

## 1. GENERAL RELATIONSHIPS AND TECHNIQUES

degree  $n$ , then  $P(\rho)$  should be sampled at intervals of at most  $\Delta/2n$  to accommodate the full bandwidth of its spectrum.

## 1.3.4.4.3.3. Solvent flattening

Crystals of proteins and nucleic acids contain large amounts of mother liquor, often in excess of 50% of the unit-cell volume, occupying connected channels. The well ordered electron density  $\rho_M(\mathbf{x})$  corresponding to the macromolecule thus occupies only a periodic subregion  $\mathcal{U}$  of the crystal. Thus

$$\rho_M = \chi_{\mathcal{U}} \times \rho_M,$$

implying the convolution identity between structure factors (Main & Woolfson, 1963):

$$F_M(\mathbf{h}) = \sum_{\mathbf{k}} \bar{\mathcal{F}} \left[ \frac{1}{\mathcal{U}} \chi_{\mathcal{U}} \right] (\mathbf{h} - \mathbf{k}) F_M(\mathbf{k})$$

which is a form of the Shannon interpolation formula (Sections 1.3.2.7.1, 1.3.4.2.1.7; Bricogne, 1974; Colman, 1974).

It is often possible to obtain an approximate ‘molecular envelope’  $\mathcal{U}$  from a poor electron-density map  $\rho$ , either interactively by computer graphics (Bricogne, 1976) or automatically by calculating a moving average of the electron density within a small sphere  $S$ . The latter procedure can be implemented in real space (Wang, 1985). However, as it is a convolution of  $\rho$  with  $\chi_S$ , it can be speeded up considerably (Leslie, 1987) by computing the moving average  $\rho_{\text{mav}}$  as

$$\rho_{\text{mav}}(\mathbf{x}) = \bar{\mathcal{F}}[\bar{\mathcal{F}}[\rho] \times \bar{\mathcal{F}}[\chi_S]](\mathbf{x}).$$

This remark is identical in substance to Booth’s method of computation of ‘bounded projections’ (Booth, 1945a) described in Section 1.3.4.2.1.8, except that the summation is kept three-dimensional.

The iterative use of the estimated envelope  $\mathcal{U}$  for the purpose of phase improvement (Wang, 1985) is a submethod of the previously developed method of molecular averaging, which is described below. Sampling rules for the Fourier analysis of envelope-truncated maps will be given there.

## 1.3.4.4.3.4. Molecular averaging by noncrystallographic symmetries

Macromolecules and macromolecular assemblies frequently crystallize with several identical subunits in the asymmetric metric unit, or in several crystal forms containing the same molecule in different arrangements. Rossmann & Blow (1963) recognized that intensity data collected from such structures are redundant (Sayre, 1952b) and that their redundancy could be a source of phase information.

The phase constraints implied by the consistency of geometrically redundant intensities were first derived by Rossmann & Blow (1963), and were generalized by Main & Rossmann (1966). Crowther (1967, 1969) reformulated them as linear eigenvalue equations between structure factors, for which he proposed an iterative matrix solution method. Although useful in practice (Jack, 1973), this reciprocal-space approach required computations of size  $\propto N^2$  for  $N$  reflections, so that  $N$  could not exceed a few thousands.

The theory was then reformulated in real space (Bricogne, 1974), showing that the most costly step in Crowther’s procedure could be carried out much more economically by averaging the electron densities of all crystallographically independent subunits, then rebuilding the crystal(s) from this averaged subunit, flattening the density in the solvent region(s) by resetting it to its

average value. This operation is a *projection* [by virtue of Section 1.3.4.2.2.2(d)]. The overall complexity was thus reduced from  $N^2$  to  $N \log N$ . The design and implementation of a general-purpose program package for averaging, reconstructing and solvent-flattening electron-density maps (Bricogne, 1976) led rapidly to the first high-resolution determinations of virus structures (Bloomer *et al.*, 1978; Harrison *et al.*, 1978), with  $N \sim 200\,000$ .

The considerable gain in speed is a consequence of the fact that the masking operations used to retrieve the various copies of the common subunit are carried out by simple *pointwise multiplication* by an indicator function  $\chi_U$  in real space, whereas they involve a *convolution* with  $\bar{\mathcal{F}}[\chi_U]$  in reciprocal space.

The averaging by noncrystallographic symmetries of an electron-density map calculated by FFT – hence sampled on a grid which is an integral subdivision of the period lattice – necessarily entails the interpolation of densities at nonintegral points of that grid. The effect of interpolation on the structure factors recalculated from an averaged map was examined by Bricogne (1976). This study showed that, if linear interpolation is used, the initial map should be calculated on a fine grid, of size  $\Delta/5$  or  $\Delta/6$  at resolution  $\Delta$  (instead of the previously used value of  $\Delta/3$ ). The analysis about to be given applies to all interpolation schemes which consist in a *convolution* of the sampled density with a fixed *interpolation kernel* function  $K$ .

Let  $\rho$  be a  $\mathbb{Z}^3$ -periodic function. Let  $K$  be the interpolation kernel in ‘normalized’ form, *i.e.* such that  $\int_{\mathbb{R}^3} K(\mathbf{x}) d^3\mathbf{x} = 1$  and scaled so as to interpolate between sample values given on a unit grid  $\mathbb{Z}^3$ ; in the case of linear interpolation,  $K$  is the ‘trilinear wedge’

$$K(\mathbf{x}) = W(x)W(y)W(z),$$

where

$$W(t) = 1 - |t| \quad \text{if } |t| \leq 1, \\ = 0 \quad \text{if } |t| \geq 1.$$

Let  $\rho$  be sampled on a grid  $\mathcal{G}_1 = \mathbf{N}_1^{-1}\mathbb{Z}^3$ , and let  $I_{\mathbf{N}_1}\rho$  denote the function interpolated from this sampled version of  $\rho$ . Then:

$$I_{\mathbf{N}_1}\rho = \left[ \rho \times \sum_{\mathbf{m} \in \mathbb{Z}^3} \delta_{(\mathbf{N}_1^{-1}\mathbf{m})} \right] * [(\mathbf{N}_1^{-1})^\# K],$$

where  $[(\mathbf{N}_1^{-1})^\# K](\mathbf{x}) = K(\mathbf{N}_1\mathbf{x})$ , so that

$$\begin{aligned} \bar{\mathcal{F}}[I_{\mathbf{N}_1}\rho] &= \left[ \bar{\mathcal{F}}[\rho] * \sum_{\mathbf{k}_1 \in \mathbb{Z}^3} \delta_{(\mathbf{N}_1^T \mathbf{k}_1)} \right] \\ &\times \left[ \frac{1}{|\det \mathbf{N}_1|} (\mathbf{N}_1^T)^\# \bar{\mathcal{F}}[K] \right] \\ &= \left[ \sum_{\mathbf{k}_1 \in \mathbb{Z}^3} \tau_{\mathbf{N}_1^T \mathbf{k}_1} \bar{\mathcal{F}}[\rho] \right] \times (\mathbf{N}_1^T)^\# \bar{\mathcal{F}}[K]. \end{aligned}$$

The transform of  $I_{\mathbf{N}_1}\rho$  thus consists of

(i) a ‘main band’ corresponding to  $\mathbf{k}_1 = \mathbf{0}$ , which consists of the true transform  $\bar{\mathcal{F}}[\rho](\xi)$  attenuated by multiplication by the central region of  $\bar{\mathcal{F}}[K][(\mathbf{N}_1^{-1})^T \xi]$ ; in the case of linear interpolation, for example,

$$\bar{\mathcal{F}}[K](\xi, \eta, \zeta) = \left( \frac{\sin \pi \xi}{\pi \xi} \right)^2 \left( \frac{\sin \pi \eta}{\pi \eta} \right)^2 \left( \frac{\sin \pi \zeta}{\pi \zeta} \right)^2;$$