded by P. A. Penczek.

## 2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

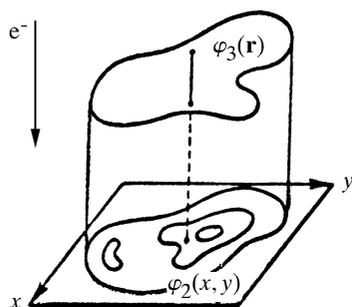


Fig. 2.5.6.1. A three-dimensional object  $\varphi_3$  and its two-dimensional projection  $\varphi_2$ . The electron beam penetrates the specimen in the direction of the  $z$  axis.

The projection direction is defined by a unit vector  $\boldsymbol{\tau}(\theta, \psi)$  and it is formed on the plane  $\mathbf{x}$  perpendicular to  $\boldsymbol{\tau}$ . The set of various projections  $\varphi_2(\mathbf{x}_\tau) = \varphi_2(\mathbf{x}_i)$  may be assigned by a discrete or continuous set of points on a unit sphere  $|\boldsymbol{\tau}| = 1$  (Fig. 2.5.6.2).

In Fourier space, the relation between an object and its projection is referred to as the central section theorem: the Fourier transformation of projection  $\varphi_2$  of a 3D object  $\varphi_3$  is the central (*i.e.*, passing through the origin of reciprocal space) 2D plane cross section of a 3D transform perpendicular to the projection vector (Bracewell, 1956; DeRosier & Klug, 1968; Crowther, Amos *et al.*, 1970). In Cartesian coordinates, a 3D Fourier transform is

$$\begin{aligned} \mathcal{F}[\varphi_3(\mathbf{r})] &= \Phi_3(u, v, w) \\ &= \iiint \varphi_3(x, y, z) \exp\{2\pi i(ux + vy + wz)\} dx dy dz. \end{aligned} \quad (2.5.6.6)$$

The transform of projection  $\varphi_2(x, y)$  along  $z$  is

$$\begin{aligned} \mathcal{F}[\varphi_2(x, y)] &= \Phi_3(u, v, 0) \\ &= \iiint \varphi_3(x, y, z) \exp\{2\pi i(ux + vy + 0z)\} dx dy dz \\ &= \iint \varphi_3(x, y, z) dz \exp\{2\pi i(ux + vy)\} dx dy \\ &= \iint \varphi_2(x, y) \exp\{2\pi i(ux + vy)\} dx dy \\ &= \Phi_2(u, v). \end{aligned} \quad (2.5.6.7)$$

In the general case of projecting along the vector  $\boldsymbol{\tau}$ , the central section theorem is

$$\mathcal{F}[\varphi_2(\mathbf{x}_\tau)] = \Phi_3(\mathbf{u}_\tau), \quad \mathbf{u}_\tau \perp \boldsymbol{\tau}. \quad (2.5.6.8)$$

From this theorem it follows that the inversion of the 3D ray transform is possible if there is a continuous set of projections  $\varphi_\tau$  corresponding to the motion of the vector  $\boldsymbol{\tau}(\theta, \psi)$  over any continuous line connecting the opposite points on the unit sphere (Fig. 2.5.6.2) (Orlov's condition: Orlov, 1976). This result is evidenced by the fact that in this case the cross sections  $\mathcal{F}[\varphi_2(\mathbf{x}_\tau)]$  that are perpendicular to  $\boldsymbol{\tau}$  in Fourier space continuously cover the whole Fourier space, *i.e.*, they yield  $\mathcal{F}[\varphi_3(\mathbf{r})]$  and thereby determine  $\varphi_3(\mathbf{r}) = \mathcal{F}^{-1}[\Phi_3(\mathbf{u})]$ .

In single-particle reconstruction, imaged objects are randomly and nonuniformly oriented on the substrate at different angles (Frank, 2006) and the distribution of their orientations is beyond our control; therefore, the practical impact of Orlov's condition is limited. In fact, it is more useful to determine *a posteriori*, *i.e.*, after the 3D reconstruction of the macromolecule is computed, how well the Fourier space was covered. This can be done by calculating the distribution of the 3D spectral signal-to-noise ratio (Penczek, 2002).

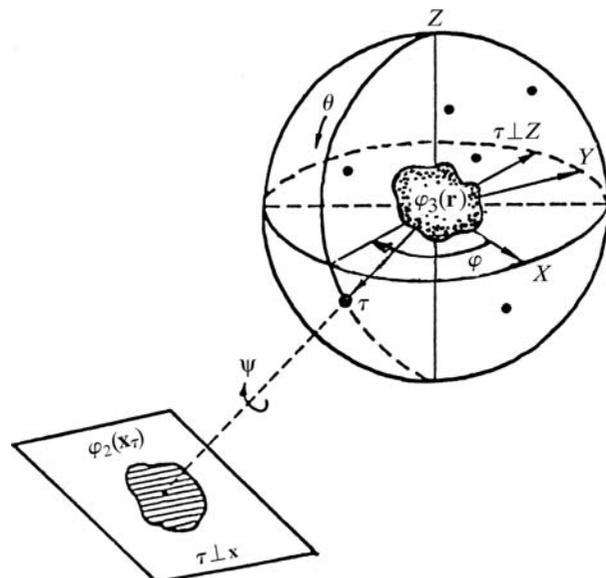


Fig. 2.5.6.2. The projection sphere and projection  $\varphi_2(\mathbf{x}_\tau)$  of  $\varphi_3(\mathbf{r})$  along  $\boldsymbol{\tau}$  onto the plane  $\boldsymbol{\tau} \perp \mathbf{x}$ . The case  $\boldsymbol{\tau} \perp Z$  represents orthoaxial projection. Points indicate an arbitrary distribution of projection directions  $\boldsymbol{\tau}$ .

### 2.5.6.2. 3D reconstruction in the general case

In the general case of the 3D reconstruction of  $\varphi_3(\mathbf{r})$  from projections  $\varphi_2(\mathbf{x}_\tau)$ , the projection vector  $\boldsymbol{\tau}$  occupies arbitrary positions on the projection sphere (Fig. 2.5.6.2). First, let us consider the case of a 2D function  $\rho_2(\mathbf{x})$  and its ray transform  $\varphi_1(x, \psi)$ . We introduce an operation of backprojection  $b$ , which is stretching along  $\boldsymbol{\tau}_\psi$  each 1D projection  $\varphi_1(x_\psi)$  (Fig. 2.5.6.3). When the result is integrated over the full angular range of projections  $\varphi_1(x, \psi)$ , we obtain the projection synthesis defined as

$$b(x, y) = \int_0^\pi \varphi_1(x \cos \psi + y \sin \psi, \psi) d\psi. \quad (2.5.6.9)$$

However, the backprojection operator is not the inverse of a 2D ray transform, as the resulting image  $b$  is blurred by the point-spread function  $(x^2 + y^2)^{-1/2}$  (Vainshtein, 1971):

$$b(x, y) = \rho(x, y) * (x^2 + y^2)^{-1/2}. \quad (2.5.6.10)$$

By noting that the Fourier transform of  $(x^2 + y^2)^{-1/2}$  is  $(u^2 + v^2)^{-1/2}$  and by using the convolution theorem  $\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]$ , we obtain the 'backprojection-filtering' inversion formula:

$$\begin{aligned} \rho(x, y) &= b(x, y) * (x^2 + y^2)^{1/2} = \mathcal{F}^{-1}[\mathbf{u}|\mathcal{F}[b]] \\ &= \text{Filtration}_{|\mathbf{u}|}[\text{Backprojection}(\varphi_1)]. \end{aligned} \quad (2.5.6.11)$$

The more commonly used 'filtered-backprojection' inversion is based on the 2D version of the central section theorem (2.5.6.8):

$$\mathcal{F}[\varphi_1(x_\psi)] = \Upsilon_2(\mathbf{u}_\psi) = \Upsilon_2(R, \Psi), \quad (2.5.6.12)$$

where  $\mathcal{F}[\rho_2] = \Upsilon_2$ . With this in mind,  $\rho_2(\mathbf{x})$  can be related to its ray transform by evaluating the Fourier transform of  $\rho_2$  in polar coordinates:

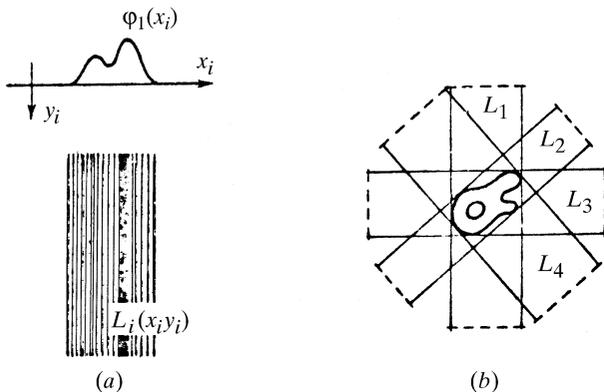


Fig. 2.5.6.3. (a) Formation of a backprojection function; (b) projection synthesis (2.5.6.9) is a superposition of these functions.

$$\begin{aligned}
 \rho_2(\mathbf{x}) &= \int \Upsilon_2(\mathbf{u}) \exp(-2\pi i \mathbf{u} \mathbf{x}) d\mathbf{u} \\
 &= \int_0^\pi \int_{-\infty}^\infty \Upsilon_2(R, \Psi) \exp(-2\pi i \mathbf{x} \boldsymbol{\tau}) |R| dR d\Psi \\
 &= \int_0^\pi \int_{-\infty}^\infty \mathcal{F}[\varphi_1(x_\psi)] \exp(-2\pi i \mathbf{x} \boldsymbol{\tau}) |R| dR d\Psi \\
 &= \int_0^\pi \mathcal{F}^{-1}[|R| \mathcal{F}[\varphi_1(x_\psi)]] d\Psi \\
 &= \text{Backprojection}[\text{Filtration}_{|R|}(\varphi_1)]. \quad (2.5.6.13)
 \end{aligned}$$

In three dimensions, the backprojection stretches each 2D projection  $\varphi_2[\mathbf{x}, \boldsymbol{\tau}(\theta, \psi)]$  along the projection direction  $\boldsymbol{\tau}(\theta, \psi)$ . A 3D synthesis is the integral over the hemisphere (Fig. 2.5.6.2)

$$b(\mathbf{r}) = \int_{\omega} \varphi_2(\mathbf{x}, \omega_\tau) d\omega_\tau = \varphi_3(\mathbf{r}) * (x^2 + y^2 + z^2)^{-1}. \quad (2.5.6.14)$$

Thus, in three dimensions the image  $b$  obtained using the backprojection operator is blurred by the point-spread function  $1/(x^2 + y^2 + z^2)$  (Vainshtein, 1971). It is possible to derive inversion formulae analogous to (2.5.6.11) and (2.5.6.13).

The inversion formulae demonstrate that it is possible to invert the ray transform for continuous functions and for a uniform distribution of projections. In electron microscopy, the projections are never distributed uniformly in three dimensions. Indeed, a uniform distribution is not even desirable, as only certain subsets of projection directions are required for the successful inversion of a 3D ray transform, as follows from the central section theorem (2.5.6.8). In practice, we always deal with sampled data and with discrete, random and nonuniform distributions of projection directions. Therefore, the inversion formulae can be considered only as a starting point for the development of the numerical (and practical!) reconstruction algorithms. According to (2.5.6.10) and (2.5.6.14), a simple backprojection step results in reconstruction that corresponds to a convolution of the original function with a point-spread function that depends only on the distribution of projections, but not on the structure itself. Taking into account the linearity of the backprojection operation, one has to conclude that for any practically encountered distribution of projections it should be possible to derive the corresponding point-spread function and then, using either deconvolution or Fourier filtration (with a ‘weighting function’), arrive at a good approximation of the structure. This observation forms the basis of the weighted backprojection algorithm (Section 2.5.6.5). Similarly, the central section theorem gives rise to direct Fourier inversion algorithms (Section 2.5.6.6). Nevertheless, since the data are discrete, the most straightforward methodology is to discretize and approach

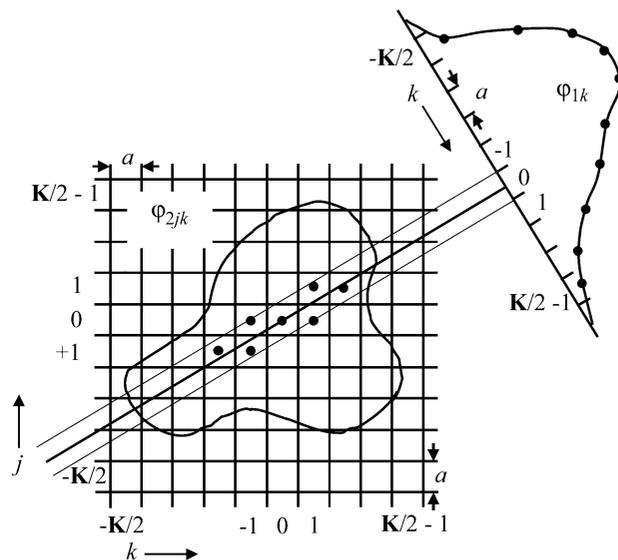


Fig. 2.5.6.4. Discretization in two dimensions ( $d = 2$ ). The assumption is that the mass is located at the centre of the voxel.

the reconstruction problem as an algebraic problem (Section 2.5.6.4).

### 2.5.6.3. Discretization and interpolation

In digital image processing, space is represented by a multi-dimensional discrete lattice. (It is sometimes expedient to use cylindrical or spherical coordinates, but these also have to be appropriately discretized.) The 2D projections  $\varphi_2(\mathbf{x})$  are sampled on a Cartesian grid  $\{\mathbf{k}a : \mathbf{k} \in \mathbf{Z}^d, (-\mathbf{K}/2) \leq \mathbf{k} \leq (\mathbf{K}/2)\}$ , where  $d$  is the dimensionality of the grid ( $d = 2$  for projections,  $d = 3$  for the reconstructed object),  $\mathbf{K} \in \mathbf{Z}_+^d$  is the size of the grid and  $a$  is the grid spacing (Fig. 2.5.6.4). In single-particle reconstruction, the units of  $a$  are usually ångströms and we also assume that the data were appropriately sampled, i.e.,  $a \leq (1/2u_{\max})$ . Thus, the pixel size is less than or equal to the inverse of twice the maximum spatial frequency present in the data. Since the latter is not known in advance, a more practical rule is to select the pixel size at about one third of the expected resolution of the final structure, so that the adverse effects of interpolation are reduced.

The input electron microscopy data (projections of the macromolecule) are discretized on a 2D Cartesian grid, but each projection has a particular orientation in polar coordinates. Except for a few cases (projection directions parallel to the axes of the coordinate system of the 3D structure), an interpolation is required to relate measured samples to the voxel (volume pixel) values on the 3D Cartesian grid of the reconstructed structure (Fig. 2.5.6.4). The step of backprojection can be visualized as a set of rays with base  $a^{d-1}$  extended from projections and the ray values being added to the intersected voxels on the grid of the reconstructed structure (Figs. 2.5.6.3 and 2.5.6.4). One can select schemes that aim at approximation of the physical reality of the data collection, for example to weight the contributions by the areas of the voxels intersected by the ray or by the lengths of the lines that transverse the voxel (Huesman *et al.*, 1977). In order to reduce the time of calculations, in electron microscopy one usually assumes that all the mass is located at the centre of the voxel and the additional accuracy is achieved by application of tri- (or bi-)linear interpolation. The exception is the algebraic reconstruction technique with blobs algorithm (Marabini *et al.*, 1998), where the voxels are represented by smooth spherically symmetric volume elements [for example, the Keiser–Bessel function (2.5.6.44)].

In real space, both the projection and backprojection steps can be implemented in two different ways: as voxel driven or as ray driven (Laurette *et al.*, 2000). If we consider a projection, in the

## 2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

voxel-driven approach the volume is scanned voxel by voxel. The nearest projection bin to the projection of each voxel is found, and the values in this bin and three neighbouring bins are increased by the corresponding voxel value multiplied by the weights calculated using bilinear interpolation. In the ray-driven approach, the volume is scanned along the projection rays. The value of the projection bin is increased by the values in the volume calculated in equidistant steps along the rays using trilinear interpolation. Because voxel- and ray-driven methods apply interpolation to projections or to voxels, respectively, the interpolation artifacts will be different in each case. Therefore, when calculating reconstructions using iterative algorithms that alternate between projection and backprojection steps, it is important to maintain consistency; that is, to use the same method for both steps. In either case, the computational complexity of each method is of the order of  $K^3$ , although the voxel-driven approach is faster due to the smaller number of neighbouring points used in the interpolation.

In the reconstruction methods based on the direct Fourier inversion of the 3D ray transform, the interpolation is performed in Fourier space. Unfortunately, it is difficult to design an accurate and fast interpolation scheme for the discrete Fourier space. Bilinear interpolation introduces local errors and when applied in real space it results in attenuation of high-frequency information. When applied in Fourier space, bilinear interpolation results in errors evenly spread over the whole frequency range, thus resulting in potentially severe nonlocal errors in real space. In order to eliminate this error it would be tempting to use interpolation based on Shannon's sampling theorem [Shannon, 1949; reprinted in *Proc. IEEE*, (1998), **86**, 447–457], which states that a properly sampled band-limited signal can be fully recovered from its discrete samples. For the signal represented by  $K^3$  equispaced Fourier samples  $\Phi_{3hkl}$ , the value of the Fourier transform  $\Phi_3$  at the arbitrary location  $(u, v, w)$  is given by (Crowther, Amos *et al.*, 1970)

$$\Phi_3(u, v, w) = \sum_{h=0}^{K-1} \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \Phi_{3hkl} w_h(u) w_k(v) w_l(w), \quad (2.5.6.15)$$

where (Yuen & Fraser, 1979; Lanzavecchia & Bellon, 1994)

$$w_k(u) = \begin{cases} \frac{\sin\{K\pi[u - (k/K)]\}}{\sin\{\pi[u - (k/K)]\}}, & K \text{ odd;} \\ \frac{\sin\{K\pi[u - (k/K)]\}}{\tan\{\pi[u - (k/K)]\}}, & K \text{ even.} \end{cases} \quad (2.5.6.16)$$

In cryo-EM, samples of  $\Phi_3$  are given at arbitrary 3D locations, as derived from Fourier transforms of 2D projection data (central section theorem) and one seeks to recover  $\Phi_{3hkl}$  on the 3D Cartesian grid. Upon the inverse Fourier transform, it will yield the reconstructed object. The problem can be solved as an overdetermined system of linear equations (Crowther, DeRosier & Klug, 1970). Indeed, if we write  $\Phi_3$  and  $\Phi_{3hkl}$  as 1D arrays  $\mathbf{\Phi}_3$  and  $\mathbf{\Phi}_{3(hkl)}$ , respectively (the former has  $K^2 \times$  [number of projections] elements, while the latter has  $K^3$  elements), and we denote by  $\mathbf{W}$  the appropriately dimensioned matrix of the interpolants, the least-squares solution is

$$\mathbf{\Phi}_{3(hkl)} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{\Phi}_3. \quad (2.5.6.17)$$

The above method is impractical because of the large size of the matrix  $\mathbf{W}$ . In some cases, when due to symmetries the projection data are distributed approximately evenly (as in the case of icosahedral structures), the problem can be solved to a good degree of accuracy by performing the interpolation (2.5.6.17)

independently along each of the three frequency axes (Crowther, DeRosier & Klug, 1970). Thus, in this case the solution to the problem of interpolation in Fourier space becomes a solution to the reconstruction problem.

For a more general single-particle reconstruction application a moving Shannon window interpolation has been proposed (Lanzavecchia & Bellon, 1994, 1998). It is based on an attenuated version of the window function and in one dimension has the form

$$\Phi_1(u) = \sum_{k=n}^{m+n-1} \Phi_{1k} \frac{\sin\{\pi[u - (k/K)]\}}{\sin\{(\pi/2)[u - (k/K)]\}} \times (\cos\{(\pi/2)[u - (k/K)]\})^A, \quad (2.5.6.18)$$

where  $n$  is the window size ( $n \ll K$ ) and  $A$  is an integer that is even for  $K$  odd and *vice versa*. In multidimensional cases, a product of  $\Phi$ 's from (2.5.6.18) is used. In the case of interpolation between equispaced samples of  $\Phi_{2hk}$ , excellent results have been reported for  $n = 11$  (Lanzavecchia *et al.*, 1996). However, the application to the reconstruction problem, *i.e.*, to resampling of the nonuniformly distributed Fourier data onto a 3D Cartesian grid, does not yield satisfactory results. Although general conditions under which interpolation using (2.5.6.18) can be done are known (Clark *et al.*, 1985), they are not met in practice and the results are at best nonexact. In addition, the relatively large window size required ( $n = 11$ ) results in impractical calculation times.

### 2.5.6.4. The algebraic and iterative methods

The algebraic methods have been derived based on the observation that when the projection equation (2.5.6.5) is discretized, it forms a set of linear equations. Thus, pixels from all available  $N$  projections are placed (in an arbitrary order) in a vector  $\varphi_{2jk}^n$ ,  $n = 1, \dots, N \rightarrow \mathbf{f}$  and the voxels of the 3D object in a vector  $\varphi_{3jkl} \rightarrow \mathbf{g}$  (in an order derived from the order of  $\mathbf{f}$  by algebraic relations). Note we left the exact sizes of  $\mathbf{f}$  and  $\mathbf{g}$  undetermined, as the major advantage of algebraic methods is that we can include in the reconstruction only pixels located within an arbitrary support in two dimensions and this support can be different for each projection; similarly, the support of the object in three dimensions can be arbitrary. Thus, the number of elements in  $\mathbf{f}$  and  $\mathbf{g}$  are at most  $K^2 N$  and  $K^3$ , respectively, with  $K^2$  being the number of pixels within chosen support. In the algebraic formulation the operation of projection is defined by the *projection matrix*  $\mathbf{P}$  whose elements  $p_{\eta\xi}$  are the interpolation weights. Their values are determined by the interpolation scheme used, but for the bi- and trilinear interpolations  $0 \leq p_{\eta\xi} \leq 1$ . The algebraic version of (2.5.6.5) is

$$\mathbf{f} = \mathbf{P}\mathbf{g}. \quad (2.5.6.19)$$

Matrix  $\mathbf{P}$  is rectangular and since in single-particle reconstruction the number of projections exceeds the linear size of the object ( $N \gg K$ ) the system of equations is overdetermined. It can be solved in a least-squares sense:

$$\tilde{\mathbf{g}} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{f}, \quad (2.5.6.20)$$

which yields a unique structure  $\tilde{\mathbf{g}}$  that corresponds to the minimum of  $|\mathbf{P}\mathbf{g} - \mathbf{f}|^2$ . As in the case of direct inversion in Fourier space (2.5.6.17), the approach (2.5.6.20) is impractical because of the very large size of the projection matrix. Indeed, the size of  $\mathbf{P}$  is  $K^2 N \times K^3$ , which for a modest number of projections  $N = 10\,000$  and image size  $K^2 = 64^2 = 4096$  yields a matrix of size  $\sim (4 \times 10^7) \times (2 \times 10^6)!$  Nevertheless, in some cases

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

the projections of the structure are orthoaxial, the full 3D reconstruction reduces to a series of independent 2D reconstructions, and it becomes possible to solve (2.5.6.19) by using the singular value decomposition (SVD) of the matrix  $\mathbf{P}$ :

$$\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (2.5.6.21)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  contain eigenvectors of matrices  $\mathbf{P}\mathbf{P}^T$  and  $\mathbf{P}^T\mathbf{P}$ , respectively, and  $\mathbf{\Sigma}$  is a diagonal matrix containing the first  $r$  nonnegative square roots of the eigenvalues of  $\mathbf{P}\mathbf{P}^T$  (Golub & Van Loan, 1996). The solution is given by

$$\mathbf{g} = \mathbf{V}\mathbf{\Sigma}_l^{-1}\mathbf{U}^T\mathbf{f}, \quad (2.5.6.22)$$

where  $\mathbf{\Sigma}_l^{-1}$  is the inverse of the matrix containing the  $l \leq r$  largest elements of  $\mathbf{\Sigma}$  with the remaining values set to zero. By selecting  $l$  appropriately we achieve a measure of regularization. The advantage of the approach is that for a given geometry the decomposition has to be calculated once; thus, the method becomes very efficient if the reconstruction has to be performed repeatedly for the same distribution of projections or if additional symmetries, such as helical, are taken into account (Holmes *et al.*, 2003).

In the general 3D case, a least-squares solution can be found using one of the iterative approaches that take advantage of the fact that the projection matrix is sparse. Indeed, if bilinear interpolation is used, a row of  $\mathbf{P}$  will contain only about  $4N$  nonzero elements (out of the total  $\sim K^2N$ ). The main idea is that the matrix  $\mathbf{P}$  is not explicitly calculated or stored; instead, its elements are calculated repeatedly during iterations as needed. In the simultaneous iterative reconstruction technique (SIRT) we find the minimum of

$$L(\mathbf{g}) = |\mathbf{P}\mathbf{g} - \mathbf{f}|^2 \quad (2.5.6.23)$$

by selecting the initial 3D structure  $\mathbf{g}^0$  (usually set to zero) and by iteratively updating its current approximation  $\mathbf{g}^{i+1}$  using the gradient of the objective function (2.5.6.23)  $\nabla L(\mathbf{g})$ ,

$$\mathbf{g}^{i+1} = \mathbf{g}^i - \lambda^i \mathbf{P}^T(\mathbf{P}\mathbf{g}^i - \mathbf{f}) = \mathbf{g}^i - \lambda^i(\mathbf{P}^T\mathbf{P}\mathbf{g}^i - \mathbf{P}^T\mathbf{f}). \quad (2.5.6.24)$$

Setting the relaxation parameter  $\lambda^i = \lambda = \text{constant}$  yields Richardson's algorithm (Gilbert, 1972). In single-particle reconstruction applications convergence (*i.e.*, a solution that is not dominated by the noise) is usually reached in  $\sim 100$  iterations.

SIRT is extensively used in single-particle reconstruction (Frank *et al.*, 1996) because it yields superior results over a wide range of experimental conditions (Penczek *et al.*, 1992) and, when presented with angular gaps in the distribution of projections, produces the least disturbing artifacts. Note that strictly speaking, filtered backprojection and direct Fourier inversion algorithms are not applicable to data that do not cover the Fourier space fully. In addition, SIRT offers considerable flexibility in EM applications. First, it is possible to accelerate the convergence by adjusting relaxation parameters: setting  $\lambda^i = \arg \min_{\lambda \geq 0} L(\mathbf{g}^i - \lambda \nabla L(\mathbf{g}^i))$  results in a steepest decent algorithm. Secondly, even faster convergence (in  $\sim 10$  iterations) is achieved by solving (2.5.6.23) using the conjugate-gradient method, but this requires addition of a regularizing term to (2.5.6.23) in order to prevent excessive enhancement of noise. Such a term has the form  $|\mathbf{B}\mathbf{g}|^2$ , where matrix  $\mathbf{B}$  is a discrete approximation of a Laplacian or higher-order derivatives. Thirdly, it is possible to take into account the underfocus settings of the microscope by including the contrast transfer function (CTF) of the microscope

in (2.5.6.23) and by solving the problem for the structure  $\mathbf{g}$  that in effect will be corrected for the CTF,

$$L(\mathbf{g}) = (1 - \eta) \sum_{\kappa} |\mathbf{S}_{\kappa}\mathbf{P}_{\kappa}\mathbf{g} - \mathbf{f}_{\kappa}|^2 + \eta|\mathbf{B}\mathbf{g}|^2. \quad (2.5.6.25)$$

Here we assumed that the projection data were grouped according to the defocus settings indexed by  $\kappa$  and we introduced a Lagrange multiplier  $\eta$  whose value determines the smoothness of the solution  $\mathbf{g}$  (Zhu *et al.*, 1997).  $\mathbf{S}_{\kappa}$  represents the space-invariant point-spread function of the microscope for the  $\kappa$ th defocus group; thus  $\mathbf{S}_{\kappa}$  has a block-Toeplitz structure (Biemond *et al.*, 1990) and  $\mathbf{S}_{\kappa} = \mathbf{S}_{\kappa}^T$ . If we assume that there is no astigmatism (which means that the point-spread function is rotationally invariant),  $\mathbf{S}_{\kappa}$  can be either applied to the projection of the structure or to the 3D structure itself. With these assumptions, (2.5.6.25) can be solved using the iterative scheme (Penczek *et al.*, 1997)

$$\mathbf{g}^{i+1} = \mathbf{g}^i - \lambda \left\{ (1 - \eta) \left[ \sum_{\kappa} (\mathbf{S}_{\kappa}^T \mathbf{P}_{\kappa}^T \mathbf{P}_{\kappa} \mathbf{S}_{\kappa} \mathbf{g}^i) - \sum_{\kappa} (\mathbf{S}_{\kappa}^T \mathbf{P}_{\kappa}^T \mathbf{f}_{\kappa}) \right] + \eta \mathbf{B}^T \mathbf{B} \mathbf{g}^i \right\}. \quad (2.5.6.26)$$

The second sum can be precalculated and stored as a 3D volume in the computer memory. Thus, the input projections have to be read only once and are never accessed again during the course of the iterations, which eliminates the need to store them in the memory [as is also the case in (2.5.6.24)]. In addition, the product  $\mathbf{B}^T\mathbf{B}$  is the 3D Laplacian, which can be applied efficiently without actually creating the matrix  $\mathbf{B}^T\mathbf{B}$ . Finally, because of the large size of the matrix  $\mathbf{S}_{\kappa}$  it is more convenient to apply it in Fourier space and to modify the Fourier transform of the volume instead of using matrix multiplication or real-space convolution. Algorithms (2.5.6.24)–(2.5.6.26) are implemented in the *SPIDER* package (Frank *et al.*, 1996) and (2.5.6.24) in the *SPARX* package (Hohn *et al.*, 2007).

The algebraic reconstruction technique (ART) predates SIRT; in the context of tomographic reconstructions it was proposed by Gordon and co-workers (Gordon *et al.*, 1970) and later it was recognized as a version of Kaczmarz's method for iteratively solving (2.5.6.23) (Kaczmarz, 1993). We write (2.5.6.19) as a set of systems of equations, each relating single pixels  $f_n$ ,  $n = 1, \dots, NK^2$  in projections with the 3D structure,

$$f_n = \mathbf{p}_n^T \mathbf{g}, \quad n = 1, \dots, NK^2. \quad (2.5.6.27)$$

Note (2.5.6.27) and (2.5.6.19) are equivalent, as  $\mathbf{f} = [f_1 f_2 \dots f_N]^T$  and  $\mathbf{P} = [\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_N]^T$ . With this notation and a relaxation parameter  $0 < \mu < 2$ , ART comprises the following steps:

- (1) set  $i = 0$  and the initial structure  $\mathbf{g}^0$ ;
- (2) for  $n = 1, \dots, NK^2$ , set

$$\mathbf{g}^{iNK^2+n} = \mathbf{g}^{iNK^2+n-1} - \mu(\mathbf{p}_n^T \mathbf{g}^{iNK^2+n-1} - f_n)(\mathbf{p}_n / |\mathbf{p}_n|); \quad (2.5.6.28)$$

- (3) set  $i = i + 1$ ; go to step (2).

Although the mathematical forms of update equations in SIRT (2.5.6.24) and in ART (2.5.6.28) are very similar, there are profound differences between them. In SIRT, all voxels in the structure are corrected simultaneously after projections and backprojections of the current update of the structure are calculated. In ART, the projection/backprojection step (2.5.6.28) involves only correction with respect to an individual pixel in a single projection immediately followed by the update of the structure. This results in a much faster convergence of ART as compared to SIRT. Further acceleration can be achieved by

selecting the order in which pixels enter the correction step (2) in (2.5.6.28). It was observed that if a pixel is selected such that its projection direction is perpendicular to the projection direction of the previous pixel, the convergence is achieved faster (Hamaker & Solmon, 1978; Herman & Meyer, 1993). Interestingly, a random order works almost equally well (Natterer & Wübbeling, 2001).

In single-particle reconstruction, ART has been introduced in the form of ‘ART with blobs’ (Marabini *et al.*, 1998) and is available in the *Xmipp* package (Sorzano *et al.*, 2004). In this implementation, the reconstruction structure is represented by a linear combination of spherically symmetric, smooth, spatially limited basis functions, such as Kaiser–Bessel window functions (Lewitt, 1990, 1992; Matej & Lewitt, 1996). Introduction of blobs significantly reduces the number of iterations necessary to reach an acceptable solution (Marabini *et al.*, 1998).

The major advantage of iterative reconstruction methods is the ability to take advantage of *a priori* knowledge, *i.e.*, any information about the protein structure that was not initially included in the data processing, and introduce it into the reconstruction process in the form of constraints. Examples of such constraints include similarity to the experimental (measured) data, positivity of the protein mass density (only valid in conjunction with the CTF correction), bounded spatial support *etc.* Formally, the process of enforcing selected constraints is best described in the framework of the projections onto convex sets (POCS) theory (Sezan & Stark, 1982; Youla & Webb, 1982; Sezan, 1992) introduced into EM by Carazo and co-workers (Carazo & Carrascosa, 1986, 1987; Carazo, 1992).

#### 2.5.6.5. Filtered backprojection

The method of filtered backprojection (FBP) is based on inversion formulae (2.5.6.11) (in two dimensions) or (2.5.6.14) (in three dimensions). It comprises the following steps: (i) a Fourier transform of each projection is computed; (ii) Fourier transforms of projections are multiplied by filters that account for a particular distribution of projections in Fourier space; (iii) the filtered projections are inversely Fourier transformed; (iv) real-space backprojection of processed projections yields the reconstruction. The method is particularly attractive due to the fact that the reconstruction calculated using simple real-space backprojection can be made efficient if the filter function is easy to compute.

In two dimensions with uniformly distributed projections the weighting function  $c(|R|, \Psi)$  in Fourier space is the ‘ramp function’  $|R|$  [(2.5.6.13)]. In two dimensions with nonuniformly distributed projections, when the analytical form of the distribution of projections is not known, an appropriate approximation to  $c(|R|, \Psi)$  has to be found. A good choice is to select weights such that the backprojection integral becomes approximated by a Riemann sum (Penczek *et al.*, 1996),

$$\begin{aligned} c(|R|, \Psi) = |R| dR d\Psi \rightarrow c(R_j, \Psi_i) &= R_j \frac{1}{2\pi} \frac{\Psi_{i+1} - \Psi_{i-1}}{2} \\ &= R_j \frac{\Delta\Psi_i}{4\pi}. \end{aligned} \quad (2.5.6.29)$$

For a given set of angles the weights  $c(R_j, \Psi_i)$  are easily computed (Fig. 2.5.6.5).

In three dimensions, the weighting (2.5.6.29) is applicable in a single-axis tilt data-collection geometry, where the 3D reconstruction can be calculated as a series of independent 2D reconstructions. In the general 3D case, the analogue of weighting (2.5.6.29) cannot be used, as the data are given in the form of 2D projections and it is not immediately apparent what fraction of the 3D Fourier volume is occupied by Fourier pixels in projections. However, the analogue of weighting (2.5.6.29) can be used in the inversion of 3D Radon transforms or in a direction

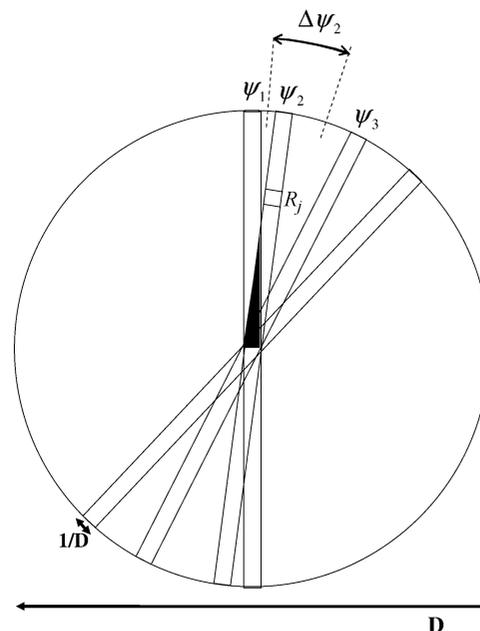


Fig. 2.5.6.5. Nonuniform distribution of projections. The projection weights for the reconstruction algorithms are chosen such that the backprojection integral becomes approximated by a Riemann sum and are equal to the angular length of an arc  $\Delta\Psi_i$  (2.5.6.29). In Fourier space, projections of an object with real-space radius  $D$  form rectangles with width  $1/D$ . In the exact filter backprojection reconstruction method, the weights are derived based on the amount the overlap between projections (2.5.6.29).

inversion of 3D ray transforms that is based on an intermediate step of converting 2D projection data to 1D projection data, as described in Section 2.5.6.6.

In order to arrive at a workable solution, the weighting functions applicable to 2D projections are constructed based on an explicitly or implicitly formulated concept of the ‘local density’ of projections. This concept was introduced by Bracewell (Bracewell & Riddle, 1967), who suggested for a 2D case of non-uniformly distributed projections a heuristic weighting function,

$$c(R_j, \Psi_i) = \frac{R_j}{\sum_l \exp[-\text{constant}(|\Psi_i - \Psi_l| \bmod \pi)^2]}. \quad (2.5.6.30)$$

The weighting function (2.5.6.30) can be easily extended to three dimensions; however, it has a major disadvantage that for a uniform distribution of projections it does not approximate well the weighting function (2.5.6.29), which we consider optimal.

Radermacher *et al.* (1986) proposed a derivation of a general weighting function using a deconvolution kernel calculated for a given (nonuniform) distribution of projections and, in modification of (2.5.6.14), a finite length of backprojection (Fig. 2.5.6.3). Such a ‘truncated’ backprojection is

$$\hat{b}_i(\mathbf{r}) = \varphi_2(\mathbf{x}_\tau) * l(\mathbf{r}), \quad \boldsymbol{\tau} \perp \mathbf{x} \quad (2.5.6.31)$$

with projection directions  $\boldsymbol{\tau}(\theta, \psi)_i$  and

$$\begin{aligned} l(\mathbf{r}) &= \delta(\mathbf{x}_\tau) t(z_\tau), \\ t(z_\tau) &= \begin{cases} 1 & -(D/2) \leq z \leq (D/2), \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \quad (2.5.6.32)$$

where  $D$  is the diameter of the object or the length of the backprojection  $l$ . By taking the Fourier transform of (2.5.6.31) and using the central section theorem (2.5.6.8), we obtain a 3D Fourier transform of the backprojected projection,

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

$$\Phi_3(u, v, w) = \Phi_2(\mathbf{u}_\tau)D \operatorname{sinc}(D\pi w_\tau), \quad (2.5.6.33)$$

and the 3D reconstruction is obtained by the inverse Fourier transform of the sum of contributions given by (2.5.6.33),

$$\mathcal{F}[\hat{\mathbf{b}}(\mathbf{r})] = \sum_i \Phi_2(\mathbf{u}_\tau)D \operatorname{sinc}(D\pi w_\tau). \quad (2.5.6.34)$$

$z_\tau$  and  $w_\tau$  are variables in real and Fourier spaces, respectively, both extending in the direction of the projection direction  $\tau(\theta, \psi)_i$ .

In this analysis, the transfer function of the backprojection algorithm is obtained by setting  $\Phi_2(\mathbf{u}_\tau) = 1$ , that is by finding the response of the algorithm to the input composed of delta functions in real space. This yields the inversion formula for a general weighted backprojection algorithm:

$$\varphi_3(\mathbf{r}) = \mathcal{F}[\hat{\mathbf{b}}(\mathbf{r})]c(u, v, w), \quad (2.5.6.35)$$

with the weighting function given by

$$c(u, v, w) = \frac{1}{\sum_i D \operatorname{sinc}(D\pi w_\tau)}. \quad (2.5.6.36)$$

The general weighting function (2.5.6.36) is consistent with analytical solutions (2.5.6.11) and (2.5.6.14), as it can be shown that by assuming infinite support ( $D \rightarrow \infty$ ) and continuous and uniform distribution of projection directions, in two dimensions one obtains  $c(u, v) = (u^2 + v^2)^{-1/2}$  (Radermacher, 2000).

The derivation of (2.5.6.36) is based on the analysis of continuous functions and its direct application to discrete data results in reconstruction artifacts; therefore, Radermacher (1992) proposed attenuating the sinc functions in (2.5.6.36) by exponent functions with decay depending on the diameter of the structure  $D$ , or simply replacing the sinc functions by exponent functions. This, however, reduces the concept of the weighting function corresponding to the deconvolution to the concept of the weighting function representing the ‘local density’ of projections (2.5.6.30). The general weighted backprojection reconstruction is implemented in *SPIDER* (using exponent-based weighting functions) (Frank *et al.*, 1996), *Xmipp* (Sorzano *et al.*, 2004) and *SPARX* (Hohn *et al.*, 2007).

Harauz & van Heel (1986) proposed basing the calculation of the density of projections, thus the weighting function, on the overlap of Fourier transforms of projections in Fourier space. Although the concept is general, it can be easily approached in two dimensions. If the diameter of the object is  $D$ , the width of a Fourier transform of a projection is  $2/D$  (Fig. 2.5.6.5), as follows from the central section theorem (2.5.6.8). Harauz and van Heel postulated that the weighting should be inversely proportional to the sum of geometrical overlaps between a given central section and the remaining central sections. For a pair of projections  $il$ , this overlap is

$$o_{il}(R) = T[DR \sin(\Psi_i - \Psi_l)], \quad (2.5.6.37)$$

where  $T$  represents the triangle function (selected by the authors because it can be calculated efficiently). Also, owing to the Friedel symmetry of central sections, the angles in (2.5.6.37) are restricted such that  $0 \leq \Psi_i - \Psi_l \leq (\pi/2)$ . In this formulation, the overlap is limited to the maximum frequency,

$$R_{il}^{\max} = \frac{2}{D \sin(\Psi_i - \Psi_l)}. \quad (2.5.6.38)$$

Thus, the overlap function becomes

$$o_{il}(R) = \begin{cases} 1 - (R/R_{il}^{\max}) & 0 \leq R < u_{il}^{\max} \\ 0 & R > u_{il}^{\max} \end{cases}. \quad (2.5.6.39)$$

In effect, the weighting function, called by the authors an ‘exact filter’, is

$$c(R, \Psi_i) = \frac{1}{1 + \sum_{l, l \neq i} o_{il}(R)}. \quad (2.5.6.40)$$

The weighting function (2.5.6.40) easily extends to three dimensions; however, the calculation of the overlap between central sections in three dimensions (represented by slabs) is more elaborate (Harauz & van Heel, 1986). The method is conceptually simple and computationally efficient. However, (2.5.6.40) does not approximate the correct weighting well for a uniform distribution of projections [*i.e.*, it should yield  $c(R, \Psi_i) = R$ ]. This, as can be seen by integrating (2.5.6.39) over the whole angular range, is not the case. The exact filter backprojection reconstruction is implemented in the *IMAGIC* (van Heel *et al.*, 1996), *SPIDER* (Frank *et al.*, 1996) and *SPARX* (Hohn *et al.*, 2007) packages.

The 3D reconstruction methods based on filtered backprojection are commonly used in single-particle reconstruction. The reasons are: their versatility, ease of implementation, and – in comparison with iterative methods – good computational efficiency. Unlike in iterative methods, there are no parameters to adjust, although it has been noted that the results depend on the value of the diameter  $D$  of the structure in all three weighting functions [(2.5.6.30), (2.5.6.36) and (2.5.6.38)], so the performance of the reconstruction algorithm can be optimized for a particular data-collection geometry by changing the value of  $D$  (Paul *et al.*, 2004). However, because computation of the weighting function involves calculation of pairwise distances between projections, the computational complexity is proportional to the square of the number of projections and for large data sets these methods become inefficient. It also has to be noted that the weighting functions (2.5.6.30), (2.5.6.36) and (2.5.6.40) remain approximations of the correct weighting function (2.5.6.29).

### 2.5.6.6. Direct Fourier inversion

Direct Fourier methods are based on the central section theorem (2.5.6.8). A set of the 2D Fourier transforms of projections yields an approximation to  $\Phi_3$  on a nonuniform 3D grid, and a subsequent numerical 3D inverse Fourier transform gives an approximation to  $\varphi_3$ . If the 3D inverse Fourier transform could be realized by means of the 3D inverse fast Fourier transform (FFT), one would have a very fast reconstruction algorithm. Unfortunately, the preprocessing step yields  $\Phi_3$  on a nonuniform grid. In effect, the 3D inverse FFT is not applicable and an additional step of recovering samples of  $\Phi_3$  on a uniform grid from the available samples on a nonuniform grid is necessary.

One possibility is to resample the nonuniformly sampled version of  $\Phi_3$  onto a 3D Cartesian grid by some form of interpolation. For example, Grigorieff used a modified trilinear interpolation scheme in the *FREALIGN* package (Grigorieff, 1998). Simple interpolation methods have been found to give inaccurate results, although more sophisticated interpolation schemes can go a long way to improve the accuracy (Lanza-

## 2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

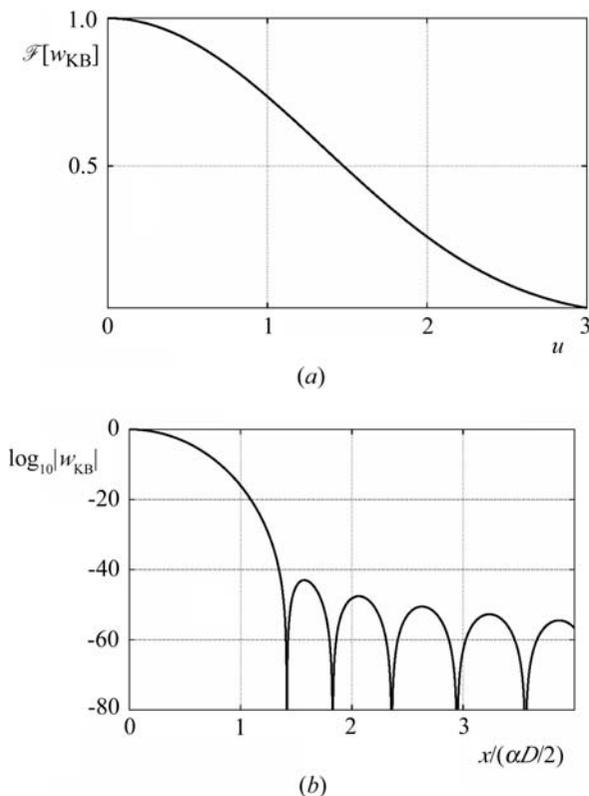


Fig. 2.5.6.6. Kaiser–Bessel window function used in the gridding reconstruction algorithm GDFR. (a) In Fourier space, the window function is effectively zero outside the support of six Fourier pixels. (b) In real space, the zeros of the window function are beyond the radius of the reconstructed object.

vecchia *et al.*, 1993). Unfortunately, the recommended window size makes them impractical for most applications.

The most accurate Fourier reconstruction methods are those that employ nonuniform Fourier transforms, particularly the 3D gridding method (O’Sullivan, 1985; Schomberg & Timmer, 1995). The gridding-based direct Fourier reconstruction algorithm (GDFR) (Penczek *et al.*, 2004) was developed specifically for single-particle reconstruction. It comprises three steps:

(1) The first step, known as ‘gridding’, involves calculating for the Fourier transform of each projection the convolution

$$\sum_i \mathcal{F}[w] * (c\mathcal{F}[\varphi_2]), \quad (2.5.6.41)$$

where  $c$  are ‘gridding weights’ designed to compensate for the nonuniform distribution of the grid points and  $\mathcal{F}[w]$  is an appropriately chosen convolution kernel. After processing all projections, this step yields samples of  $\mathcal{F}[w] * \mathcal{F}[\varphi_3]$  on a Cartesian grid.

(2) 3D inverse FFT is used to compute

$$w\varphi_3 = \mathcal{F}^{-1}[\mathcal{F}[w] * \mathcal{F}[\varphi_3]] \quad (2.5.6.42)$$

on a Cartesian grid.

(3) The weights are removed using division:

$$\varphi_3 = w\varphi_3/w. \quad (2.5.6.43)$$

The method involves a number of parameters. First, we need to decide the oversampling factor for the padding with zeros of input projections before FFTs are computed. In GDFR, the

oversampling is set to two, although smaller factors can also yield good results. Second, we need a *window function*  $\mathcal{F}[w]$  whose support in Fourier space is ‘small’. In order to assure good computational efficiency of the convolution step (2.5.6.41), in GDFR this support was set to six Fourier voxels. In addition, in order to prevent division by zeros in (2.5.6.43), the *weighting function*  $w = \mathcal{F}^{-1}[\mathcal{F}[w]]$  must be positive within the support of the reconstructed object. A recommendable window function is the separable, bell-shaped Kaiser–Bessel window (O’Sullivan, 1985; Jackson *et al.*, 1991; Schomberg & Timmer, 1995) (Fig. 2.5.6.6a):

$$\mathcal{F}[w_{\text{KB}}(\mathbf{u})] = \begin{cases} \prod_{v=1}^d \frac{I_0\{\pi D\alpha s_v [1 - (u_v/s_v)^2]^{1/2}\}}{2s_v}, & \mathbf{u} \in [-s, s], \\ 0, & \mathbf{u} \notin [-s, s], \end{cases} \quad (2.5.6.44)$$

where  $I_0$  is the zero-order modified Bessel function and  $\alpha = 1.25$  is a parameter. The weighting function associated with  $\mathcal{F}[w_{\text{KB}}]$  is

$$w_{\text{KB}}(\mathbf{x}) = \prod_{v=1}^d \frac{\sinh[\pi D\alpha s_v (1 - \{x_v/[\alpha(D/2)]^2\})^{1/2}]}{\pi D\alpha s_v (1 - \{x_v/[\alpha(D/2)]^2\})^{1/2}}. \quad (2.5.6.45)$$

Finally, the gridding weights  $c$  are chosen such that in discrete implementation of (2.5.6.41) and (2.5.6.42) we obtain a Riemann sum approximating the respective integral (Schomberg, 2002). The Voronoi diagram (Okabe *et al.*, 2000) of the sampling points provides a good partitioning of sampling space, so the gridding weights  $c$  are computed by constructing a Voronoi diagram of the grid points and by choosing the weights as the volumes of the Voronoi cells (Fig. 2.5.6.7a). In cryo-EM, the number of projections, thus the number of sampling points in Fourier space, is extremely large. Thus, although the calculation of the gridding weights *via* the 3D Voronoi diagram of the nonuniformly spaced grid points for  $\mathcal{F}[\varphi_3]$  would lead to an accurate direct Fourier method, the method would be very slow and would require excessive computer memory. To circumvent this problem, in GDFR (Penczek *et al.*, 2004) the 2D reverse gridding method is used to compute the Fourier transform of each projection on a 2D polar grid. In this way,  $\mathcal{F}[\varphi_3]$  is obtained on a 3D spherical grid, where the grid points are located both on centred spheres and on straight lines through the origin. Accordingly, it becomes possible to partition the sampled region into suitable sampling cells *via* the computation of a 2D Voronoi diagram on a unit sphere, rather than a 3D Voronoi diagram in Euclidean space (Fig. 2.5.6.7b). This significantly reduces the memory requirements and improves the computational efficiency of the method, particularly when a fast  $O(n \log n)$  algorithm for computing the Voronoi diagram on the unit sphere is employed (Renka, 1997).

The gridding weights in GDFR correspond to the weighting functions in filtered backprojection algorithms. Moreover, setting the gridding weights to be equal to angular regions on the unit sphere directly corresponds to the equating of weighting function (2.5.6.29) in 2D filtered backprojection to the length of an arc on a unitary circle. Thus, both methods yield nearly optimum projection density functions.

The reversed gridding method is obtained by reversing the sequence of steps (1)–(3) of the gridding method:

(1) the input image is divided by the weighting function  $\varphi_2/w$ ;

(2) the image is padded with zeros and 2D FFT is used to compute  $\mathcal{F}[\varphi_2/w]$ ;

(3) gridding is used to compute  $\mathcal{F}[w] * \mathcal{F}[\varphi_2/w]$  on an arbitrary nonuniform grid.

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

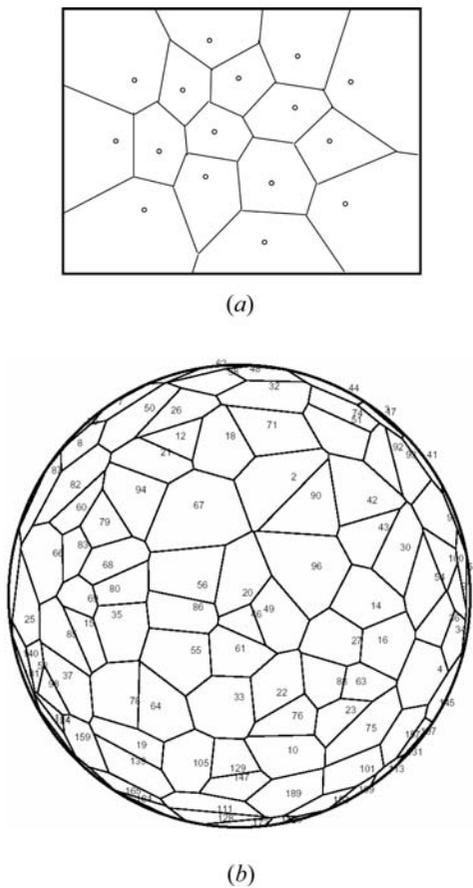


Fig. 2.5.6.7. Partitions of the sampling Fourier space using Voronoi diagrams in direct inversion algorithms. (a) Central section of a 3D Fourier volume with sampling points originating from 2D Fourier transforms of projections. Although 2D projections are sampled on a uniform (Cartesian) grid, the arbitrary rotations of projections in 3D space yields a nonuniform distribution of points in three dimensions. In effect, the 3D reconstruction by direct inversion using 3D FFT is not possible. (b) Voronoi diagram on a sphere in the GDFR algorithm. Using the reverse gridding method, the 2D Fourier transforms of projections are resampled onto 1D central lines using a constant angular step. In 3D Fourier space, they are located on central sections and their angular directions are evenly distributed on grand circles. However, since central sections have nonuniform distributions, the distribution of angular directions (sampling points on the unitary sphere) is also nonuniform and effectively random.

Note the reverse gridding does not require explicit gridding weights, as in the third step they are constant (Penczek *et al.*, 2004).

In GDFR, the third step of the reversed gridding results in a set of 1D Fourier central lines  $\mathcal{F}[\varphi_{1l}]$  calculated using a constant angular step. Clearly, upon inverse Fourier transform they amount to a Radon transform  $\varphi_1(u, \psi)$  of the projection and, upon repeating the process for all available projections, they yield a Radon transform  $\varphi_1(u, \theta)$  of a 3D object  $\varphi_3$  (albeit non-uniformly sampled with respect to Eulerian angles). Thus GDFR, in addition to being a method of inverting a 3D ray transform, is also a highly accurate method of inverting a 3D Radon transform. GDFR is implemented in the *SPIDER* package (Frank *et al.*, 1996).

The results of a comparison of selected reconstruction algorithms are shown in Fig. 2.5.6.8. The tests were performed using simulated data with the projections of a phantom 3D structure calculated using the inverse gridding method (Penczek *et al.*, 2004). The GDFR yields a virtually perfect reconstruction that agrees with the phantom over the whole frequency range (with the exception of the highest frequencies, but these cannot be reproduced due to geometrical limitations). GD3D is also a gridding-based reconstruction algorithm, in which the gridding weights are calculated directly from contributions of the

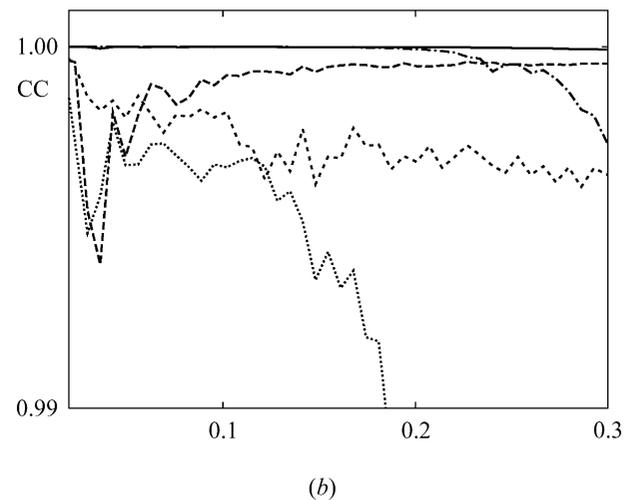
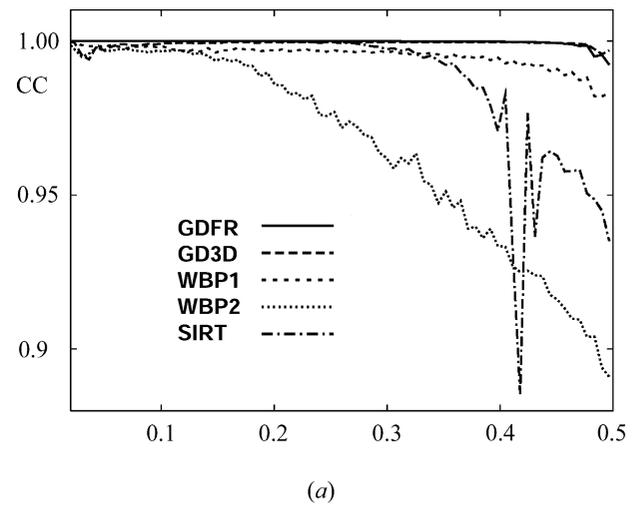


Fig. 2.5.6.8. (a) Plot of correlation coefficients (CCs) calculated between Fourier transforms of the reconstructed structure and the original phantom as a function of the magnitude of the spatial frequency using five reconstruction algorithms. GDFR: gridding direct Fourier reconstruction with Voronoi weights; GD3D: gridding reconstruction with simplified gridding weights; WBP1: general weighted backprojection with exponent-based weighting function; WBP2: exact filter weighted backprojection; SIRT: simultaneous iterative reconstruction algorithm. Noise-free projection data were computed in Fourier space using the reverse gridding method. (b) Rescaled version of the low-frequency range of (a). Note the different scales in (a) and (b). The horizontal axis is scaled in absolute frequency units with 0.5 equal to the Nyquist frequency.

weighting function to the 3D Fourier pixels and are applied to the voxels in the 3D Fourier volume (Jackson *et al.*, 1991). This yields an approximation to the 'local density' of contributing projections. The simplified gridding weights in GD3D result in deterioration of the reconstruction in the low- and intermediate-frequency ranges. The general weighted backprojection with exponent-based weighting function (WBP1) performs quite well, although reproduction in the low-frequency range is inferior. Similar artifacts are present in the reconstruction using the exact filter weighted backprojection (WBP2), which, in addition, performs disappointingly at higher frequencies. This relatively poor performance is attributed to the nonoptimal weighting schemes used in both methods. It is also interesting to note that the backprojection step is identical in both algorithms; they only differ by the weighting function. The significant difference in their performance attests to the importance of good weighting schemes for high-quality 3D reconstructions from nonuniformly distributed projections. SIRT yields a reconstruction that in the low- and intermediate-frequency ranges matches in quality the reconstruction obtained with GDFR. However, there is a significant loss of quality at high frequencies. This is due to the

## 2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

inferior linear interpolation scheme used in SIRT (this impediment is shared with WBP1 and WBP2). [The dip of the SIRT fidelity curve at a spatial frequency of 0.42 in Fig. 2.5.6.8(a) is due to an inconsistency between the interpolation methods used internally in SIRT to generate intermediate projections of the object to be reconstructed and the Fourier-space-based method used to generate the test data.] In terms of computational efficiency, gridding-based algorithms outperform weighted back-projection algorithms by a small factor while SIRT is approximately ten times slower (depending on the number of iterations used).

### 2.5.6.7. 3D reconstruction of symmetric objects

Many objects imaged by EM have symmetries; the two types that are often met are helical (phage tails, *F*-actin, microtubules, myosin thick filaments, and bacterial pili and flagella) and point-group symmetries (many macromolecular assemblies, virus capsids). If the object has helical symmetry, it is convenient to use cylindrical coordinates and a dedicated reconstruction algorithm; particularly for the initial analysis of the data that has to be done in Fourier space. Diffraction from such structures with  $c$  periodicity and scattering density  $\varphi(r, \psi, z)$  is defined by the Fourier–Bessel transform:

$$\begin{aligned} \Phi(R, \Psi, Z) &= \sum_{-\infty}^{+\infty} \exp\left[in\left(\Psi + \frac{\pi}{2}\right)\right] \int_0^{\infty} \int_0^{2\pi} \int_0^l \varphi(r, \psi, z) \\ &\quad \times J_n(2\pi r R) \exp[-i(n\psi + 2\pi z Z)] r \, dr \, d\psi \, dz \\ &= \sum_n G_n(R, Z) \exp\left[in\left(\Psi + \frac{\pi}{2}\right)\right]. \end{aligned} \quad (2.5.6.46)$$

The inverse transform has the form

$$\rho(r, \psi, z) = \sum_n \int g_n(r, Z) \exp(in\psi) \exp(2\pi iz Z) \, dZ, \quad (2.5.6.47)$$

so that  $g_n$  and  $G_n$  are the mutual Bessel transforms

$$G_n(R, Z) = \int_0^{\infty} g_n(r, Z) J_n(2\pi r R) 2\pi r \, dr \quad (2.5.6.48)$$

and

$$g_n(r, Z) = \int_0^{\infty} G_n(R, Z) J_n(2\pi r R) 2\pi R \, dR. \quad (2.5.6.49)$$

Owing to helical symmetry, (2.5.6.48) and (2.5.6.49) contain only those of the Bessel functions that satisfy the selection rule (Cochran *et al.*, 1952)

$$l = mp + (nq/N), \quad (2.5.6.50)$$

where  $N$ ,  $q$  and  $p$  are the helix symmetry parameters,  $m = 0, \pm 1, \pm 2, \dots$ . Each layer  $l$  is practically determined by the single function  $J_n$  with the lowest  $n$ ; the contributions of other functions are neglected. Thus, the Fourier transform of one projection of a helical structure, with an account of symmetry and phases, gives the 3D transform (2.5.6.49). However, biological specimens tend to be flexible and disordered, and exact helical symmetry is rarely observed. A possible approach to dealing with flexibility is to computationally straighten filaments (Egelman, 1986), but this has the potential for introducing artifacts. Another difficulty with helical analysis is that the indexing of a pattern can

be ambiguous and the wrong symmetry can be chosen (Egelman & Stasiak, 1988). Further complications exist when the filament does not have a precisely defined helical symmetry, such as *F*-actin, which has a random variable twist (Egelman *et al.*, 1982). To address these problems, Egelman developed a real-space refinement method for the reconstruction of helical filaments that is capable of determining the helical symmetry of an unknown structure (Egelman, 2000). In this approach, the Fourier–Bessel reconstruction algorithm is replaced by a general reconstruction algorithm discussed in Section 2.5.6.6 with the real-space projections supplied as input. Thus, the symmetry is not enforced within the reconstruction algorithm; instead, it is determined and imposed subsequently by real-space averaging.

The presence of point-group symmetry in the structure means that any projection yields as many symmetry-related (but differently oriented) copies of itself as the number of symmetry operations in the group. For example, for three dimensions, each projection enters the reconstruction process at six different projection directions, while for icosahedral symmetry ( $I$ ) the number is sixty! This high multiplicity makes it possible to write a reconstruction program that explicitly takes into account the symmetry and can perform the task much faster than generic algorithms. Such programs are usually part of dedicated software packages that were specifically designed for the determination of icosahedral structures (Fuller *et al.*, 1996; Lawton & Prasad, 1996; Liang *et al.*, 2002). These icosahedral reconstruction programs are heavily indebted to a set of FORTRAN routines written by R. A. Crowther at the Medical Research Council Laboratory of Molecular Biology (MRC LMB) and perform the 3D reconstruction using the Fourier–Bessel transform strategy outlined above [(2.5.6.46)–(2.5.6.49)] (Crowther, Amos *et al.*, 1970; Crowther & Amos, 1972).

The general reconstruction methods (algebraic, filtered back-projection, direct Fourier inversion) easily accommodate symmetries in the data. In major single-particle reconstruction packages, reconstruction programs are implemented such that the point-group symmetry is a parameter of the program. The symmetry operation is internally taken into account during calculation of the weighting function and it is also applied to the set of Eulerian angles assigned to each projection so the multiple copies are implicitly created and processed. This approach results in an extended time of calculations, but it is entirely general. In direct Fourier inversion algorithms the numerical inaccuracy of the symmetrization performed in Fourier space will result in nonsymmetric artifacts in real space. Thus, in *SPIDER* (Frank *et al.*, 1996) an additional real-space symmetrization is performed after the reconstruction is completed. Finally, it has to be noted that although it might be tempting to calculate a 3D reconstruction without enforcement of symmetry and to symmetrize the resulting structure, this approach is incorrect. This fact can be seen from the way weighting functions are constructed: weighting functions calculated with symmetries taken into account are not equal to the weighting function calculated for a unique set of projections and subsequently symmetrized.

### 2.5.7. Single-particle reconstruction

BY P. A. PENCZEK

#### 2.5.7.1. Formation of projection images in single-particle reconstruction

Cryo-electron microscopy (cryo-EM) in combination with the single-particle approach is a new method of structure determination for large macromolecular assemblies. Currently, resolution in the range 10 to 30 Å can be reached routinely, although in a number of pilot studies it has been possible to obtain structures at 4 to 8 Å. Theoretically, electron microscopy can yield data