1. GENERAL RELATIONSHIPS AND TECHNIQUES

degree n, then $P(\wp)$ 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 $\wp_{\mathrm{M}}(\mathbf{x})$ corresponding to the macromolecule thus occupies only a periodic subregion $\mathscr{U}$ of the crystal. Thus

$$\wp_{\mathrm{M}} = \chi_{\mathscr{U}} \times \wp_{\mathrm{M}},$$

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

$$F_{\mathrm{M}}(\mathbf{h}) = \sum_{\mathbf{k}} \bar{\mathscr{F}}\left[\frac{1}{\mathscr{U}}\chi_{\mathscr{U}}\right](\mathbf{h} - \mathbf{k})F_{\mathrm{M}}(\mathbf{k})$$

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' $\mathscr{U}$ from a poor electron-density map $\wp$, 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 $\wp$ with χ_S, it can be speeded up considerably (Leslie, 1987) by computing the moving average $\wp_{\mathrm{mav}}$ as

$$\wp_{\mathrm{mav}}(\mathbf{x}) = \mathscr{F}[\bar{\mathscr{F}}[\wp] \times \bar{\mathscr{F}}[\chi_S]](\mathbf{x}).$$

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

The iterative use of the estimated envelope $\mathscr{U}$ 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 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, 1952*b*) 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 χ_U in real space, whereas they involve a *convolution* with $\bar{\mathscr{F}}[\chi_U]$ 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 Δ (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 $\wp$ be a $\mathbb{Z}^3$-periodic function. Let K be the interpolation kernel in 'normalized' form, *i.e.* such that $\int_{\mathbb{R}^3} K(\mathbf{x}) \, \mathrm{d}^3\mathbf{x} = 1$ 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

$$
\begin{aligned}
W(t) &= 1 - |t| & &\text{if } |t| \le 1, \\
&= 0 & &\text{if } |t| \ge 1.
\end{aligned}
$$

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

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

where $[(N_1^{-1})^{\#}K](\mathbf{x}) = K(\mathbf{N}_1\mathbf{x})$, so that

$$
\begin{aligned}
\bar{\mathscr{F}}[I_{\mathbf{N}_1}\wp] &= \left[\bar{\mathscr{F}}[\wp] * |\det \mathbf{N}_1| \sum_{\mathbf{k}_1\in\mathbb{Z}^3} \delta_{(\mathbf{N}_1^T\mathbf{k}_1)}\right] \\
&\quad \times \left[\frac{1}{|\det \mathbf{N}_1|}(N_1^T)^{\#}\bar{\mathscr{F}}[K]\right] \\
&= \left[\sum_{\mathbf{k}_1\in\mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_1^T\mathbf{k}_1}\bar{\mathscr{F}}[\wp]\right] \times (N_1^T)^{\#}\bar{\mathscr{F}}[K].
\end{aligned}
$$

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

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

(ii) a series of 'ghost bands' corresponding to $\mathbf{k}_1 \neq \mathbf{0}$, which consist of translates of $\bar{\mathscr{F}}[\rho]$ multiplied by the tail regions of $(N_1^T)^{\#}\bar{\mathscr{F}}[K]$.

Thus $I_{\mathbf{N}_1}\rho$ is not band-limited even if ρ is. Supposing, however, that ρ is band-limited and that grid $\mathscr{G}_1$ satisfies the Shannon sampling criterion, we see that there will be no overlap between the different bands: $\bar{\mathscr{F}}[\rho]$ may therefore be recovered from the main band by compensating its attenuation, which is approximately a temperature-factor correction.

For numerical work, however, $I_{\mathbf{N}_1}\rho$ must be resampled onto another grid $\mathscr{G}_2$, which causes its transform to become periodized into

$$|\det \mathbf{N}_2| \sum_{\mathbf{k}_2 \in \mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_2^T \mathbf{k}_2} \left\{ \left[\sum_{\mathbf{k}_1 \in \mathbb{Z}^3} \boldsymbol{\tau}_{\mathbf{N}_1^T \mathbf{k}_1} \bar{\mathscr{F}}[\rho] \right] (N_1^T)^{\#} \bar{\mathscr{F}}[K] \right\}.$$

This now causes the main band $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{0}$ to become contaminated by the ghost bands ($\mathbf{k}_1 \neq \mathbf{0}$) of the translates ($\mathbf{k}_2 \neq \mathbf{0}$) of $I_{\mathbf{N}_1}\rho$.

Aliasing errors may be minimized by increasing the sampling rate in grid $\mathscr{G}_1$ well beyond the Shannon minimum, which rapidly reduces the r.m.s. content of the ghost bands.

The sampling rate in grid $\mathscr{G}_2$ needs only exceed the Shannon minimum to the extent required to accommodate the increase in bandwidth due to convolution with $\bar{\mathscr{F}}[\chi_U]$, which is the reciprocal-space counterpart of envelope truncation (or solvent flattening) in real space.

1.3.4.4.3.5. Molecular-envelope transforms via Green's theorem

Green's theorem stated in terms of distributions (Section 1.3.2.3.9.1) is particularly well suited to the calculation of the Fourier transforms $\bar{\mathscr{F}}[\chi_U]$ of indicator functions. Let f be the indicator function χ_U and let S be the boundary of U (assumed to be a smooth surface). The jump σ_0 in the value of f across S along the outer normal vector is $\sigma_0 = -1$, the jump σ_ν in the normal derivative of f across S is $\sigma_\nu = 0$, and the Laplacian of f as a function is (almost everywhere) 0 so that $T_{\Delta f} = 0$. Green's theorem then reads:

$$\Delta(T_f) = T_{\Delta f} + \sigma_\nu \delta_{(S)} + \partial_\nu [\sigma_0 \delta_{(S)}]$$
$$= -\partial_\nu [\delta_{(S)}].$$

The function $e_{\mathbf{H}}(\mathbf{X}) = \exp(2\pi i \mathbf{H} \cdot \mathbf{X})$ satisfies the identity $\Delta e_{\mathbf{H}} = -4\pi^2 \|\mathbf{H}\|^2 e_{\mathbf{H}}$. Therefore, in Cartesian coordinates:

$$\bar{F}[\chi_U](\mathbf{H}) = \langle T_{\chi_U}, e_{\mathbf{H}} \rangle$$
$$= -\frac{1}{4\pi^2 \|\mathbf{H}\|^2} \langle T_{\chi_U}, \Delta e_{\mathbf{H}} \rangle$$
$$= -\frac{1}{4\pi^2 \|\mathbf{H}\|^2} \langle \Delta(T_{\chi_U}), e_{\mathbf{H}} \rangle \quad [\text{Section } 1.3.2.3.9.1(a)]$$
$$= -\frac{1}{4\pi^2 \|\mathbf{H}\|^2} \langle -\partial_\nu [\delta_{(S)}], e_{\mathbf{H}} \rangle$$
$$= -\frac{1}{4\pi^2 \|\mathbf{H}\|^2} \int_S \partial_\nu e_{\mathbf{H}} \, \mathrm{d}^2 S \quad [\text{Section } 1.3.2.3.9.1(c)]$$
$$= -\frac{1}{4\pi^2 \|\mathbf{H}\|^2} \int_S 2\pi i \mathbf{H} \cdot \mathbf{n} \exp(2\pi i \mathbf{H} \cdot \mathbf{X}) \, \mathrm{d}^2 S,$$

i.e.

$$\bar{\mathscr{F}}[\chi_U](\mathbf{H}) = \frac{1}{2\pi i \|\mathbf{H}\|^2} \int_S \mathbf{H} \cdot \mathbf{n} \exp(2\pi i \mathbf{H} \cdot \mathbf{X}) \, \mathrm{d}^2 S,$$

where $\mathbf{n}$ is the outer normal to S. This formula was used by von Laue (1936) for a different purpose, namely to calculate the transforms of crystal shapes (see also Ewald, 1940). If the surface S is given by a triangulation, the surface integral becomes a sum over all faces, since $\mathbf{n}$ is constant on each face. If U is a solid sphere with radius R, an integration by parts gives immediately:

$$\frac{1}{\mathrm{vol}(U)} \bar{\mathscr{F}}[\chi_U](\mathbf{H}) = \frac{3}{X^3} [\sin X - X \cos X]$$
$$\text{with } X = 2\pi \|\mathbf{H}\| R.$$

1.3.4.4.4. Structure factors from model atomic parameters

An atomic model of a crystal structure consists of a list of symmetry-unique atoms described by their positions, their thermal agitation and their chemical identity (which can be used as a pointer to form-factor tables). Form factors are usually parameterized as sums of Gaussians, and thermal agitation by a Gaussian temperature factor or tensor. The formulae given in Section 1.3.4.2.2.6 for Gaussian atoms are therefore adequate for most purposes. High-resolution electron-density studies use more involved parameterizations.

Early calculations were carried out by means of Bragg–Lipson charts (Bragg & Lipson, 1936) which gave a graphical representation of the symmetrized trigonometric sums Ξ of Section 1.3.4.2.2.9. The approximation of form factors by Gaussians goes back to the work of Vand *et al.* (1957) and Forsyth & Wells (1959). Agarwal (1978) gave simplified expansions suitable for medium-resolution modelling of macromolecular structures.

This method of calculating structure factors is expensive because *each* atom sends contributions of essentially equal magnitude to *all* structure factors in a resolution shell. The calculation is therefore of size $\propto N\mathscr{N}$ for N atoms and $\mathscr{N}$ reflections. Since N and $\mathscr{N}$ are roughly proportional at a given resolution, this method is very costly for large structures.

Two distinct programming strategies are available (Rollett, 1965) according to whether the fast loop is on all atoms for each reflection, or on all reflections for each atom. The former method was favoured in the early times when computers were unreliable. The latter was shown by Burnett & Nordman (1974) to be more amenable to efficient programming, as no multiplication is required in calculating the arguments of the sine/cosine terms: these can be accumulated by integer addition, and used as subscripts in referencing a trigonometric function table.

1.3.4.4.5. Structure factors via model electron-density maps

Robertson (1936b) recognized the similarity between the calculation of structure factors by Fourier summation and the calculation of Fourier syntheses, the main difference being of course that atomic coordinates do not usually lie exactly on a grid obtained by integer subdivision of the crystal lattice. He proposed to address this difficulty by the use of his sorting board, which could extend the scale of subdivision and thus avoid phase errors. In this way the calculation of structure factors became amenable to Beevers–Lipson strip methods, with considerable gain of speed.

Later, Beevers & Lipson (1952) proposed that trigonometric functions attached to atomic positions falling between the grid points on which Beevers–Lipson strips were based should be obtained by linear interpolation from the values found on the strips for the closest grid points. This amounts (Section 1.3.4.4.3.4) to using atoms in the shape of a trilinear wedge, whose