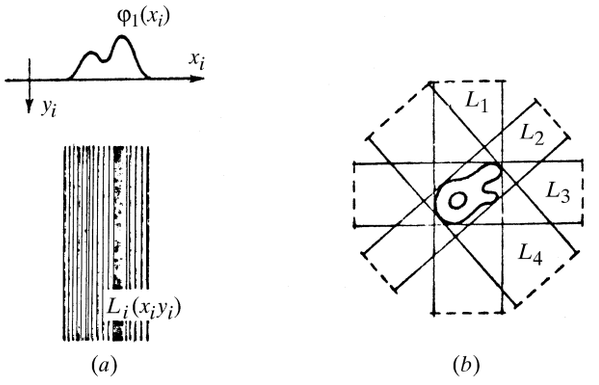

Fig. 2.5.6.3. (*a*) Formation of a backprojection function; (*b*) projection synthesis (2.5.6.9) is a superposition of these functions.

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

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

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

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

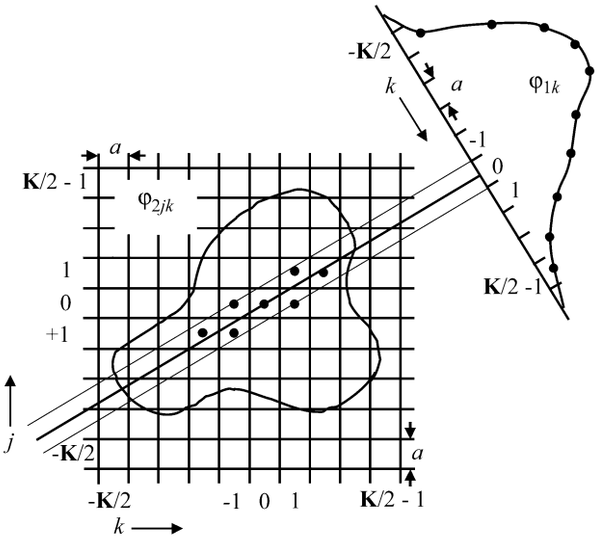

Fig. 2.5.6.4. Discretization in two dimensions ($d = 2$). The assumption is that the mass is located at the centre of the voxel.

2.5.6.3. Discretization and interpolation

In digital image processing, space is represented by a multidimensional 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 ångstroms 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

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 Φ_{3hkl}, the value of the Fourier transform Φ_3 at the arbitrary location (u, v, w) is given by (Crowther, Amos *et al.*, 1970)

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

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

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

In cryo-EM, samples of Φ_3 are given at arbitrary 3D locations, as derived from Fourier transforms of 2D projection data (central section theorem) and one seeks to recover Φ_{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 Φ_3 and Φ_{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

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

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

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

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 Φ's from (2.5.6.18) is used. In the case of interpolation between equispaced samples of Φ_{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 $\varphi_{2jk}^n, n = 1, \ldots, N \rightarrow \mathbf{f}$ 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^2N 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

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

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:

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

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 $K^2N \times K^3$, 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