

1. GENERAL RELATIONSHIPS AND TECHNIQUES

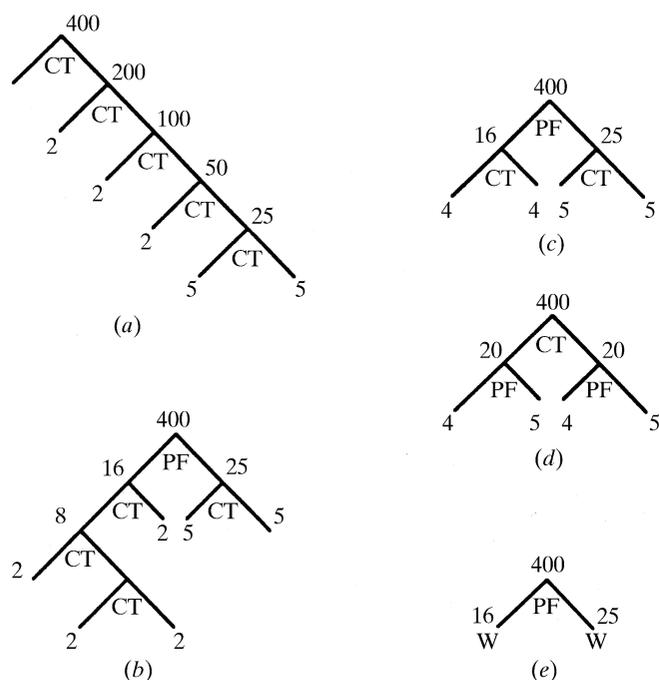


Fig. 1.3.3.1. A few global algorithms for computing a 400-point DFT. CT: Cooley–Tukey factorization. PF: prime factor (or Good) factorization. W: Winograd algorithm.

The simplest DFT may then be carried out into a global algorithm in many different ways. The diagrams in Fig. 1.3.3.1 illustrate a few of the options available to compute a 400-point DFT. They may differ greatly in their arithmetic operation counts.

1.3.3.3.2. Computer architecture considerations

To obtain a truly useful measure of the computational complexity of a DFT algorithm, its arithmetic operation count must be tempered by computer architecture considerations. Three main types of trade-offs must be borne in mind:

- (i) reductions in floating-point (f.p.) arithmetic count are obtained by reindexing, hence at the cost of an increase in integer arithmetic on addresses, although some shortcuts may be found (Uhrich, 1969; Burrus & Eschenbacher, 1981);
- (ii) reduction in the f.p. multiplication count usually leads to a large increase in the f.p. addition count (Morris, 1978);
- (iii) nesting can increase execution speed, but causes a loss of modularity and hence complicates program development (Silverman, 1977; Kolba & Parks, 1977).

Many of the mathematical developments above took place in the context of single-processor serial computers, where f.p. addition is substantially cheaper than f.p. multiplication but where integer address arithmetic has to compete with f.p. arithmetic for processor cycles. As a result, the alternatives to the Cooley–Tukey algorithm hardly ever led to particularly favourable trade-offs, thus creating the impression that there was little to gain by switching to more exotic algorithms.

The advent of new machine architectures with vector and/or parallel processing features has greatly altered this picture (Pease, 1968; Korn & Lambiotte, 1979; Fornberg, 1981; Swartzrauber, 1984):

- (i) *pipelining* equalizes the cost of f.p. addition and f.p. multiplication, and the ideal ‘blend’ of the two types of operations depends solely on the number of adder and multiplier units available in each machine;
- (ii) integer address arithmetic is delegated to specialized arithmetic and logical units (ALUs) operating concurrently with

the f.p. units, so that complex reindexing schemes may be used without loss of overall efficiency.

Another major consideration is that of data flow [see *e.g.* Nawab & McClellan (1979)]. Serial machines only have few registers and few paths connecting them, and allow little or no overlap between computation and data movement. New architectures, on the other hand, comprise banks of vector registers (or ‘cache memory’) besides the usual internal registers, and dedicated ALUs can service data transfers between several of them simultaneously and concurrently with computation.

In this new context, the devices described in Sections 1.3.3.2 and 1.3.3.3 for altering the balance between the various types of arithmetic operations, and reshaping the data flow during the computation, are invaluable. The field of machine-dependent DFT algorithm design is thriving on them [see *e.g.* Temperton (1983*a,b,c*, 1985); Agarwal & Cooley (1986, 1987)].

1.3.3.3.3. The Johnson–Burrus family of algorithms

In order to explore systematically all possible algorithms for carrying out a given DFT computation, and to pick the one best suited to a given machine, attempts have been made to develop:

- (i) a high-level notation of describing all the ingredients of a DFT computation, including data permutation and data flow;
- (ii) a formal calculus capable of operating on these descriptions so as to represent all possible reorganizations of the computation;
- (iii) an automatic procedure for evaluating the performance of a given algorithm on a specific architecture.

Task (i) can be accomplished by systematic use of a tensor product notation to represent the various stages into which the DFT can be factored (reindexing, small transforms on subsets of indices, twiddle factors, digit-reversal permutations).

Task (ii) may for instance use the Winograd **CBA** normal form for each small transform, then apply the rules governing the rearrangement of tensor product \otimes and ordinary product \times operations on matrices. The matching of these rearrangements to the architecture of a vector and/or parallel computer can be formalized algebraically [see *e.g.* Chapter 2 of Tolimieri *et al.* (1989)].

Task (iii) is a complex search which requires techniques such as dynamic programming (Bellman, 1958).

Johnson & Burrus (1983) have proposed and tested such a scheme to identify the optimal trade-offs between prime factor nesting and Winograd nesting of small Winograd transforms. In step (ii), they further decomposed the pre-addition matrix **A** and post-addition matrix **C** into several factors, so that the number of design options available becomes very large: the N -point DFT when N has four factors can be calculated in over 10^{12} distinct ways.

This large family of nested algorithms contains the prime factor algorithm and the Winograd algorithms as particular cases, but usually achieves greater efficiency than either by reducing the f.p. multiplication count while keeping the number of f.p. additions small.

There is little doubt that this systematic approach will be extended so as to incorporate all available methods of restructuring the DFT.

1.3.4. Crystallographic applications of Fourier transforms

1.3.4.1. Introduction

The central role of the Fourier transformation in X-ray crystallography is a consequence of the kinematic approximation used in the description of the scattering of X-rays by a distribution of electrons (Bragg, 1915; Duane, 1925; Havighurst, 1925*a,b*; Zachariasen, 1945; James, 1948*a*, Chapters 1 and 2; Lipson & Cochran, 1953, Chapter 1; Bragg, 1975).

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

Let $\rho(\mathbf{X})$ be the density of electrons in a sample of matter contained in a finite region V which is being illuminated by a parallel monochromatic X-ray beam with wavevector \mathbf{K}_0 . Then the far-field amplitude scattered in a direction corresponding to wavevector $\mathbf{K} = \mathbf{K}_0 + \mathbf{H}$ is proportional to

$$\begin{aligned} F(\mathbf{H}) &= \int_V \rho(\mathbf{X}) \exp(2\pi i \mathbf{H} \cdot \mathbf{X}) \, d^3 \mathbf{X} \\ &= \tilde{\mathcal{F}}[\rho](\mathbf{H}) \\ &= \langle \rho_{\mathbf{x}}, \exp(2\pi i \mathbf{H} \cdot \mathbf{X}) \rangle. \end{aligned}$$

In certain model calculations, the ‘sample’ may contain not only volume charges, but also point, line and surface charges. These singularities may be accommodated by letting ρ be a distribution, and writing

$$F(\mathbf{H}) = \tilde{\mathcal{F}}[\rho](\mathbf{H}) = \langle \rho_{\mathbf{x}}, \exp(2\pi i \mathbf{H} \cdot \mathbf{X}) \rangle.$$

F is still a well behaved function (analytic, by Section 1.3.2.4.2.10) because ρ has been assumed to have compact support.

If the sample is assumed to be an infinite crystal, so that ρ is now a *periodic* distribution, the customary limiting process by which it is shown that F becomes a discrete series of peaks at reciprocal-lattice points (see e.g. von Laue, 1936; Ewald, 1940; James, 1948a p. 9; Lipson & Taylor, 1958, pp. 14–27; Ewald, 1962, pp. 82–101; Warren, 1969, pp. 27–30) is already subsumed under the treatment of Section 1.3.2.6.

1.3.4.2. Crystallographic Fourier transform theory

1.3.4.2.1. Crystal periodicity

1.3.4.2.1.1. Period lattice, reciprocal lattice and structure factors

Let ρ be the distribution of electrons in a crystal. Then, by definition of a crystal, ρ is Λ -periodic for some period lattice Λ (Section 1.3.2.6.5) so that there exists a motif distribution ρ^0 with compact support such that

$$\rho = R * \rho^0,$$

where $R = \sum_{\mathbf{x} \in \Lambda} \delta(\mathbf{x})$. The lattice Λ is usually taken to be the finest for which the above representation holds.

Let Λ have a basis $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ over the integers, these basis vectors being expressed in terms of a standard orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ as

$$\mathbf{a}_k = \sum_{j=1}^3 a_{jk} \mathbf{e}_j.$$

Then the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

is the period matrix of Λ (Section 1.3.2.6.5) with respect to the unit lattice with basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, and the volume V of the unit cell is given by $V = |\det \mathbf{A}|$.

By Fourier transformation

$$\tilde{\mathcal{F}}[\rho] = R^* \times \tilde{\mathcal{F}}[\rho^0],$$

where $R^* = \sum_{\mathbf{H} \in \Lambda^*} \delta(\mathbf{H})$ is the lattice distribution associated to the reciprocal lattice Λ^* . The basis vectors $(\mathbf{a}_1^*, \mathbf{a}_2^*, \mathbf{a}_3^*)$ have coordinates in $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ given by the columns of $(\mathbf{A}^{-1})^T$, whose expression in terms of the cofactors of \mathbf{A} (see Section 1.3.2.6.5) gives the familiar formulae involving the cross product of vectors for $n = 3$. The \mathbf{H} -distribution F of scattered amplitudes may be written

$$F = \tilde{\mathcal{F}}[\rho]_{\mathbf{H}} = \sum_{\mathbf{H} \in \Lambda^*} \tilde{\mathcal{F}}[\rho^0](\mathbf{H}) \delta(\mathbf{H}) = \sum_{\mathbf{H} \in \Lambda^*} F_{\mathbf{H}} \delta(\mathbf{H})$$

and is thus a weighted reciprocal-lattice distribution, the weight $F_{\mathbf{H}}$ attached to each node $\mathbf{H} \in \Lambda^*$ being the value at \mathbf{H} of the transform $\tilde{\mathcal{F}}[\rho^0]$ of the motif ρ^0 . Taken in conjunction with the assumption that the scattering is elastic, i.e. that \mathbf{H} only changes the direction but not the magnitude of the incident wavevector \mathbf{K}_0 , this result yields the usual forms (Laue or Bragg) of the diffraction conditions: $\mathbf{H} \in \Lambda^*$, and simultaneously \mathbf{H} lies on the Ewald sphere.

By the reciprocity theorem, ρ^0 can be recovered if F is known for all $\mathbf{H} \in \Lambda^*$ as follows [Section 1.3.2.6.5, e.g. (iv)]:

$$\rho_{\mathbf{x}} = \frac{1}{V} \sum_{\mathbf{H} \in \Lambda^*} F_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{X}).$$

These relations may be rewritten in terms of standard, or ‘fractional crystallographic’, coordinates by putting

$$\mathbf{X} = \mathbf{A} \mathbf{x}, \quad \mathbf{H} = (\mathbf{A}^{-1})^T \mathbf{h},$$

so that a unit cell of the crystal corresponds to $\mathbf{x} \in \mathbb{R}^3 / \mathbb{Z}^3$, and that $\mathbf{h} \in \mathbb{Z}^3$. Defining ρ and ρ^0 by

$$\rho = \frac{1}{V} A^{\#} \rho, \quad \rho^0 = \frac{1}{V} A^{\#} \rho^0$$

so that

$$\rho(\mathbf{X}) \, d^3 \mathbf{X} = \rho(\mathbf{x}) \, d^3 \mathbf{x}, \quad \rho^0(\mathbf{X}) \, d^3 \mathbf{X} = \rho^0(\mathbf{x}) \, d^3 \mathbf{x},$$

we have

$$\tilde{\mathcal{F}}[\rho]_{\mathbf{h}} = \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \delta(\mathbf{h}),$$

$$\begin{aligned} F(\mathbf{h}) &= \langle \rho_{\mathbf{x}}^0, \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) \rangle \\ &= \int_{\mathbb{R}^3 / \mathbb{Z}^3} \rho^0(\mathbf{x}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) \, d^3 \mathbf{x} \quad \text{if } \rho^0 \in L_{\text{loc}}^1(\mathbb{R}^3 / \mathbb{Z}^3), \end{aligned}$$

$$\rho_{\mathbf{x}} = \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}).$$

These formulae are valid for an arbitrary motif distribution ρ^0 , provided the convergence of the Fourier series for ρ is considered from the viewpoint of distribution theory (Section 1.3.2.6.10.3).

The experienced crystallographer may notice the absence of the familiar factor $1/V$ from the expression for ρ just given. This is because we use the (mathematically) natural unit for ρ , the *electron per unit cell*, which matches the dimensionless nature of the crystallographic coordinates \mathbf{x} and of the associated volume element $d^3 \mathbf{x}$. The traditional factor $1/V$ was the result of the somewhat inconsistent use of \mathbf{x} as an argument but of $d^3 \mathbf{X}$ as a volume element to obtain ρ in electrons per unit volume (e.g. \AA^{-3}). A fortunate consequence of the present convention is that nuisance factors of V or $1/V$, which used to abound in convolution or scalar product formulae, are now absent.

It should be noted at this point that the crystallographic terminology regarding $\tilde{\mathcal{F}}$ and $\tilde{\mathcal{F}}$ differs from the standard mathematical terminology introduced in Section 1.3.2.4.1 and applied to periodic distributions in Section 1.3.2.6.4: F is the *inverse* Fourier transform of ρ rather than its Fourier transform, and the calculation of ρ is called a Fourier *synthesis* in crystallography even though it is mathematically a Fourier analysis. The origin of this discrepancy may be traced to the fact that the mathematical theory of the Fourier transformation originated with the study of *temporal* periodicity, while crystallography deals with *spatial* periodicity; since the expression for the phase factor of a plane wave is $\exp[2\pi i(\nu t - \mathbf{K} \cdot \mathbf{X})]$, the difference in sign between the

1. GENERAL RELATIONSHIPS AND TECHNIQUES

contributions from time *versus* spatial displacements makes this conflict unavoidable.

1.3.4.2.1.2. Structure factors in terms of form factors

In many cases, ρ^0 is a sum of translates of atomic electron-density distributions. Assume there are n distinct chemical types of atoms, with N_j identical isotropic atoms of type j described by an electron distribution ρ_j about their centre of mass. According to quantum mechanics each ρ_j is a smooth rapidly decreasing function of \mathbf{x} , i.e. $\rho_j \in \mathcal{S}$, hence $\rho^0 \in \mathcal{S}$ and (ignoring the effect of thermal agitation)

$$\rho^0(\mathbf{x}) = \sum_{j=1}^n \left[\sum_{k_j=1}^{N_j} \rho_j(\mathbf{x} - \mathbf{x}_{k_j}) \right],$$

which may be written (Section 1.3.2.5.8)

$$\rho^0 = \sum_{j=1}^n \left[\rho_j * \left(\sum_{k_j=1}^{N_j} \delta_{(\mathbf{x}_{k_j})} \right) \right].$$

By Fourier transformation:

$$F(\mathbf{h}) = \sum_{j=1}^n \left\{ \tilde{\mathcal{F}}[\rho_j](\mathbf{h}) \times \left[\sum_{k_j=1}^{N_j} \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_{k_j}) \right] \right\}.$$

Defining the form factor f_j of atom j as a function of \mathbf{h} to be

$$f_j(\mathbf{h}) = \tilde{\mathcal{F}}[\rho_j](\mathbf{h})$$

we have

$$F(\mathbf{h}) = \sum_{j=1}^n f_j(\mathbf{h}) \times \left[\sum_{k_j=1}^{N_j} \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_{k_j}) \right].$$

If $\mathbf{X} = \mathbf{A}\mathbf{x}$ and $\mathbf{H} = (\mathbf{A}^{-1})^T \mathbf{h}$ are the real- and reciprocal-space coordinates in \AA and \AA^{-1} , and if $\rho_j(\|\mathbf{X}\|)$ is the spherically symmetric electron-density function for atom type j , then

$$f_j(\mathbf{H}) = \int_0^\infty 4\pi \|\mathbf{X}\|^2 \rho_j(\|\mathbf{X}\|) \frac{\sin(2\pi \|\mathbf{H}\| \|\mathbf{X}\|)}{2\pi \|\mathbf{H}\| \|\mathbf{X}\|} d\|\mathbf{X}\|.$$

More complex expansions are used for electron-density studies (see Chapter 1.2 in this volume). Anisotropic Gaussian atoms may be dealt with through the formulae given in Section 1.3.2.4.4.2.

1.3.4.2.1.3. Fourier series for the electron density and its summation

The convergence of the Fourier series for ρ

$$\rho(\mathbf{x}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

is usually examined from the classical point of view (Section 1.3.2.6.10). The summation of multiple Fourier series meets with considerable difficulties, because there is no natural order in \mathbb{Z}^n to play the role of the natural order in \mathbb{Z} (Ash, 1976). In crystallography, however, the structure factors $F(\mathbf{h})$ are often obtained within spheres $\|\mathbf{H}\| \leq \Delta^{-1}$ for increasing resolution (decreasing Δ). Therefore, successive estimates of ρ are most naturally calculated as the corresponding partial sums (Section 1.3.2.6.10.1):

$$S_\Delta(\rho)(\mathbf{x}) = \sum_{\|(\mathbf{A}^{-1})^T \mathbf{h}\| \leq \Delta^{-1}} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}).$$

This may be written

$$S_\Delta(\rho)(\mathbf{x}) = (D_\Delta * \rho)(\mathbf{x}),$$

where D_Δ is the ‘spherical Dirichlet kernel’

$$D_\Delta(\mathbf{x}) = \sum_{\|(\mathbf{A}^{-1})^T \mathbf{h}\| \leq \Delta^{-1}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}).$$

D_Δ exhibits numerous negative ripples around its central peak. Thus the ‘series termination errors’ incurred by using $S_\Delta(\rho)$ instead of ρ consist of negative ripples around each atom, and may lead to a Gibbs-like phenomenon (Section 1.3.2.6.10.1) near a molecular boundary.

As in one dimension, Cesàro sums (arithmetic means of partial sums) have better convergence properties, as they lead to a convolution by a ‘spherical Fejér kernel’ which is everywhere positive. Thus Cesàro summation will always produce positive approximations to a positive electron density. Other positive summation kernels were investigated by Pepinsky (1952) and by Waser & Schomaker (1953).

1.3.4.2.1.4. Friedel’s law, anomalous scatterers

If the wavelength λ of the incident X-rays is far from any absorption edge of the atoms in the crystal, there is a constant phase shift in the scattering, and the electron density may be considered to be *real-valued*. Then

$$\begin{aligned} F(\mathbf{h}) &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}) d^3 \mathbf{x} \\ &= \overline{\int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \exp[2\pi i (-\mathbf{h}) \cdot \mathbf{x}] d^3 \mathbf{x}} \\ &= \overline{F(-\mathbf{h})} \text{ since } \overline{\rho(\mathbf{x})} = \rho(\mathbf{x}). \end{aligned}$$

Thus if

$$F(\mathbf{h}) = |F(\mathbf{h})| \exp(i\varphi(\mathbf{h})),$$

then

$$|F(-\mathbf{h})| = |F(\mathbf{h})| \quad \text{and} \quad \varphi(-\mathbf{h}) = -\varphi(\mathbf{h}).$$

This is Friedel’s law (Friedel, 1913). The set $\{F_{\mathbf{h}}\}$ of Fourier coefficients is said to have *Hermitian symmetry*.

If λ is close to some absorption edge(s), the proximity to resonance induces an extra phase shift, whose effect may be represented by letting $\rho(\mathbf{x})$ take on *complex values*. Let

$$\rho(\mathbf{x}) = \rho^R(\mathbf{x}) + i\rho^I(\mathbf{x})$$

and correspondingly, by termwise Fourier transformation

$$F(\mathbf{h}) = F^R(\mathbf{h}) + iF^I(\mathbf{h}).$$

Since $\rho^R(\mathbf{x})$ and $\rho^I(\mathbf{x})$ are both real, $F^R(\mathbf{h})$ and $F^I(\mathbf{h})$ are both Hermitian symmetric, hence

$$F(-\mathbf{h}) = \overline{F^R(\mathbf{h})} + i\overline{F^I(\mathbf{h})},$$

while

$$\overline{F(\mathbf{h})} = \overline{F^R(\mathbf{h})} - i\overline{F^I(\mathbf{h})}.$$

Thus $F(-\mathbf{h}) \neq \overline{F(\mathbf{h})}$, so that Friedel’s law is violated. The components $F^R(\mathbf{h})$ and $F^I(\mathbf{h})$, which do obey Friedel’s law, may be expressed as:

$$F^R(\mathbf{h}) = \frac{1}{2}[F(\mathbf{h}) + \overline{F(-\mathbf{h})}],$$

$$F^I(\mathbf{h}) = \frac{1}{2i}[F(\mathbf{h}) - \overline{F(-\mathbf{h})}].$$

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

1.3.4.2.1.5. *Parseval's identity and other L^2 theorems*
By Section 1.3.2.4.3.3 and Section 1.3.2.6.10.2,

$$\sum_{\mathbf{h} \in \mathbb{Z}^3} |F(\mathbf{h})|^2 = \int_{\mathbb{R}^3/\mathbb{Z}^3} |\rho(\mathbf{x})|^2 d^3\mathbf{x} = V \int_{\mathbb{R}^3/\Lambda} |\rho(\mathbf{X})|^2 d^3\mathbf{X}.$$

Usually $\rho(\mathbf{x})$ is real and positive, hence $|\rho(\mathbf{x})| = \rho(\mathbf{x})$, but the identity remains valid even when $\rho(\mathbf{x})$ is made complex-valued by the presence of anomalous scatterers.

If $\{G_{\mathbf{h}}\}$ is the collection of structure factors belonging to another electron density $\sigma = A^{\#}\sigma$ with the same period lattice as ρ , then

$$\begin{aligned} \sum_{\mathbf{h} \in \mathbb{Z}^3} \overline{F(\mathbf{h})}G(\mathbf{h}) &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho(\mathbf{x})}\sigma(\mathbf{x}) d^3\mathbf{x} \\ &= V \int_{\mathbb{R}^3/\Lambda} \rho(\mathbf{X})\sigma(\mathbf{X}) d^3\mathbf{X}. \end{aligned}$$

Thus, norms and inner products may be evaluated either from structure factors or from 'maps'.

1.3.4.2.1.6. *Convolution, correlation and Patterson function*

Let $\rho = r * \rho^0$ and $\sigma = r * \sigma^0$ be two electron densities referred to crystallographic coordinates, with structure factors $\{F_{\mathbf{h}}\}_{\mathbf{h} \in \mathbb{Z}^3}$ and $\{G_{\mathbf{h}}\}_{\mathbf{h} \in \mathbb{Z}^3}$, so that

$$\begin{aligned} \rho_{\mathbf{x}} &= \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}), \\ \sigma_{\mathbf{x}} &= \sum_{\mathbf{h} \in \mathbb{Z}^3} G(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}). \end{aligned}$$

The distribution $\omega = r * (\rho^0 * \sigma^0)$ is well defined, since the generalized support condition (Section 1.3.2.3.9.7) is satisfied. The forward version of the convolution theorem implies that if

$$\omega_{\mathbf{x}} = \sum_{\mathbf{h} \in \mathbb{Z}^3} W(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}),$$

then

$$W(\mathbf{h}) = F(\mathbf{h})G(\mathbf{h}).$$

If either ρ^0 or σ^0 is infinitely differentiable, then the distribution $\psi = \rho \times \sigma$ exists, and if we analyse it as

$$\psi_{\mathbf{x}} = \sum_{\mathbf{h} \in \mathbb{Z}^3} Y(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}),$$

then the backward version of the convolution theorem reads:

$$Y(\mathbf{h}) = \sum_{\mathbf{k} \in \mathbb{Z}^3} F(\mathbf{h})G(\mathbf{h} - \mathbf{k}).$$

The *cross correlation* $\kappa[\rho, \sigma]$ between ρ and σ is the \mathbb{Z}^3 -periodic distribution defined by:

$$\kappa = \check{\rho}^0 * \sigma.$$

If ρ^0 and σ^0 are locally integrable,

$$\begin{aligned} \kappa[\rho, \sigma](\mathbf{t}) &= \int_{\mathbb{R}^3} \rho^0(\mathbf{x})\sigma(\mathbf{x} + \mathbf{t}) d^3\mathbf{x} \\ &= \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x})\sigma(\mathbf{x} + \mathbf{t}) d^3\mathbf{x}. \end{aligned}$$

Let

$$\kappa(\mathbf{t}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} K(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{t}).$$

The combined use of the shift property and of the forward convolution theorem then gives immediately:

$$K(\mathbf{h}) = \overline{F(\mathbf{h})}G(\mathbf{h});$$

hence the Fourier series representation of $\kappa[\rho, \sigma]$:

$$\kappa[\rho, \sigma](\mathbf{t}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} \overline{F(\mathbf{h})}G(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{t}).$$

Clearly, $\kappa[\rho, \sigma] = (\kappa[\sigma, \rho])^*$, as shown by the fact that permuting F and G changes $K(\mathbf{h})$ into its complex conjugate.

The *auto-correlation* of ρ is defined as $\kappa[\rho, \rho]$ and is called the *Patterson function* of ρ . If ρ consists of point atoms, *i.e.*

$$\rho^0 = \sum_{j=1}^N Z_j \delta_{(\mathbf{x}_j)},$$

then

$$\kappa[\rho, \rho] = r * \left[\sum_{j=1}^N \sum_{k=1}^N Z_j Z_k \delta_{(\mathbf{x}_j - \mathbf{x}_k)} \right]$$

contains information about interatomic vectors. It has the Fourier series representation

$$\kappa[\rho, \rho](\mathbf{t}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} |F(\mathbf{h})|^2 \exp(-2\pi i \mathbf{h} \cdot \mathbf{t}),$$

and is therefore calculable from the diffraction intensities alone. It was first proposed by Patterson (1934, 1935*a,b*) as an extension to crystals of the radially averaged correlation function used by Warren & Gingrich (1934) in the study of powders.

1.3.4.2.1.7. *Sampling theorems, continuous transforms, interpolation*

Shannon's sampling and interpolation theorem (Section 1.3.2.7.1) takes two different forms, according to whether the property of finite bandwidth is assumed in real space or in reciprocal space.

(1) The most usual setting is in reciprocal space (see Sayre, 1952*c*). Only a finite number of diffraction intensities can be recorded and phased, and for physical reasons the cutoff criterion is the resolution $\Delta = 1/\|\mathbf{H}\|_{\max}$. Electron-density maps are thus calculated as partial sums (Section 1.3.4.2.1.3), which may be written in Cartesian coordinates as

$$S_{\Delta}(\rho)(\mathbf{X}) = \sum_{\mathbf{H} \in \Lambda^*, \|\mathbf{H}\| \leq \Delta^{-1}} F(\mathbf{H}) \exp(-2\pi i \mathbf{H} \cdot \mathbf{X}).$$

$S_{\Delta}(\rho)$ is band-limited, the support of its spectrum being contained in the solid sphere Σ_{Δ} defined by $\|\mathbf{H}\| \leq \Delta^{-1}$. Let χ_{Δ} be the indicator function of Σ_{Δ} . The transform of the normalized version of χ_{Δ} is (see below, Section 1.3.4.4.3.5)

$$\begin{aligned} I_{\Delta}(\mathbf{X}) &= \frac{3\Delta^3}{4\pi} \mathcal{F}[\chi_{\Delta}](\mathbf{X}) \\ &= \frac{3}{u^3} (\sin u - u \cos u) \quad \text{where } u = 2\pi \frac{\|\mathbf{X}\|}{\Delta}. \end{aligned}$$

By Shannon's theorem, it suffices to calculate $S_{\Delta}(\rho)$ on an integral subdivision Γ of the period lattice Λ such that the sampling criterion is satisfied (*i.e.* that the translates of Σ_{Δ} by vectors of Γ^* do not overlap). Values of $S_{\Delta}(\rho)$ may then be calculated at an arbitrary point \mathbf{X} by the interpolation formula:

$$S_{\Delta}(\rho)(\mathbf{X}) = \sum_{\mathbf{Y} \in \Gamma} I_{\Delta}(\mathbf{X} - \mathbf{Y}) S_{\Delta}(\rho)(\mathbf{Y}).$$

(2) The reverse situation occurs whenever the support of the motif ρ^0 does not fill the whole unit cell, *i.e.* whenever there exists a region M (the 'molecular envelope'), strictly smaller than the unit cell, such that the translates of M by vectors of r do not overlap and that

1. GENERAL RELATIONSHIPS AND TECHNIQUES

$$\chi_M \times \rho^0 = \rho^0.$$

It then follows that $\rho = r * (\chi_M \times \rho)$. Defining the ‘interference function’ G as the normalized indicator function of M according to

$$G(\boldsymbol{\eta}) = \frac{1}{\text{vol}(M)} \tilde{\mathcal{F}}[\chi_M](\boldsymbol{\eta})$$

we may invoke Shannon’s theorem to calculate the value $\tilde{\mathcal{F}}[\rho^0](\boldsymbol{\xi})$ at an arbitrary point $\boldsymbol{\xi}$ of reciprocal space from its sample values $F(\mathbf{h}) = \tilde{\mathcal{F}}[\rho^0](\mathbf{h})$ at points of the reciprocal lattice as

$$\tilde{\mathcal{F}}[\rho^0](\boldsymbol{\xi}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} \mathbf{G}(\boldsymbol{\xi} - \mathbf{h}) F(\mathbf{h}).$$

This aspect of Shannon’s theorem constitutes the mathematical basis of phasing methods based on geometric redundancies created by solvent regions and/or noncrystallographic symmetries (Bricogne, 1974). The connection between Shannon’s theorem and the phase problem was first noticed by Sayre (1952*b*). He pointed out that the Patterson function of ρ , written as $\kappa[\rho, \rho] = r * (\check{\rho}^0 * \rho^0)$, may be viewed as consisting of a motif $\kappa^0 = \check{\rho}^0 * \rho^0$ (containing all the internal interatomic vectors) which is periodized by convolution with r . As the translates of κ^0 by vectors of \mathbb{Z}^3 do overlap, the sample values of the intensities $|F(\mathbf{h})|^2$ at nodes of the reciprocal lattice do not provide enough data to interpolate intensities $|F(\boldsymbol{\xi})|^2$ at arbitrary points of reciprocal space. Thus the loss of phase is intimately related to the impossibility of intensity interpolation, implying in return that any indication of intensity values attached to non-integral points of the reciprocal lattice is a potential source of phase information.

1.3.4.2.1.8. Sections and projections

It was shown at the end of Section 1.3.2.5.8 that the convolution theorem establishes, under appropriate assumptions, a duality between sectioning a smooth function (viewed as a multiplication by a δ -function in the sectioning coordinate) and projecting its transform (viewed as a convolution with the function $\mathbf{1}$ everywhere equal to 1 as a function of the projection coordinate). This duality follows from the fact that \mathcal{F} and $\tilde{\mathcal{F}}$ map $\mathbf{1}_{x_i}$ to δ_{x_i} and δ_{x_i} to $\mathbf{1}_{x_i}$ (Section 1.3.2.5.6), and from the tensor product property (Section 1.3.2.5.5).

In the case of periodic distributions, projection and section must be performed with respect to directions or subspaces which are integral with respect to the period lattice if the result is to be periodic; furthermore, projections must be performed only on the contents of *one* repeating unit along the direction of projection, or else the result would diverge. The same relations then hold between principal central sections and projections of the electron density and the dual principal central projections and sections of the weighted reciprocal lattice, *e.g.*

$$\rho(x_1, 0, 0) \leftrightarrow \sum_{h_1, h_2} F(h_1, h_2, h_3),$$

$$\rho(x_1, x_2, 0) \leftrightarrow \sum_{h_3} F(h_1, h_2, h_3),$$

$$\rho_{1,2}(x_3) = \int_{\mathbb{R}^2/\mathbb{Z}^2} \rho(x_1, x_2, x_3) dx_1 dx_2 \leftrightarrow F(0, 0, h_3),$$

$$\rho_1(x_2, x_3) = \int_{\mathbb{R}/\mathbb{Z}} \rho(x_1, x_2, x_3) dx_1 \leftrightarrow F(0, h_2, h_3)$$

etc.

When the sections are principal but not central, it suffices to use the shift property of Section 1.3.2.5.5. When the sections or projections are not principal, they can be made principal by changing to new primitive bases B and B^* for Λ and Λ^* , respectively, the transition matrices \mathbf{P} and \mathbf{P}^* to these new bases

being related by $\mathbf{P}^* = (\mathbf{P}^{-1})^T$ in order to preserve duality. This change of basis must be such that one of these matrices (say, \mathbf{P}) should have a given integer vector \mathbf{u} as its first column, \mathbf{u} being related to the line or plane defining the section or projection of interest.

The problem of constructing a matrix \mathbf{P} given \mathbf{u} received an erroneous solution in Volume II of *International Tables* (Patterson, 1959), which was subsequently corrected in 1962. Unfortunately, the solution proposed there is complicated and does not suggest a general approach to the problem. It therefore seems worthwhile to record here an effective procedure which solves this problem in any dimension n (Watson, 1970).

Let

$$\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

be a primitive integral vector, *i.e.* g.c.d. $(u_1, \dots, u_n) = 1$. Then an $n \times n$ integral matrix \mathbf{P} with $\det \mathbf{P} = 1$ having \mathbf{u} as its first column can be constructed by induction as follows. For $n = 1$ the result is trivial. For $n = 2$ it can be solved by means of the Euclidean algorithm, which yields z_1, z_2 such that $u_1 z_2 - u_2 z_1 = 1$, so that we

may take $\mathbf{P} = \begin{pmatrix} u_1 & z_1 \\ u_2 & z_2 \end{pmatrix}$. Note that, if $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ is a solution,

then $\mathbf{z} + m\mathbf{u}$ is another solution for any $m \in \mathbb{Z}$. For $n \geq 3$, write

$\mathbf{u} = \begin{pmatrix} u_1 \\ d\mathbf{z} \end{pmatrix}$ with $d = \text{g.c.d.}(u_2, \dots, u_n)$ so that both $\mathbf{z} = \begin{pmatrix} z_2 \\ \vdots \\ z_n \end{pmatrix}$

and $\begin{pmatrix} u_1 \\ d \end{pmatrix}$ are primitive. By the inductive hypothesis there is an integral 2×2 matrix \mathbf{V} with $\begin{pmatrix} u_1 \\ d \end{pmatrix}$ as its first column, and an

integral $(n-1) \times (n-1)$ matrix \mathbf{Z} with \mathbf{z} as its first column, with $\det \mathbf{V} = 1$ and $\det \mathbf{Z} = 1$.

Now put

$$\mathbf{P} = \begin{pmatrix} 1 & \\ & \mathbf{Z} \end{pmatrix} \begin{pmatrix} \mathbf{V} \\ \mathbf{I}_{n-2} \end{pmatrix},$$

i.e.

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & z_2 & * & \dots & * \\ 0 & z_3 & * & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & z_n & * & \dots & * \end{pmatrix} \begin{pmatrix} u_1 & * & 0 & \dots & 0 \\ d & * & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

The first column of \mathbf{P} is

$$\begin{pmatrix} u_1 \\ dz_2 \\ \vdots \\ dz_n \end{pmatrix} = \mathbf{u},$$

and its determinant is 1, QED.

The incremental step from dimension $n-1$ to dimension n is the construction of 2×2 matrix \mathbf{V} , for which there exist infinitely many solutions labelled by an integer m_{n-1} . Therefore, the collection of matrices \mathbf{P} which solve the problem is labelled by $n-1$ arbitrary integers $(m_1, m_2, \dots, m_{n-1})$. This freedom can be used to adjust the shape of the basis B .

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

Once \mathbf{P} has been chosen, the calculation of general sections and projections is transformed into that of *principal* sections and projections by the changes of coordinates:

$$\mathbf{x} = \mathbf{P}\mathbf{x}', \quad \mathbf{h} = \mathbf{P}^*\mathbf{h}'$$

and an appeal to the tensor product property.

Booth (1945a) made use of the convolution theorem to form the Fourier coefficients of 'bounded projections', which provided a compromise between 2D and 3D Fourier syntheses. If it is desired to compute the projection on the (x, y) plane of the electron density lying between the planes $z = z_1$ and $z = z_2$, which may be written as

$$[\rho \times (\mathbf{1}_x \otimes \mathbf{1}_y \otimes \chi_{[z_1, z_2]})] * (\delta_x \otimes \delta_y \otimes \mathbf{1}_z)$$

The transform is then

$$[F * (\delta_h \otimes \delta_k \otimes \tilde{\mathcal{F}}[\chi_{[z_1, z_2]}])] \times (\mathbf{1}_h \otimes \mathbf{1}_k \otimes \delta_l)$$

giving for coefficient (h, k) :

$$\sum_{l \in \mathbb{Z}} F(h, k, l) \exp\{2\pi i l [(z_1 + z_2)/2]\} \times \frac{\sin \pi l (z_1 - z_2)}{\pi l}$$

1.3.4.2.1.9. Differential syntheses

Another particular instance of the convolution theorem is the duality between differentiation and multiplication by a monomial (Sections 1.3.2.4.2.8, 1.3.2.5.8).

In the present context, this result may be written

$$\begin{aligned} \tilde{\mathcal{F}} \left[\frac{\partial^{m_1+m_2+m_3} \rho}{\partial X_1^{m_1} \partial X_2^{m_2} \partial X_3^{m_3}} \right] (\mathbf{H}) \\ = (-2\pi i)^{m_1+m_2+m_3} H_1^{m_1} H_2^{m_2} H_3^{m_3} F(\mathbf{A}^T \mathbf{H}) \end{aligned}$$

in Cartesian coordinates, and

$$\tilde{\mathcal{F}} \left[\frac{\partial^{m_1+m_2+m_3} \rho}{\partial X_1^{m_1} \partial X_2^{m_2} \partial X_3^{m_3}} \right] (\mathbf{h}) = (-2\pi i)^{m_1+m_2+m_3} h_1^{m_1} h_2^{m_2} h_3^{m_3} F(\mathbf{h})$$

in crystallographic coordinates.

A particular case of the first formula is

$$-4\pi^2 \sum_{\mathbf{H} \in \Lambda^*} \|\mathbf{H}\|^2 F(\mathbf{A}^T \mathbf{H}) \exp(-2\pi i \mathbf{H} \cdot \mathbf{X}) = \Delta \rho(\mathbf{X}),$$

where

$$\Delta \rho = \sum_{j=1}^3 \frac{\partial^2 \rho}{\partial X_j^2}$$

is the Laplacian of ρ .

The second formula has been used with $|\mathbf{m}| = 1$ or 2 to compute 'differential syntheses' and refine the location of maxima (or other stationary points) in electron-density maps. Indeed, the values at \mathbf{x} of the gradient vector $\nabla \rho$ and Hessian matrix $(\nabla \nabla^T) \rho$ are readily obtained as

$$\begin{aligned} (\nabla \rho)(\mathbf{x}) &= \sum_{\mathbf{h} \in \mathbb{Z}^3} (-2\pi i \mathbf{h}) F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}), \\ [(\nabla \nabla^T) \rho](\mathbf{x}) &= \sum_{\mathbf{h} \in \mathbb{Z}^3} (-4\pi^2 \mathbf{h} \mathbf{h}^T) F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}), \end{aligned}$$

and a step of Newton iteration towards the nearest stationary point of ρ will proceed by

$$\mathbf{x} \mapsto \mathbf{x} - \{[(\nabla \nabla^T) \rho](\mathbf{x})\}^{-1} (\nabla \rho)(\mathbf{x}).$$

The modern use of Fourier transforms to speed up the computation of derivatives for model refinement will be described in Section 1.3.4.4.7.

The converse property is also useful: it relates the derivatives of the continuous transform $\tilde{\mathcal{F}}[\rho^0]$ to the moments of ρ^0 :

$$\frac{\partial^{m_1+m_2+m_3} \tilde{\mathcal{F}}[\rho^0]}{\partial X_1^{m_1} \partial X_2^{m_2} \partial X_3^{m_3}} (\mathbf{H}) = \tilde{\mathcal{F}}[(2\pi i)^{m_1+m_2+m_3} X_1^{m_1} X_2^{m_2} X_3^{m_3} \rho_x^0](\mathbf{H}).$$

For $|\mathbf{m}| = 2$ and $\mathbf{H} = \mathbf{0}$, this identity gives the well known relation between the Hessian matrix of the transform $\tilde{\mathcal{F}}[\rho^0]$ at the origin of reciprocal space and the inertia tensor of the motif ρ^0 . This is a particular case of the moment-generating properties of $\tilde{\mathcal{F}}$, which will be further developed in Section 1.3.4.5.2.

1.3.4.2.1.10. Toeplitz forms, determinantal inequalities and Szegő's theorem

The classical results presented in Section 1.3.2.6.9 can be readily generalized to the case of triple Fourier series; no new concept is needed, only an obvious extension of the notation.

Let ρ be real-valued, so that Friedel's law holds and $F(-\mathbf{h}) = F(\mathbf{h})$. Let \mathbf{H} be a finite set of indices comprising the origin: $\mathbf{H} = \{\mathbf{h}_0 = \mathbf{0}, \mathbf{h}_1, \dots, \mathbf{h}_n\}$. Then the Hermitian form in $n+1$ complex variables

$$T_{\mathbf{H}}[\rho](\mathbf{u}) = \sum_{j, k=0}^n F(\mathbf{h}_j - \mathbf{h}_k) \bar{u}_j u_k$$

is called the Toeplitz form of order \mathbf{H} associated to ρ . By the convolution theorem and Parseval's identity,

$$T_{\mathbf{H}}[\rho](\mathbf{u}) = \int_{\mathbb{R}^3/\mathbb{Z}^3} \rho(\mathbf{x}) \left| \sum_{j=0}^n u_j \exp(2\pi i \mathbf{h}_j \cdot \mathbf{x}) \right|^2 d^3 \mathbf{x}.$$

If ρ is almost everywhere non-negative, then for all \mathbf{H} the forms $T_{\mathbf{H}}[\rho]$ are positive semi-definite and therefore all Toeplitz determinants $D_{\mathbf{H}}[\rho]$ are non-negative, where

$$D_{\mathbf{H}}[\rho] = \det \{ [F(\mathbf{h}_j - \mathbf{h}_k)] \}.$$

The Toeplitz–Carathéodory–Herglotz theorem given in Section 1.3.2.6.9.2 states that the converse is true: if $D_{\mathbf{H}}[\rho] \geq 0$ for all \mathbf{H} , then ρ is almost everywhere non-negative. This result is known in the crystallographic literature through the papers of Karle & Hauptman (1950), MacGillavry (1950), and Goedkoop (1950), following previous work by Harker & Kasper (1948) and Gillis (1948a,b).

Szegő's study of the asymptotic distribution of the eigenvalues of Toeplitz forms as their order tends to infinity remains valid. Some precautions are needed, however, to define the notion of a sequence (\mathbf{H}_k) of finite subsets of indices tending to infinity: it suffices that the \mathbf{H}_k should consist essentially of the reciprocal-lattice points \mathbf{h} contained within a domain of the form $k\Omega$ (k -fold dilation of Ω) where Ω is a convex domain in \mathbb{R}^3 containing the origin (Widom, 1960). Under these circumstances, the eigenvalues $\lambda_{\nu}^{(n)}$ of the Toeplitz forms $T_{\mathbf{H}_k}[\rho]$ become equidistributed with the sample values $\rho_{\nu}^{(n)}$ of ρ on a grid satisfying the Shannon sampling criterion for the data in \mathbf{H}_k (cf. Section 1.3.2.6.9.3).

A particular consequence of this equidistribution is that the geometric means of the $\lambda_{\nu}^{(n)}$ and of the $\rho_{\nu}^{(n)}$ are equal, and hence as in Section 1.3.2.6.9.4

$$\lim_{k \rightarrow \infty} \{D_{\mathbf{H}_k}[\rho]\}^{1/|\mathbf{H}_k|} = \exp \left\{ \int_{\mathbb{R}^3/\mathbb{Z}^3} \log \rho(\mathbf{x}) d^3 \mathbf{x} \right\},$$

where $|\mathbf{H}_k|$ denotes the number of reflections in \mathbf{H}_k . Complementary terms giving a better comparison of the two sides were obtained by Widom (1960, 1975) and Linnik (1975).

1. GENERAL RELATIONSHIPS AND TECHNIQUES

This formula played an important role in the solution of the 2D Ising model by Onsager (1944) (see Montroll *et al.*, 1963). It is also encountered in phasing methods involving the ‘Burg entropy’ (Britten & Collins, 1982; Narayan & Nityananda, 1982; Bricogne, 1982, 1984, 1988).

1.3.4.2.2. Crystal symmetry

1.3.4.2.2.1. Crystallographic groups

The description of a crystal given so far has dealt only with its invariance under the action of the (discrete Abelian) group of translations by vectors of its period lattice Λ .

Let the crystal now be embedded in Euclidean 3-space, so that it may be acted upon by the group $M(3)$ of rigid (*i.e.* distance-preserving) motions of that space. The group $M(3)$ contains a normal subgroup $T(3)$ of translations, and the quotient group $M(3)/T(3)$ may be identified with the 3-dimensional orthogonal group $O(3)$. The period lattice Λ of a crystal is a discrete uniform subgroup of $T(3)$.

The possible invariance properties of a crystal under the action of $M(3)$ are captured by the following definition: a *crystallographic group* is a subgroup Γ of $M(3)$ if

- (i) $\Gamma \cap T(3) = \Lambda$, a period lattice and a normal subgroup of Γ ;
- (ii) the factor group $G = \Gamma/\Lambda$ is finite.

The two properties are not independent: by a theorem of Bieberbach (1911), they follow from the assumption that Λ is a discrete subgroup of $M(3)$ which operates without accumulation point and with a compact fundamental domain (see Auslander, 1965). These two assumptions imply that G acts on Λ through an integral representation, and this observation leads to a complete enumeration of all distinct Γ 's. The mathematical theory of these groups is still an active research topic (see, for instance, Farkas, 1981), and has applications to Riemannian geometry (Wolf, 1967).

This classification of crystallographic groups is described elsewhere in these *Tables* (Wondratschek, 1995), but it will be surveyed briefly in Section 1.3.4.2.2.3 for the purpose of establishing further terminology and notation, after recalling basic notions and results concerning groups and group actions in Section 1.3.4.2.2.2.

1.3.4.2.2.2. Groups and group actions

The books by Hall (1959) and Scott (1964) are recommended as reference works on group theory.

(a) Left and right actions

Let G be a group with identity element e , and let X be a set. An *action* of G on X is a mapping from $G \times X$ to X with the property that, if $g \cdot x$ denotes the image of (g, x) , then

- (i) $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$ for all $g_1, g_2 \in G$ and all $x \in X$,
- (ii) $e \cdot x = x$ for all $x \in X$.

An element g of G thus induces a mapping T_g of X into itself defined by $T_g(x) = g \cdot x$, with the ‘representation property’:

- (iii) $T_{g_1 g_2} = T_{g_1} T_{g_2}$ for all $g_1, g_2 \in G$.

Since G is a group, every g has an inverse g^{-1} ; hence every mapping T_g has an inverse $T_{g^{-1}}$, so that each T_g is a permutation of X .

Strictly speaking, what has just been defined is a *left* action. A *right* action of G on X is defined similarly as a mapping $(g, x) \mapsto xg$ such that

- (i') $x(g_1 g_2) = (xg_1)g_2$ for all $g_1, g_2 \in G$ and all $x \in X$,
- (ii') $xe = x$ for all $x \in X$.

The mapping T'_g defined by $T'_g(x) = xg$ then has the ‘right-representation’ property:

$$(iii') T'_{g_1 g_2} = T'_{g_2} T'_{g_1} \quad \text{for all } g_1, g_2 \in G.$$

The essential difference between left and right actions is of course not whether the elements of G are written on the left or right of those of X : it lies in the difference between (iii) and (iii'). In a left action the product $g_1 g_2$ in G operates on $x \in X$ by g_2 operating first, then g_1 operating on the result; in a right action, g_1 operates first, then g_2 . This distinction will be of importance in Sections 1.3.4.2.2.4 and 1.3.4.2.2.5. In the sequel, we will use left actions unless otherwise stated.

(b) Orbits and isotropy subgroups

Let x be a fixed element of X . Two fundamental entities are associated to x :

- (1) the subset of G consisting of all g such that $gx = x$ is a subgroup of G , called the *isotropy subgroup* of x and denoted G_x ;
- (2) the subset of X consisting of all elements gx with g running through G is called the *orbit* of x under G and is denoted Gx .

Through these definitions, the action of G on X can be related to the internal structure of G , as follows. Let G/G_x denote the collection of distinct left cosets of G_x in G , *i.e.* of distinct subsets of G of the form gG_x . Let $|G|$, $|G_x|$, $|Gx|$ and $|G/G_x|$ denote the numbers of elements in the corresponding sets. The number $|G/G_x|$ of distinct cosets of G_x in G is also denoted $[G : G_x]$ and is called the *index* of G_x in G ; by Lagrange's theorem

$$[G : G_x] = |G/G_x| = \frac{|G|}{|G_x|}.$$

Now if g_1 and g_2 are in the same coset of G_x , then $g_2 = g_1 g'$ with $g' \in G_x$, and hence $g_1 x = g_2 x$; the converse is obviously true. Therefore, the mapping from cosets to orbit elements

$$gG_x \mapsto gx$$

establishes a one-to-one correspondence between the distinct left cosets of G_x in G and the elements of the orbit of x under G . It follows that the number of distinct elements in the orbit of x is equal to the index of G_x in G :

$$|Gx| = [G : G_x] = \frac{|G|}{|G_x|},$$

and that the elements of the orbit of x may be listed without repetition in the form

$$Gx = \{\gamma x | \gamma \in G/G_x\}.$$

Similar definitions may be given for a right action of G on X . The set of distinct right cosets $G_x g$ in G , denoted $G_x \backslash G$, is then in one-to-one correspondence with the distinct elements in the orbit xG of x .

(c) Fundamental domain and orbit decomposition

The group properties of G imply that two orbits under G are either disjoint or equal. The set X may thus be written as the *disjoint* union

$$X = \bigcup_{i \in I} Gx_i,$$

where the x_i are elements of distinct orbits and I is an indexing set labelling them. The subset $D = \{x_i\}_{i \in I}$ is said to constitute a *fundamental domain* (mathematical terminology) or an *asymmetric unit* (crystallographic terminology) for the action of G on X : it contains one representative x_i of each distinct orbit. Clearly, other fundamental domains may be obtained by choosing different representatives for these orbits.

If X is finite and if f is an arbitrary complex-valued function over X , the ‘integral’ of f over X may be written as a sum of integrals over the distinct orbits, yielding the *orbit decomposition formula*:

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

$$\begin{aligned}\sum_{x \in X} f(x) &= \sum_{i \in I} \left(\sum_{y_i \in Gx_i} f(y_i) \right) = \sum_{i \in I} \left(\sum_{\gamma_i \in G/G_{x_i}} f(\gamma_i x_i) \right) \\ &= \sum_{i \in I} \frac{1}{|G_{x_i}|} \left(\sum_{g_i \in G} f(g_i x_i) \right).\end{aligned}$$

In particular, taking $f(x) = 1$ for all x and denoting by $|X|$ the number of elements of X :

$$|X| = \sum_{i \in I} |Gx_i| = \sum_{i \in I} |G/G_{x_i}| = \sum_{i \in I} \frac{|G|}{|G_{x_i}|}.$$

(d) *Conjugation, normal subgroups, semi-direct products*

A group G acts on itself by *conjugation*, i.e. by associating to $g \in G$ the mapping C_g defined by

$$C_g(h) = ghg^{-1}.$$

Indeed, $C_g(hk) = C_g(h)C_g(k)$ and $[C_g(h)]^{-1} = C_{g^{-1}}(h)$. In particular, C_g operates on the set of subgroups of G , two subgroups H and K being called *conjugate* if $H = C_g(K)$ for some $g \in G$; for example, it is easily checked that $G_{gx} = C_g(G_x)$. The orbits under this action are the *conjugacy classes* of subgroups of G , and the isotropy subgroup of H under this action is called the *normalizer* of H in G .

If $\{H\}$ is a one-element orbit, H is called a *self-conjugate* or *normal* subgroup of G ; the cosets of H in G then form a group G/H called the *factor group* of G by H .

Let G and H be two groups, and suppose that G acts on H by *automorphisms* of H , i.e. in such a way that

$$\begin{aligned}g(h_1 h_2) &= g(h_1)g(h_2) \\ g(e_H) &= e_H \quad (\text{where } e_H \text{ is the identity element of } H). \\ g(h^{-1}) &= (g(h))^{-1}\end{aligned}$$

Then the symbols $[g, h]$ with $g \in G, h \in H$ form a group K under the product rule:

$$[g_1, h_1][g_2, h_2] = [g_1 g_2, h_1 g_1(h_2)]$$

{associativity checks; $[e_G, e_H]$ is the identity; $[g, h]$ has inverse $[g^{-1}, g^{-1}(h^{-1})]$ }. The group K is called the *semi-direct product* of H by G , denoted $K = H \rtimes G$.

The elements $[g, e_H]$ form a subgroup of K isomorphic to G , the elements $[e_G, h]$ form a normal subgroup of K isomorphic to H , and the action of G on H may be represented as an action by conjugation in the sense that

$$C_{[g, e_H]}([e_G, h]) = [e_G, g(h)].$$

A familiar example of semi-direct product is provided by the group of Euclidean motions $M(3)$ (Section 1.3.4.2.2.1). An element S of $M(3)$ may be written $S = [R, t]$ with $R \in O(3)$, the orthogonal group, and $t \in T(3)$, the translation group, and the product law

$$[R_1, t_1][R_2, t_2] = [R_1 R_2, t_1 + R_1(t_2)]$$

shows that $M(3) = T(3) \rtimes O(3)$ with $O(3)$ acting on $T(3)$ by rotating the translation vectors.

(e) *Associated actions in function spaces*

For every left action T_g of G in X , there is an associated left action $T_g^\#$ of G on the space $L(X)$ of complex-valued functions over X , defined by ‘change of variable’ (Section 1.3.2.3.9.5):

$$[T_g^\# f](x) = f((T_g)^{-1}x) = f(g^{-1}x).$$

Indeed for any g_1, g_2 in G ,

$$\begin{aligned}[T_{g_1}^\# [T_{g_2}^\# f]](x) &= [T_{g_2}^\# f]((T_{g_1})^{-1}x) = f[(T_{g_2}^{-1} T_{g_1}^{-1})x] \\ &= f((T_{g_1} T_{g_2})^{-1}x);\end{aligned}$$

since $T_{g_1} T_{g_2} = T_{g_1 g_2}$, it follows that

$$T_{g_1}^\# T_{g_2}^\# = T_{g_1 g_2}^\#.$$

It is clear that the change of variable must involve the action of g^{-1} (not g) if $T_g^\#$ is to define a *left* action; using g instead would yield a *right* action.

The linear representation operators $T_g^\#$ on $L(X)$ provide the most natural instrument for stating and exploiting symmetry properties which a function may possess with respect to the action of G . Thus a function $f \in L(X)$ will be called *G-invariant* if $f(gx) = f(x)$ for all $g \in G$ and all $x \in X$. The value $f(x)$ then depends on x only through its orbit Gx , and f is uniquely defined once it is specified on a fundamental domain $D = \{x_i\}_{i \in I}$; its integral over X is then a weighted sum of its values in D :

$$\sum_{x \in X} f(x) = \sum_{i \in I} [G : G_{x_i}] f(x_i).$$

The G -invariance of f may be written:

$$T_g^\# f = f \quad \text{for all } g \in G.$$

Thus f is invariant under each $T_g^\#$, which obviously implies that f is invariant under the linear operator in $L(X)$

$$A_G = \frac{1}{|G|} \sum_{g \in G} T_g^\#,$$

which averages an arbitrary function by the action of G . Conversely, if $A_G f = f$, then

$$T_{g_0}^\# f = T_{g_0}^\# (A_G f) = (T_{g_0}^\# A_G) f = A_G f = f \quad \text{for all } g_0 \in G,$$

so that the two statements of the G -invariance of f are equivalent. The identity

$$T_{g_0}^\# A_G = A_G \quad \text{for all } g_0 \in G$$

is easily proved by observing that the map $g \mapsto g_0 g$ (g_0 being any element of G) is a one-to-one map from G into itself, so that

$$\sum_{g \in G} T_g^\# = \sum_{g \in G} T_{g_0 g}^\#$$

as these sums differ only by the order of the terms. The same identity implies that A_G is a *projector*:

$$(A_G)^2 = A_G,$$

and hence that its eigenvalues are either 0 or 1. In summary, we may say that the invariance of f under G is equivalent to f being an eigenfunction of the associated projector A_G for eigenvalue 1.

(f) *Orbit exchange*

One final result about group actions which will be used repeatedly later is concerned with the case when X has the structure of a Cartesian product:

$$X = X_1 \times X_2 \times \dots \times X_n$$

and when G acts *diagonally* on X , i.e. acts on each X_j separately:

$$gx = g(x_1, x_2, \dots, x_n) = (gx_1, gx_2, \dots, gx_n).$$

Then complete sets (but not usually minimal sets) of representatives

1. GENERAL RELATIONSHIPS AND TECHNIQUES

of the distinct orbits for the action of G in X may be obtained in the form

$$D_k = X_1 \times \dots \times X_{k-1} \times \{x_{i_k}^{(k)}\}_{i_k \in I_k} \times X_{k+1} \times \dots \times X_n$$

for each $k = 1, 2, \dots, n$, *i.e.* by taking a fundamental domain in X_k and all the elements in X_j with $j \neq k$. The action of G on each D_k does indeed generate the whole of X : given an arbitrary element $y = (y_1, y_2, \dots, y_n)$ of X , there is an index $i_k \in I_k$ such that $y_k \in Gx_{i_k}^{(k)}$ and a coset of $G_{x_{i_k}^{(k)}}$ in G such that $y_k = \gamma x_{i_k}^{(k)}$ for any representative γ of that coset; then

$$y = \gamma(\gamma^{-1}y_1, \dots, \gamma^{-1}y_{k-1}, x_{i_k}^{(k)}, \gamma^{-1}y_{k+1}, \dots, \gamma^{-1}y_n)$$

which is of the form $y = \gamma d_k$ with $d_k \in D_k$.

The various D_k are related in a simple manner by ‘transposition’ or ‘orbit exchange’ (the latter name is due to J. W. Cooley). For instance, D_j may be obtained from D_k ($j \neq k$) as follows: for each $y_j \in X_j$ there exists $g(y_j) \in G$ and $i_j(y_j) \in I_j$ such that $y_j = g(y_j)x_{i_j(y_j)}^{(j)}$; therefore

$$D_j = \bigcup_{y_j \in X_j} [g(y_j)]^{-1} D_k,$$

since the fundamental domain of X_k is thus expanded to the whole of X_k , while X_j is reduced to its fundamental domain. In other words: orbits are simultaneously collapsed in the j th factor and expanded in the k th.

When G operates without fixed points in each X_k (*i.e.* $G_{x_k} = \{e\}$ for all $x_k \in X_k$), then each D_k is a fundamental domain for the action of G in X . The existence of fixed points in some or all of the X_k complicates the situation in that for each k and each $x_k \in X_k$ such that $G_{x_k} \neq \{e\}$ the action of G/G_{x_k} on the other factors must be examined. Shenefelt (1988) has made a systematic study of orbit exchange for space group $P622$ and its subgroups.

Orbit exchange will be encountered, in a great diversity of forms, as the basic mechanism by which intermediate results may be rearranged between the successive stages of the computation of crystallographic Fourier transforms (Section 1.3.4.3).

1.3.4.2.2.3. Classification of crystallographic groups

Let Γ be a crystallographic group, Λ the normal subgroup of its lattice translations, and G the finite factor group Γ/Λ . Then G acts on Λ by conjugation [Section 1.3.4.2.2(d)] and this action, being a mapping of a lattice into itself, is representable by matrices with integer entries.

The classification of crystallographic groups proceeds from this observation in the following three steps:

Step 1: find all possible finite abstract groups G which can be represented by 3×3 integer matrices.

Step 2: for each such G find all its inequivalent representations by 3×3 integer matrices, equivalence being defined by a change of primitive lattice basis (*i.e.* conjugation by a 3×3 integer matrix with determinant ± 1).

Step 3: for each G and each equivalence class of integral representations of G , find all inequivalent extensions of the action of G from Λ to $T(3)$, equivalence being defined by an affine coordinate change [*i.e.* conjugation by an element of $A(3)$].

Step 1 leads to the following groups, listed in association with the crystal system to which they later give rise:

$\mathbb{Z}/2\mathbb{Z}$	monoclinic
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	orthorhombic
$\mathbb{Z}/3\mathbb{Z}, (\mathbb{Z}/3\mathbb{Z}) \rtimes \{\alpha\}$	trigonal
$\mathbb{Z}/4\mathbb{Z}, (\mathbb{Z}/4\mathbb{Z}) \rtimes \{\alpha\}$	tetragonal
$\mathbb{Z}/6\mathbb{Z}, (\mathbb{Z}/6\mathbb{Z}) \rtimes \{\alpha\}$	hexagonal
$(\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}) \rtimes \{S_3\}$	cubic

and the extension of these groups by a centre of inversion. In this list \rtimes denotes a semi-direct product [Section 1.3.4.2.2(d)], α denotes the automorphism $g \mapsto g^{-1}$, and S_3 (the group of permutations on three letters) operates by permuting the copies of $\mathbb{Z}/2\mathbb{Z}$ (using the subgroup A_3 of cyclic permutations gives the tetrahedral subsystem).

Step 2 leads to a list of 73 equivalence classes called *arithmetic classes* of representations $g \mapsto \mathbf{R}_g$, where \mathbf{R}_g is a 3×3 integer matrix, with $\mathbf{R}_{g_1 g_2} = \mathbf{R}_{g_1} \mathbf{R}_{g_2}$ and $\mathbf{R}_e = \mathbf{I}_3$. This enumeration is more familiar if equivalence is relaxed so as to allow conjugation by rational 3×3 matrices with determinant ± 1 : this leads to the 32 crystal classes. The difference between an arithmetic class and its rational class resides in the choice of a lattice mode ($P, A/B/C, I, F$ or R). Arithmetic classes always refer to a primitive lattice, but may use inequivalent integral representations for a given geometric symmetry element; while crystallographers prefer to change over to a non-primitive lattice, if necessary, in order to preserve the same integral representation for a given geometric symmetry element. The matrices \mathbf{P} and $\mathbf{Q} = \mathbf{P}^{-1}$ describing the changes of basis between primitive and centred lattices are listed in Table 5.1 and illustrated in Figs. 5.3 to 5.9, pp. 76–79, of Volume A of *International Tables* (Arnold, 1995).

Step 3 gives rise to a system of congruences for the systems of non-primitive translations $\{\mathbf{t}_g\}_{g \in G}$ which may be associated to the matrices $\{\mathbf{R}_g\}_{g \in G}$ of a given arithmetic class, namely:

$$\mathbf{t}_{g_1 g_2} \equiv \mathbf{R}_{g_1} \mathbf{t}_{g_2} + \mathbf{t}_{g_1} \pmod{\Lambda},$$

first derived by Frobenius (1911). If equivalence under the action of $A(3)$ is taken into account, 219 classes are found. If equivalence is defined with respect to the action of the subgroup $A^+(3)$ of $A(3)$ consisting only of transformations with determinant $+1$, then 230 classes called *space-group types* are obtained. In particular, associating to each of the 73 arithmetic classes a trivial set of non-primitive translations ($\mathbf{t}_g = \mathbf{0}$ for all $g \in G$) yields the 73 symmetric space groups. This third step may also be treated as an abstract problem concerning group extensions, using cohomological methods [Ascher & Janner (1965); see Janssen (1973) for a summary]; the connection with Frobenius’s approach, as generalized by Zassenhaus (1948), is examined in Ascher & Janner (1968).

The finiteness of the number of space-group types in dimension 3 was shown by Bieberbach (1912) to be the case in arbitrary dimension. The reader interested in N -dimensional space-group theory for $N > 3$ may consult Brown (1969), Brown *et al.* (1978), Schwarzenberger (1980), and Engel (1986). The standard reference for integral representation theory is Curtis & Reiner (1962).

All three-dimensional space groups G have the property of being *solvable*, *i.e.* that there exists a chain of subgroups

$$G = G_r > G_{r-1} > \dots > G_1 > G_0 = \{e\},$$

where each G_{i-1} is a normal subgroup of G_i and the factor group G_i/G_{i-1} is a *cyclic* group of some order m_i ($1 \leq i \leq r$). This property may be established by inspection, or deduced from a famous theorem of Burnside [see Burnside (1911), pp. 322–323] according to which any group G such that $|G| = p^\alpha q^\beta$, with p and q distinct primes, is solvable; in the case at hand, $p = 2$ and $q = 3$.

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

The whole classification of 3D space groups can be performed swiftly by a judicious use of the solvability property (L. Auslander, personal communication).

Solvability facilitates the indexing of elements of G in terms of generators and relations (Coxeter & Moser, 1972; Magnus *et al.*, 1976) for the purpose of calculation. By definition of solvability, elements g_1, g_2, \dots, g_r may be chosen in such a way that the cyclic factor group G_i/G_{i-1} is generated by the coset $g_i G_{i-1}$. The set $\{g_1, g_2, \dots, g_r\}$ is then a system of generators for G such that the defining relations [see Brown *et al.* (1978), pp. 26–27] have the particularly simple form

$$\begin{aligned} g_1^{m_1} &= e, \\ g_i^{m_i} &= g_{i-1}^{a(i,i-1)} g_{i-2}^{a(i,i-2)} \dots g_1^{a(i,1)} \quad \text{for } 2 \leq i \leq r, \\ g_i^{-1} g_j^{-1} g_i g_j &= g_{j-1}^{b(i,j,j-1)} g_{j-2}^{b(i,j,j-2)} \dots g_1^{b(i,j,1)} \quad \text{for } 1 \leq i < j \leq r, \end{aligned}$$

with $0 \leq a(i, h) < m_h$ and $0 \leq b(i, j, h) < m_h$. Each element g of G may then be obtained uniquely as an ‘ordered word’:

$$g = g_r^{k_r} g_{r-1}^{k_{r-1}} \dots g_1^{k_1},$$

with $0 \leq k_i < m_i$ for all $i = 1, \dots, r$, using the algorithm of Jürgensen (1970). Such generating sets and defining relations are tabulated in Brown *et al.* (1978, pp. 61–76). An alternative list is given in Janssen (1973, Table 4.3, pp. 121–123, and Appendix D, pp. 262–271).

1.3.4.2.2.4. Crystallographic group action in real space

The action of a crystallographic group Γ may be written in terms of standard coordinates in $\mathbb{R}^3/\mathbb{Z}^3$ as

$$(g, \mathbf{x}) \mapsto S_g(\mathbf{x}) = \mathbf{R}_g \mathbf{x} + \mathbf{t}_g \pmod{\Lambda}, \quad g \in G,$$

with

$$S_{g_1 g_2} = S_{g_1} S_{g_2}.$$

An important characteristic of the representation $\theta : g \mapsto S_g$ is its *reducibility*, i.e. whether or not it has invariant subspaces other than $\{\mathbf{0}\}$ and the whole of $\mathbb{R}^3/\mathbb{Z}^3$. For triclinic, monoclinic and orthorhombic space groups, θ is reducible to a direct sum of three one-dimensional representations:

$$\mathbf{R}_g = \begin{pmatrix} \mathbf{R}_g^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_g^{(2)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}_g^{(3)} \end{pmatrix};$$

for trigonal, tetragonal and hexagonal groups, it is reducible to a direct sum of two representations, of dimension 2 and 1, respectively; while for tetrahedral and cubic groups, it is irreducible.

By Schur’s lemma (see *e.g.* Ledermann, 1987), any matrix which commutes with all the matrices \mathbf{R}_g for $g \in G$ must be a scalar multiple of the identity in each invariant subspace.

In the reducible cases, the reductions involve changes of basis which will be *rational*, not integral, for those arithmetic classes corresponding to non-primitive lattices. Thus the simplification of having maximally reduced representation has as its counterpart the use of non-primitive lattices.

The notions of orbit, isotropy subgroup and fundamental domain (or asymmetric unit) for the action of G on $\mathbb{R}^3/\mathbb{Z}^3$ are inherited directly from the general setting of Section 1.3.4.2.2.2. Points \mathbf{x} for which $G_{\mathbf{x}} \neq \{e\}$ are called *special positions*, and the various types of isotropy subgroups which may be encountered in crystallographic groups have been labelled by means of Wyckoff symbols. The representation operators $S_g^\#$ in $L(\mathbb{R}^3/\mathbb{Z}^3)$ have the form:

$$[S_g^\# f](\mathbf{x}) = f[S_g^{-1}(\mathbf{x})] = f[\mathbf{R}_g^{-1}(\mathbf{x} - \mathbf{t}_g)].$$

The operators $R_g^\#$ associated to the purely rotational part of each transformation $S_g^\#$ will also be used. Note the relation: $S_g^\# = \tau_{\mathbf{t}_g} R_g^\#$.

Let a crystal structure be described by the list of the atoms in its unit cell, indexed by $k \in K$. Let the electron-density distribution about the centre of mass of atom k be described by ρ_k with respect to the standard coordinates \mathbf{x} . Then the motif ρ^0 may be written as a sum of translates:

$$\rho^0 = \sum_{k \in K} \tau_{\mathbf{x}_k} \rho_k$$

and the crystal electron density is $\rho = r^* \rho^0$.

Suppose that ρ is invariant under Γ . If \mathbf{x}_{k_1} and \mathbf{x}_{k_2} are in the same orbit, say $\mathbf{x}_{k_2} = S_g(\mathbf{x}_{k_1})$, then

$$\tau_{\mathbf{x}_{k_2}} \rho_{k_2} = S_g^\#(\tau_{\mathbf{x}_{k_1}} \rho_{k_1}).$$

Therefore if \mathbf{x}_k is a special position and thus $G_{\mathbf{x}_k} \neq \{e\}$, then

$$S_g^\#(\tau_{\mathbf{x}_k} \rho_k) = \tau_{\mathbf{x}_k} \rho_k \quad \text{for all } g \in G_{\mathbf{x}_k}.$$

This identity implies that

$$\mathbf{R}_g \mathbf{x}_k + \mathbf{t}_g \equiv \mathbf{x}_k \pmod{\Lambda}$$

(the special position condition), and that

$$\rho_k = R_g^\# \rho_k,$$

i.e. that ρ_k must be invariant by the pure rotational part of $G_{\mathbf{x}_k}$. Trueblood (1956) investigated the consequences of this invariance on the thermal vibration tensor of an atom in a special position (see Section 1.3.4.2.2.6 below).

Let J be a subset of K such that $\{\mathbf{x}_j\}_{j \in J}$ contains exactly one atom from each orbit. An orbit decomposition yields an expression for ρ^0 in terms of symmetry-unique atoms:

$$\rho^0 = \sum_{j \in J} \left(\sum_{\gamma_j \in G/G_{\mathbf{x}_j}} S_{\gamma_j}^\#(\tau_{\mathbf{x}_j} \rho_j) \right)$$

or equivalently

$$\rho^0(\mathbf{x}) = \sum_{j \in J} \left\{ \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \rho_j [\mathbf{R}_{\gamma_j}^{-1}(\mathbf{x} - \mathbf{t}_{\gamma_j}) - \mathbf{x}_j] \right\}.$$

If the atoms are assumed to be Gaussian, write

$$\begin{aligned} \rho_j(\mathbf{X}) &= \frac{Z_j}{|\det \pi \mathbf{U}_j|^{1/2}} \\ &\times \exp\left(-\frac{1}{2} \mathbf{X}^T \mathbf{U}_j^{-1} \mathbf{X}\right) \text{ in Cartesian } \text{\AA} \text{ coordinates,} \end{aligned}$$

where Z_j is the total number of electrons, and where the matrix \mathbf{U}_j combines the Gaussian spread of the electrons in atom j at rest with the covariance matrix of the random positional fluctuations of atom j caused by thermal agitation.

In crystallographic coordinates:

$$\begin{aligned} \rho_j(\mathbf{x}) &= \frac{Z_j}{|\det \pi \mathbf{Q}_j|^{1/2}} \\ &\times \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{Q}_j^{-1} \mathbf{x}\right) \text{ with } \mathbf{Q}_j = \mathbf{A}^{-1} \mathbf{U}_j (\mathbf{A}^{-1})^T. \end{aligned}$$

If atom k is in a special position \mathbf{x}_k , then the matrix \mathbf{Q}_k must satisfy the identity

$$\mathbf{R}_g \mathbf{Q}_k \mathbf{R}_g^{-1} = \mathbf{Q}_k$$

1. GENERAL RELATIONSHIPS AND TECHNIQUES

for all g in the isotropy subgroup of \mathbf{x}_k . This condition may also be written in Cartesian coordinates as

$$\mathbf{T}_g \mathbf{U}_k \mathbf{T}_g^{-1} = \mathbf{U}_k,$$

where

$$\mathbf{T}_g = \mathbf{A} \mathbf{R}_g \mathbf{A}^{-1}.$$

This is a condensed form of the symmetry properties derived by Trueblood (1956).

1.3.4.2.2.5. Crystallographic group action in reciprocal space

An elementary discussion of this topic may be found in Chapter 1.4 of this volume.

Having established that the symmetry of a crystal may be most conveniently stated and handled *via* the left representation $g \mapsto S_g^\#$ of G given by its action on electron-density distributions, it is natural to transpose this action by the identity of Section 1.3.2.5.5:

$$\begin{aligned} \tilde{\mathcal{F}}[S_g^\# T]_\xi &= \tilde{\mathcal{F}}[\tau_{\mathbf{t}_g}(R_g^\# T)]_\xi \\ &= \exp(2\pi i \xi \cdot \mathbf{t}_g) [(\mathbf{R}_g^{-1})^{T\#} \tilde{\mathcal{F}}[T]]_\xi \end{aligned}$$

for any tempered distribution T , *i.e.*

$$\tilde{\mathcal{F}}[S_g^\# T](\xi) = \exp(2\pi i \xi \cdot \mathbf{t}_g) \tilde{\mathcal{F}}[T](\mathbf{R}_g^T \xi)$$

whenever the transforms are functions.

Putting $T = \rho$, a \mathbb{Z}^3 -periodic distribution, this relation defines a left action S_g^* of G on $L(\mathbb{Z}^3)$ given by

$$(S_g^* F)(\mathbf{h}) = \exp(2\pi i \xi \cdot \mathbf{t}_g) F(\mathbf{R}_g^T \mathbf{h})$$

which is conjugate to the action $S_g^\#$ in the sense that

$$\tilde{\mathcal{F}}[S_g^\# \rho] = S_g^* \tilde{\mathcal{F}}[\rho], \quad \text{i.e. } S_g^* = \tilde{\mathcal{F}} S_g^\# \tilde{\mathcal{F}}.$$

The identity $S_g^\# \rho = \rho$ expressing the G -invariance of ρ is then equivalent to the identity $S_g^* F = F$ between its structure factors, *i.e.* (Waser, 1955a)

$$F(\mathbf{h}) = \exp(2\pi i \mathbf{h} \cdot \mathbf{t}_g) F(\mathbf{R}_g^T \mathbf{h}).$$

If G is made to act on \mathbb{Z}^3 *via*

$$\theta^* : (g, \mathbf{h}) \mapsto (\mathbf{R}_g^{-1})^T \mathbf{h},$$

the usual notions of orbit, isotropy subgroup (denoted $G_{\mathbf{h}}$) and fundamental domain may be attached to this action. The above relation then shows that the spectrum $\{F(\mathbf{h})\}_{\mathbf{h} \in \mathbb{Z}^3}$ is entirely known if it is specified on a fundamental domain D^* containing one reciprocal-lattice point from each orbit of this action.

A reflection \mathbf{h} is called *special* if $G_{\mathbf{h}} \neq \{e\}$. Then for any $g \in G_{\mathbf{h}}$ we have $\mathbf{R}_g^T \mathbf{h} = \mathbf{h}$, and hence

$$F(\mathbf{h}) = \exp(2\pi i \mathbf{h} \cdot \mathbf{t}_g) F(\mathbf{h}),$$

implying that $F(\mathbf{h}) = 0$ unless $\mathbf{h} \cdot \mathbf{t}_g \equiv 0 \pmod{1}$. Special reflections \mathbf{h} for which $\mathbf{h} \cdot \mathbf{t}_g \not\equiv 0 \pmod{1}$ for some $g \in G_{\mathbf{h}}$ are thus *systematically absent*. This phenomenon is an instance of the duality between periodization and decimation of Section 1.3.2.7.2: if $\mathbf{t}_g \neq \mathbf{0}$, the projection of ρ on the direction of \mathbf{h} has period $(\mathbf{t}_g \cdot \mathbf{h})/(\mathbf{h} \cdot \mathbf{h}) < 1$, hence its transform (which is the portion of F supported by the central line through \mathbf{h}) will be decimated, giving rise to the above condition.

A reflection \mathbf{h} is called *centric* if $G_{\mathbf{h}} = G(-\mathbf{h})$, *i.e.* if the orbit of \mathbf{h} contains $-\mathbf{h}$. Then $\mathbf{R}_\gamma^T \mathbf{h} = -\mathbf{h}$ for some coset γ in $G/G_{\mathbf{h}}$, so that the following relation must hold:

$$|F(\mathbf{h})| \exp(i\varphi_{\mathbf{h}}) = \exp(2\pi i \mathbf{h} \cdot \mathbf{t}_\gamma) |F(-\mathbf{h})| \exp(i\varphi_{-\mathbf{h}}).$$

In the absence of dispersion, Friedel's law gives rise to the *phase restriction*:

$$\varphi_{\mathbf{h}} \equiv \pi \mathbf{h} \cdot \mathbf{t}_\gamma \pmod{\pi}.$$

The value of the restricted phase is independent of the choice of coset representative γ . Indeed, if γ' is another choice, then $\gamma' = g\gamma$ with $g \in G_{\mathbf{h}}$ and by the Frobenius congruences $\mathbf{t}_{\gamma'} = \mathbf{R}_g \mathbf{t}_\gamma + \mathbf{t}_g$, so that

$$\mathbf{h} \cdot \mathbf{t}_{\gamma'} \equiv (\mathbf{R}_g^T \mathbf{h}) \cdot \mathbf{t}_\gamma + \mathbf{h} \cdot \mathbf{t}_g \pmod{1}.$$

Since $g \in G_{\mathbf{h}}$, $\mathbf{R}_g^T \mathbf{h} = \mathbf{h}$ and $\mathbf{h} \cdot \mathbf{t}_g \equiv 0 \pmod{1}$ if \mathbf{h} is not a systematic absence: thus

$$\pi \mathbf{h} \cdot \mathbf{t}_{\gamma'} \equiv \pi \mathbf{h} \cdot \mathbf{t}_\gamma \pmod{\pi}.$$

The treatment of centred lattices may be viewed as another instance of the duality between periodization and decimation (Section 1.3.2.7.2): the periodization of the electron density by the non-primitive lattice translations has as its counterpart in reciprocal space the decimation of the transform by the 'reflection conditions' describing the allowed reflections, the decimation and periodization matrices being each other's contragredient.

The reader may consult the papers by Bienenstock & Ewald (1962) and Wells (1965) for earlier approaches to this material.

1.3.4.2.2.6. Structure-factor calculation

Structure factors may be calculated from a list of symmetry-unique atoms by Fourier transformation of the orbit decomposition formula for the motif ρ^0 given in Section 1.3.4.2.2.4:

$$\begin{aligned} F(\mathbf{h}) &= \tilde{\mathcal{F}}[\rho^0](\mathbf{h}) \\ &= \tilde{\mathcal{F}} \left[\sum_{j \in J} \left(\sum_{\gamma_j \in G/G_{\mathbf{x}_j}} S_{\gamma_j}^\#(\tau_{\mathbf{x}_j} \rho_j) \right) \right](\mathbf{h}) \\ &= \sum_{j \in J} \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \tilde{\mathcal{F}}[\tau_{\mathbf{x}_j} \mathbf{R}_{\gamma_j}^\# \tau_{\mathbf{x}_j} \rho_j](\mathbf{h}) \\ &= \sum_{j \in J} \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp(2\pi i \mathbf{h} \cdot \mathbf{t}_{\gamma_j}) \\ &\quad \times [(\mathbf{R}_{\gamma_j}^{-1})^{T\#} [\exp(2\pi i \xi \cdot \mathbf{x}_j) \tilde{\mathcal{F}}[\rho_j]_\xi]](\mathbf{h}) \\ &= \sum_{j \in J} \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp(2\pi i \mathbf{h} \cdot \mathbf{t}_{\gamma_j}) \\ &\quad \times \exp[2\pi i (\mathbf{R}_{\gamma_j}^T \mathbf{h}) \cdot \mathbf{x}_j] \tilde{\mathcal{F}}[\rho_j](\mathbf{R}_{\gamma_j}^T \mathbf{h}); \end{aligned}$$

i.e. finally:

$$F(\mathbf{h}) = \sum_{j \in J} \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp\{2\pi i \mathbf{h} \cdot [S_{\gamma_j}(\mathbf{x}_j)]\} \tilde{\mathcal{F}}[\rho_j](\mathbf{R}_{\gamma_j}^T \mathbf{h}).$$

In the case of Gaussian atoms, the atomic transforms are

$$\tilde{\mathcal{F}}[\rho_j](\mathbf{h}) = Z_j \exp[-\frac{1}{2} \mathbf{h}^T (4\pi^2 \mathbf{Q}_j) \mathbf{h}]$$

or equivalently

$$\tilde{\mathcal{F}}[\rho_j](\mathbf{H}) = Z_j \exp[-\frac{1}{2} \mathbf{H}^T (4\pi^2 \mathbf{U}_j) \mathbf{H}].$$

Two common forms of equivalent temperature factors (incorporating both atomic form and thermal motion) are

(i) isotropic B :

$$\tilde{\mathcal{F}}[\rho_j](\mathbf{h}) = Z_j \exp(-\frac{1}{4} B_j \mathbf{H}^T \mathbf{H}),$$

so that $\mathbf{U}_j = (B_j/8\pi^2) \mathbf{I}$, or $\mathbf{Q}_j = (B_j/8\pi^2) \mathbf{A}^{-1} (\mathbf{A}^{-1})^T$;

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

(ii) anisotropic β 's:

$$\tilde{\mathcal{F}}[\rho_j](\mathbf{h}) = Z_j \exp(-\mathbf{h}^T \boldsymbol{\beta} \mathbf{h}),$$

so that $\boldsymbol{\beta}_j = 2\pi^2 \mathbf{Q}_j = 2\pi^2 \mathbf{A}^{-1} \mathbf{U}_j (\mathbf{A}^{-1})^T$, or $\mathbf{U}_j = (1/2\pi^2) \mathbf{A} \boldsymbol{\beta}_j \mathbf{A}^T$.

In the first case, $\tilde{\mathcal{F}}[\rho_j](\mathbf{R}_\gamma^T \mathbf{h})$ does not depend on γ_j , and therefore:

$$F(\mathbf{h}) = \sum_{j \in J} Z_j \exp\{-\frac{1}{4} \boldsymbol{\beta}_j \mathbf{h}^T [\mathbf{A}^{-1} (\mathbf{A}^{-1})^T] \mathbf{h}\} \\ \times \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp\{2\pi i \mathbf{h} \cdot [S_{\gamma_j}(\mathbf{x}_j)]\}.$$

In the second case, however, no such simplification can occur:

$$F(\mathbf{h}) = \sum_{j \in J} Z_j \sum_{\gamma_j \in G/G_{\mathbf{x}_j}} \exp[-\mathbf{h}^T (\mathbf{R}_{\gamma_j} \boldsymbol{\beta}_j \mathbf{R}_{\gamma_j}^T) \mathbf{h}] \\ \times \exp\{2\pi i \mathbf{h} \cdot [S_{\gamma_j}(\mathbf{x}_j)]\}.$$

These formulae, or special cases of them, were derived by Rollett & Davies (1955), Waser (1955b), and Trueblood (1956).

The computation of structure factors by applying the discrete Fourier transform to a set of electron-density values calculated on a grid will be examined in Section 1.3.4.4.5.

1.3.4.2.2.7. Electron-density calculations

A formula for the Fourier synthesis of electron-density maps from symmetry-unique structure factors is readily obtained by orbit decomposition:

$$\rho(\mathbf{x}) = \sum_{\mathbf{h} \in \mathbb{Z}^3} F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}) \\ = \sum_{l \in L} \left[\sum_{\gamma_l \in G/G_{\mathbf{h}_l}} F(\mathbf{R}_{\gamma_l}^T \mathbf{h}_l) \exp[-2\pi i (\mathbf{R}_{\gamma_l}^T \mathbf{h}_l) \cdot \mathbf{x}] \right] \\ = \sum_{l \in L} F(\mathbf{h}_l) \left[\sum_{\gamma_l \in G/G_{\mathbf{h}_l}} \exp\{-2\pi i \mathbf{h}_l \cdot [S_{\gamma_l}(\mathbf{x})]\} \right],$$

where L is a subset of \mathbb{Z}^3 such that $\{\mathbf{h}_l\}_{l \in L}$ contains exactly one point of each orbit for the action $\theta^* : (g, \mathbf{h}) \mapsto (\mathbf{R}_g^{-1})^T \mathbf{h}$ of G on \mathbb{Z}^3 . The physical electron density per cubic ångström is then

$$\rho(\mathbf{X}) = \frac{1}{V} \rho(\mathbf{A}\mathbf{x})$$

with V in Å^3 .

In the absence of anomalous scatterers in the crystal and of a centre of inversion $-\mathbf{I}$ in Γ , the spectrum $\{F(\mathbf{h})\}_{\mathbf{h} \in \mathbb{Z}^3}$ has an extra symmetry, namely the Hermitian symmetry expressing Friedel's law (Section 1.3.4.2.1.4). The action of a centre of inversion may be added to that of Γ to obtain further simplification in the above formula: under this extra action, an orbit $G\mathbf{h}_l$ with $\mathbf{h}_l \neq \mathbf{0}$ is either mapped into itself or into the disjoint orbit $G(-\mathbf{h}_l)$; the terms corresponding to $+\mathbf{h}_l$ and $-\mathbf{h}_l$ may then be grouped *within* the common orbit in the first case, and *between* the two orbits in the second case.

Case 1: $G(-\mathbf{h}_l) = G\mathbf{h}_l$, \mathbf{h}_l is *centric*. The cosets in $G/G_{\mathbf{h}_l}$ may be partitioned into two disjoint classes by picking one coset in each of the two-coset orbits of the action of $-\mathbf{I}$. Let $(G/G_{\mathbf{h}_l})^+$ denote one such class: then the *reduced orbit*

$$\{\mathbf{R}_{\gamma_l}^T \mathbf{h}_l | \gamma_l \in (G/G_{\mathbf{h}_l})^+\}$$

contains exactly once the Friedel-unique half of the full orbit $G\mathbf{h}_l$, and thus

$$|(G/G_{\mathbf{h}_l})^+| = \frac{1}{2} |G/G_{\mathbf{h}_l}|.$$

Grouping the summands for $+\mathbf{h}_l$ and $-\mathbf{h}_l$ yields a real-valued summand

$$2F(\mathbf{h}_l) \sum_{\gamma_l \in (G/G_{\mathbf{h}_l})^+} \cos[2\pi \mathbf{h}_l \cdot [S_{\gamma_l}(\mathbf{x})] - \varphi_{\mathbf{h}_l}].$$

Case 2: $G(-\mathbf{h}_l) \neq G\mathbf{h}_l$, \mathbf{h}_l is *acentric*. The two orbits are then disjoint, and the summands corresponding to $+\mathbf{h}_l$ and $-\mathbf{h}_l$ may be grouped together into a single real-valued summand

$$2F(\mathbf{h}_l) \sum_{\gamma_l \in G/G_{\mathbf{h}_l}} \cos[2\pi \mathbf{h}_l \cdot [S_{\gamma_l}(\mathbf{x})] - \varphi_{\mathbf{h}_l}].$$

In order to reindex the collection of all summands of ρ , put

$$L = L_c \cup L_a,$$

where L_c labels the Friedel-unique centric reflections in L and L_a the acentric ones, and let L_a^+ stand for a subset of L_a containing a unique element of each pair $\{+\mathbf{h}_l, -\mathbf{h}_l\}$ for $l \in L_a$. Then

$$\rho(\mathbf{x}) = F(\mathbf{0}) \\ + \sum_{c \in L_c} \left[2F(\mathbf{h}_c) \sum_{\gamma_c \in (G/G_{\mathbf{h}_c})^+} \cos[2\pi \mathbf{h}_c \cdot [S_{\gamma_c}(\mathbf{x})] - \varphi_{\mathbf{h}_c}] \right] \\ + \sum_{a \in L_a^+} \left[2F(\mathbf{h}_a) \sum_{\gamma_a \in G/G_{\mathbf{h}_a}} \cos[2\pi \mathbf{h}_a \cdot [S_{\gamma_a}(\mathbf{x})] - \varphi_{\mathbf{h}_a}] \right].$$

1.3.4.2.2.8. Parseval's theorem with crystallographic symmetry

The general statement of Parseval's theorem given in Section 1.3.4.2.1.5 may be rewritten in terms of symmetry-unique structure factors and electron densities by means of orbit decomposition.

In reciprocal space,

$$\sum_{\mathbf{h} \in \mathbb{Z}^3} \overline{F_1(\mathbf{h})} F_2(\mathbf{h}) = \sum_{l \in L} \sum_{\gamma_l \in G/G_{\mathbf{h}_l}} \overline{F_1(\mathbf{R}_{\gamma_l}^T \mathbf{h}_l)} F_2(\mathbf{R}_{\gamma_l}^T \mathbf{h}_l);$$

for each l , the summands corresponding to the various γ_l are equal, so that the left-hand side is equal to

$$F_1(\mathbf{0}) F_2(\mathbf{0}) \\ + \sum_{c \in L_c} 2|(G/G_{\mathbf{h}_c})^+| |F_1(\mathbf{h}_c)| |F_2(\mathbf{h}_c)| \cos[\varphi_1(\mathbf{h}_c) - \varphi_2(\mathbf{h}_c)] \\ + \sum_{a \in L_a^+} 2|G/G_{\mathbf{h}_a}| |F_1(\mathbf{h}_a)| |F_2(\mathbf{h}_a)| \cos[\varphi_1(\mathbf{h}_a) - \varphi_2(\mathbf{h}_a)].$$

In real space, the triple integral may be rewritten as

$$\int_{\mathbb{R}^3/\mathbb{Z}^3} \overline{\rho_1(\mathbf{x})} \rho_2(\mathbf{x}) d^3 \mathbf{x} = |G| \int_D \overline{\rho_1(\mathbf{x})} \rho_2(\mathbf{x}) d^3 \mathbf{x}$$

(where D is the asymmetric unit) if ρ_1 and ρ_2 are smooth densities, since the set of special positions has measure zero. If, however, the integral is approximated as a sum over a G -invariant grid defined by decimation matrix \mathbf{N} , special positions on this grid must be taken into account:

$$\frac{1}{|\mathbf{N}|} \sum_{\mathbf{k} \in \mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3} \overline{\rho_1(\mathbf{x})} \rho_2(\mathbf{x}) \\ = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{x} \in D} [G : G_{\mathbf{x}}] \overline{\rho_1(\mathbf{x})} \rho_2(\mathbf{x}) \\ = \frac{|G|}{|\mathbf{N}|} \sum_{\mathbf{x} \in D} \frac{1}{|G_{\mathbf{x}}|} \overline{\rho_1(\mathbf{x})} \rho_2(\mathbf{x}),$$

where the discrete asymmetric unit D contains exactly one point in each orbit of G in $\mathbb{Z}^3/\mathbf{N}\mathbb{Z}^3$.

1. GENERAL RELATIONSHIPS AND TECHNIQUES

1.3.4.2.2.9. Convolution theorems with crystallographic symmetry

The standard convolution theorems derived in the absence of symmetry are readily seen to follow from simple properties of functions $e^\pm(\mathbf{h}, \mathbf{x}) = \exp(\pm 2\pi i \mathbf{h} \cdot \mathbf{x})$ (denoted simply e in formulae which are valid for both signs), namely:

- (i) $e(\mathbf{h}, \mathbf{x}) \times e(\mathbf{k}, \mathbf{x}) = e(\mathbf{h} + \mathbf{k}, \mathbf{x})$,
- (ii) $e(\mathbf{h}, \mathbf{x}) \times e(\mathbf{h}, \mathbf{y}) = e(\mathbf{h}, \mathbf{x} + \mathbf{y})$.

These relations imply that the families of functions

$$\{\mathbf{x} \mapsto e(\mathbf{h}, \mathbf{x})\}_{\mathbf{h} \in \mathbb{Z}^3} \quad \text{in real space}$$

and

$$\{\mathbf{h} \mapsto e(\mathbf{h}, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^3/\mathbb{Z}^3} \quad \text{in reciprocal space}$$

both generate an *algebra* of functions, *i.e.* a vector space endowed with an internal multiplication, since (i) and (ii) show how to ‘linearize products’.

Friedel’s law (when applicable) on the one hand, and the Fourier relation between intensities and the Patterson function on the other hand, both follow from the property

$$(iii) \quad \overline{e(\mathbf{h}, \mathbf{x})} = e(-\mathbf{h}, \mathbf{x}) = e(\mathbf{h}, -\mathbf{x}).$$

When crystallographic symmetry is present, the convolution theorems remain valid in their original form if written out in terms of ‘expanded’ data, but acquire a different form when rewritten in terms of symmetry-unique data only. This rewriting is made possible by the extra relation (Section 1.3.4.2.2.5)

$$(iv) \quad S_{g^{-1}}^\# e(\mathbf{h}, \mathbf{x}) \equiv e[\mathbf{h}, S_g(\mathbf{x})] = e(\mathbf{h}, \mathbf{t}_g) e(\mathbf{R}_g^T \mathbf{h}, \mathbf{x})$$

or equivalently

$$(iv') \quad S_g^\# e(\mathbf{h}, \mathbf{x}) \equiv e[\mathbf{h}, S_g^{-1}(\mathbf{x})] \\ = e[(-\mathbf{R}_g^{-1})^T \mathbf{h}, \mathbf{t}_g] e[(\mathbf{R}_g^{-1})^T \mathbf{h}, \mathbf{x}].$$

The kernels of symmetrized Fourier transforms are not the functions e but rather the symmetrized sums

$$\Xi^\pm(\mathbf{h}, \mathbf{x}) = \sum_{g \in G} e^\pm[\mathbf{h}, S_g(\mathbf{x})] = \sum_{g \in G} e^\pm[\mathbf{h}, S_g^{-1}(\mathbf{x})]$$

for which the linearization formulae are readily obtained using (i), (ii) and (iv) as

- (i)_G $\Xi^\pm(\mathbf{h}, \mathbf{x}) \Xi^\pm(\mathbf{k}, \mathbf{x}) = \sum_{g \in G} e^\pm(\mathbf{k}, \mathbf{t}_g) \Xi^\pm(\mathbf{h} + \mathbf{R}_g^T \mathbf{k}, \mathbf{x})$,
- (ii)_G $\Xi^\pm(\mathbf{h}, \mathbf{x}) \Xi^\pm(\mathbf{h}, \mathbf{y}) = \sum_{g \in G} \Xi^\pm[\mathbf{h}, \mathbf{x} + S_g(\mathbf{y})]$,

where the choice of sign in \pm must be the same throughout each formula.

Formulae (i)_G defining the ‘structure-factor algebra’ associated to G were derived by Bertaut (1955c, 1956b,c, 1959a,b) and Bertaut & Waser (1957) in another context.

The forward convolution theorem (in discrete form) then follows. Let

$$F_1(\mathbf{h}) = \sum_{\mathbf{y} \in D} \frac{1}{|G_{\mathbf{y}}|} \rho_1(\mathbf{y}) \Xi^+(\mathbf{h}, \mathbf{y}), \\ F_2(\mathbf{h}) = \sum_{\mathbf{z} \in D} \frac{1}{|G_{\mathbf{z}}|} \rho_2(\mathbf{z}) \Xi^+(\mathbf{h}, \mathbf{z}),$$

then

$$F_1(\mathbf{h}) F_2(\mathbf{h}) = \sum_{\mathbf{x} \in D} \frac{1}{|G_{\mathbf{x}}|} \sigma(\mathbf{x}) \Xi^+(\mathbf{h}, \mathbf{x})$$

with

$$\sigma(\mathbf{x}) = \frac{1}{|N|} \sum_{\mathbf{z} \in D} \sum_{g \in G} \frac{|G_{\mathbf{x}}|}{|G_{\mathbf{x}-S_g(\mathbf{z})}| \times |G_{\mathbf{z}}|} \rho_1[\mathbf{x} - S_g(\mathbf{z})] \rho_2(\mathbf{z}).$$

The backward convolution theorem is derived similarly. Let

$$\rho_1(\mathbf{x}) = \sum_{\mathbf{k} \in D^*} \frac{1}{|G_{\mathbf{k}}|} F_1(\mathbf{k}) \Xi^-(\mathbf{k}, \mathbf{x}), \\ \rho_2(\mathbf{x}) = \sum_{\mathbf{l} \in D^*} \frac{1}{|G_{\mathbf{l}}|} F_2(\mathbf{l}) \Xi^-(\mathbf{l}, \mathbf{x}),$$

then

$$\rho_1(\mathbf{x}) \rho_2(\mathbf{x}) = \sum_{\mathbf{h} \in D^*} \frac{1}{|G_{\mathbf{h}}|} F(\mathbf{h}) \Xi^-(\mathbf{h}, \mathbf{x})$$

with

$$F(\mathbf{h}) = \sum_{\mathbf{l} \in D^*} \sum_{g \in G} \frac{|G_{\mathbf{h}}|}{|G_{\mathbf{h}-\mathbf{R}_g^T(\mathbf{l})}| \times |G_{\mathbf{l}}|} e^-(\mathbf{l}, \mathbf{t}_g) F_1(\mathbf{h} - \mathbf{R}_g^T \mathbf{l}) F_2(\mathbf{l}).$$

Both formulae are simply orbit decompositions of their symmetry-free counterparts.

1.3.4.2.2.10. Correlation and Patterson functions

Consider two model electron densities ρ_1 and ρ_2 with the same period lattice \mathbb{Z}^3 and the same space group G . Write their motifs in terms of atomic electron densities (Section 1.3.4.2.2.4) as

$$\rho_1^0 = \sum_{j_1 \in J_1} \left(\sum_{\gamma_{j_1} \in G/G_{x_{j_1}^{(1)}}} S_{\gamma_{j_1}}^\# (\tau_{x_{j_1}^{(1)}} \rho_{j_1}^{(1)}) \right), \\ \rho_2^0 = \sum_{j_2 \in J_2} \left(\sum_{\gamma_{j_2} \in G/G_{x_{j_2}^{(2)}}} S_{\gamma_{j_2}}^\# (\tau_{x_{j_2}^{(2)}} \rho_{j_2}^{(2)}) \right),$$

where J_1 and J_2 label the symmetry-unique atoms placed at positions $\{x_{j_1}^{(1)}\}_{j_1 \in J_1}$ and $\{x_{j_2}^{(2)}\}_{j_2 \in J_2}$, respectively.

To calculate the correlation between ρ_1 and ρ_2 we need the following preliminary formulae, which are easily established: if $S(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t}$ and f is an arbitrary function on \mathbb{R}^3 , then

$$(R^\# f)^\checkmark = R^\# \checkmark f, \quad (\tau_{\mathbf{x}} f)^\checkmark = \tau_{-\mathbf{x}} \checkmark f, \quad R^\# (\tau_{\mathbf{x}} f) = \tau_{\mathbf{R}\mathbf{x}} f,$$

hence

$$S^\# (\tau_{\mathbf{x}} f) = \tau_{S(\mathbf{x})} R^\# f \quad \text{and} \quad [S^\# (\tau_{\mathbf{x}} f)]^\checkmark = \tau_{-S(\mathbf{x})} R^\# \checkmark f;$$

and

$$S_1^\# f_1 * S_2^\# f_2 = S_1^\# [f_1 * (S_1^{-1} S_2)^\# f_2] = S_2^\# [(S_2^{-1} S_1)^\# f_1 * f_2].$$

The cross correlation $\rho_1^0 * \rho_2^0$ between motifs is therefore

$$\rho_1^0 * \rho_2^0 = \sum_{j_1} \sum_{j_2} \sum_{\gamma_{j_1}} \sum_{\gamma_{j_2}} [S_{\gamma_{j_1}}^\# (\tau_{x_{j_1}^{(1)}} \rho_{j_1}^{(1)})]^\checkmark * [S_{\gamma_{j_2}}^\# (\tau_{x_{j_2}^{(2)}} \rho_{j_2}^{(2)})] \\ = \sum_{j_1} \sum_{j_2} \sum_{\gamma_{j_1}} \sum_{\gamma_{j_2}} \tau_{S_{\gamma_{j_2}}(x_{j_2}^{(2)}) - S_{\gamma_{j_1}}(x_{j_1}^{(1)})} [(R_{\gamma_{j_1}}^\# \checkmark \rho_{j_1}^{(1)}) * (R_{\gamma_{j_2}}^\# \checkmark \rho_{j_2}^{(2)})]$$

which contains a peak of shape $(R_{\gamma_{j_1}}^\# \checkmark \rho_{j_1}^{(1)}) * (R_{\gamma_{j_2}}^\# \checkmark \rho_{j_2}^{(2)})$ at the interatomic vector $S_{\gamma_{j_2}}(x_{j_2}^{(2)}) - S_{\gamma_{j_1}}(x_{j_1}^{(1)})$ for each $j_1 \in J_1$, $j_2 \in J_2$, $\gamma_{j_1} \in G/G_{x_{j_1}^{(1)}}$, $\gamma_{j_2} \in G/G_{x_{j_2}^{(2)}}$.

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

The cross-correlation $r * \rho_1^0 * \rho_2^0$ between the original electron densities is then obtained by further periodizing by \mathbb{Z}^3 .

Note that these expressions are valid for any choice of ‘atomic’ density functions $\rho_{j_1}^{(1)}$ and $\rho_{j_2}^{(2)}$, which may be taken as molecular fragments if desired (see Section 1.3.4.4.8).

If G contains elements g such that \mathbf{R}_g has an eigenspace E_1 with eigenvalue 1 and an invariant complementary subspace E_2 , while \mathbf{t}_g has a non-zero component $\mathbf{t}_g^{(1)}$ in E_1 , then the Patterson function $r * \rho^0 * \rho^0$ will contain *Harker peaks* (Harker, 1936) of the form

$$S_g(\mathbf{x}) - \mathbf{x} = \mathbf{t}_g^{(1)} \oplus (S_g^{(2)}(\mathbf{x}) - \mathbf{x})$$

[where $S_g^{(s)}$ represent the action of g in E_2] in the translate of E_1 by $\mathbf{t}_g^{(1)}$.

1.3.4.3. Crystallographic discrete Fourier transform algorithms

1.3.4.3.1. Historical introduction

In 1929, W. L. Bragg demonstrated the practical usefulness of the Fourier transform relation between electron density and structure factors by determining the structure of diopside from three principal projections calculated numerically by 2D Fourier summation (Bragg, 1929). It was immediately realized that the systematic use of this powerful method, and of its extension to three dimensions, would entail considerable amounts of numerical computation which had to be organized efficiently. As no other branch of applied science had yet needed this type of computation, crystallographers had to invent their own techniques.

The first step was taken by Beevers & Lipson (1934) who pointed out that a 2D summation could be factored into successive 1D summations. This is essentially the tensor product property of the Fourier transform (Sections 1.3.2.4.2.4, 1.3.3.3.1), although its aspect is rendered somewhat complicated by the use of sines and cosines instead of complex exponentials. Computation is economized to the extent that the cost of an $N \times N$ transform grows with N as $2N^3$ rather than N^4 . Generalization to 3D is immediate, reducing computation size from N^6 to $3N^4$ for an $N \times N \times N$ transform. The complication introduced by using expressions in terms of sines and cosines is turned to advantage when symmetry is present, as certain families of terms are systematically absent or are simply related to each other; multiplicity corrections must, however, be introduced. The necessary information was tabulated for each space group by Lonsdale (1936), and was later incorporated into Volume I of *International Tables*.

The second step was taken by Beevers & Lipson (1936) and Lipson & Beevers (1936) in the form of the invention of the ‘Beevers–Lipson strips’, a practical device which was to assist a whole generation of crystallographers in the numerical computation of crystallographic Fourier sums. The strips comprise a set of ‘cosine strips’ tabulating the functions

$$A \cos\left(\frac{2\pi hm}{60}\right) \quad (A = 1, 2, \dots, 99; h = 1, 2, \dots, 99)$$

and a set of ‘sine strips’ tabulating the functions

$$B \sin\left(\frac{2\pi hm}{60}\right) \quad (B = 1, 2, \dots, 99; h = 1, 2, \dots, 99)$$

for the 16 arguments $m = 0, 1, \dots, 15$. Function values are rounded to the nearest integer, and those for other arguments m may be obtained by using the symmetry properties of the sine and cosine functions. A Fourier summation of the form

$$Y(m) = \sum_{j=1}^n \left[A_j \cos\left(\frac{2\pi h_j m}{60}\right) + B_j \sin\left(\frac{2\pi h_j m}{60}\right) \right]$$

is then performed by selecting the n cosine strips labelled (A_j, h_j) and the n sine strips labelled (B_j, h_j) , placing them in register, and adding the tabulated values columnwise. The number 60 was chosen as the l.c.m. of 12 (itself the l.c.m. of the orders of all possible non-primitive translations) and of 10 (for decimal convenience). The limited accuracy imposed by the two-digit tabulation was later improved by Robertson’s sorting board (Robertson, 1936*a,b*) or by the use of separate strips for each decimal digit of the amplitude (Booth, 1948*b*), which allowed three-digit tabulation while keeping the set of strips within manageable size. Cochran (1948*a*) found that, for most structures under study at the time, the numerical inaccuracies of the method were less than the level of error in the experimental data. The sampling rate was subsequently increased from 60 to 120 (Beevers, 1952) to cope with larger unit cells.

Further gains in speed and accuracy were sought through the construction of special-purpose mechanical, electro-mechanical, electronic or optical devices. Two striking examples are the mechanical computer RUFUS built by Robertson (1954, 1955, 1961) on the principle of previous strip methods (see also Robertson, 1932) and the electronic analogue computer X-RAC built by Pepinsky, capable of real-time calculation and display of 2D and 3D Fourier syntheses (Pepinsky, 1947; Pepinsky & Sayre, 1948; Pepinsky *et al.*, 1961; see also Suryan, 1957). The optical methods of Lipson & Taylor (1951, 1958) also deserve mention. Many other ingenious devices were invented, whose descriptions may be found in Booth (1948*b*), Niggli (1961), and Lipson & Cochran (1968).

Later, commercial punched-card machines were programmed to carry out Fourier summations or structure-factor calculations (Shaffer *et al.*, 1946*a,b*; Cox *et al.*, 1947, 1949; Cox & Jeffrey, 1949; Donohue & Schomaker, 1949; Grems & Kasper, 1949; Hodgson *et al.*, 1949; Greenhalgh & Jeffrey, 1950; Kitz & Marchington, 1953).

The modern era of digital electronic computation of Fourier series was initiated by the work of Bennett & Kendrew (1952), Mayer & Trueblood (1953), Ahmed & Cruickshank (1953*b*), Sparks *et al.* (1956) and Fowweather (1955). Their Fourier-synthesis programs used Beevers–Lipson factorization, the program by Sparks *et al.* being the first 3D Fourier program useable for all space groups (although these were treated as $P1$ or $P\bar{1}$ by data expansion). Ahmed & Barnes (1958) then proposed a general programming technique to allow full use of symmetry elements (orthorhombic or lower) in the 3D Beevers–Lipson factorization process, including multiplicity corrections. Their method was later adopted by Shoemaker & Sly (1961), and by crystallographic program writers at large.

The discovery of the FFT algorithm by Cooley & Tukey in 1965, which instantly transformed electrical engineering and several other disciplines, paradoxically failed to have an immediate impact on crystallographic computing. A plausible explanation is that the calculation of large 3D Fourier maps was a relatively infrequent task which was not thought to constitute a bottleneck, as crystallographers had learned to settle most structural questions by means of cheaper 2D sections or projections. It is significant in this respect that the first use of the FFT in crystallography by Barrett & Zwick (1971) should have occurred as part of an iterative scheme for improving protein phases by density modification in real space, which required a much greater number of Fourier transformations than any previous method. Independently, Bondot (1971) had attracted attention to the merits of the FFT algorithm.

The FFT program used by Barrett & Zwick had been written for signal-processing applications. It was restricted to sampling rates of the form 2^n , and was not designed to take advantage of crystallographic symmetry at any stage of the calculation; Bantz & Zwick (1974) later improved this situation somewhat.

1. GENERAL RELATIONSHIPS AND TECHNIQUES

It was the work of Ten Eyck (1973) and Immirzi (1973, 1976) which led to the general adoption of the FFT in crystallographic computing. Immirzi treated all space groups as $P1$ by data expansion. Ten Eyck based his program on a versatile multi-radix FFT routine (Gentleman & Sande, 1966) coupled with a flexible indexing scheme for dealing efficiently with multidimensional transforms. He also addressed the problems of incorporating symmetry elements of order 2 into the factorization of 1D transforms, and of transposing intermediate results by other symmetry elements. He was thus able to show that in a large number of space groups (including the 74 space groups having orthorhombic or lower symmetry) it is possible to calculate only the unique results from the unique data within the logic of the FFT algorithm. Ten Eyck wrote and circulated a package of programs for computing Fourier maps and re-analysing them into structure factors in some simple space groups ($P1$, $P\bar{1}$, $P2$, $P2/m$, $P2_1$, $P22_2$, $P2_12_12_1$, $Pmmm$). This package was later augmented by a handful of new space-group-specific programs contributed by other crystallographers ($P2_12_12$, $I222$, $P3_121$, $P4_12_12$). The writing of such programs is an undertaking of substantial complexity, which has deterred all but the bravest: the usual practice is now to expand data for a high-symmetry space group to the largest subgroup for which a specific FFT program exists in the package, rather than attempt to write a new program. Attempts have been made to introduce more modern approaches to the calculation of crystallographic Fourier transforms (Auslander, Feig & Winograd, 1982; Auslander & Shenefelt, 1987; Auslander *et al.*, 1988) but have not gone beyond the stage of preliminary studies.

The task of fully exploiting the FFT algorithm in crystallographic computations is therefore still unfinished, and it is the purpose of this section to provide a systematic treatment such as that (say) of Ahmed & Barnes (1958) for the Beevers–Lipson algorithm.

Ten Eyck's approach, based on the reducibility of certain space groups, is extended by the derivation of a universal transposition formula for intermediate results. It is then shown that space groups which are not completely reducible may nevertheless be treated by three-dimensional Cooley–Tukey factorization in such a way that their symmetry may be fully exploited, whatever the shape of their asymmetric unit. Finally, new factorization methods with built-in symmetries are presented. The unifying concept throughout this presentation is that of 'group action' on indexing sets, and of 'orbit exchange' when this action has a composite structure; it affords new ways of rationalizing the use of symmetry, or of improving computational speed, or both.

1.3.4.3.2. Defining relations and symmetry considerations

A finite set of reflections $\{F_{\mathbf{h}}\}_{\mathbf{h} \in L}$ can be periodized without aliasing by the translations of a suitable sublattice $\mathbf{N}^T \Lambda^*$ of the reciprocal lattice Λ^* ; the converse operation in real space is the sampling of ρ at points \mathbf{X} of a grid of the form $\mathbf{N}^{-1} \Lambda$ (Section 1.3.2.7.3). In standard coordinates, $\{F_{\mathbf{h}}\}_{\mathbf{h} \in L}$ is periodized by $\mathbf{N}^T \mathbb{Z}^3$, and ρ is sampled at points $\mathbf{x} \in \mathbf{N}^{-1} \mathbb{Z}^3$.

In the absence of symmetry, the unique data are
 – the $F_{\mathbf{h}}$ indexed by $\mathbf{h} \in \mathbb{Z}^3 / \mathbf{N}^T \mathbb{Z}^3$ in reciprocal space;
 – the $\rho_{\mathbf{x}}$ indexed by $\mathbf{x} \in (\mathbf{N}^{-1} \mathbb{Z}^3) / \mathbb{Z}^3$; or equivalently the $\rho_{\mathbf{m}}$ indexed by $\mathbf{m} \in \mathbb{Z}^3 / \mathbf{N} \mathbb{Z}^3$, where $\mathbf{x} = \mathbf{N}^{-1} \mathbf{m}$.

They are connected by the ordinary DFT relations:

$$F_{\mathbf{h}} = \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{x} \in (\mathbf{N}^{-1} \mathbb{Z}^3) / \mathbb{Z}^3} \rho_{\mathbf{x}} \exp(2\pi i \mathbf{h} \cdot \mathbf{x})$$

or

$$F_{\mathbf{h}} = \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{m} \in \mathbb{Z}^3 / \mathbf{N} \mathbb{Z}^3} \rho_{\mathbf{m}} \exp[2\pi i \mathbf{h} \cdot (\mathbf{N}^{-1} \mathbf{m})]$$

and

$$\rho_{\mathbf{x}} = \sum_{\mathbf{h} \in \mathbb{Z}^3 / \mathbf{N}^T \mathbb{Z}^3} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

or

$$\rho_{\mathbf{m}} = \sum_{\mathbf{h} \in \mathbb{Z}^3 / \mathbf{N}^T \mathbb{Z}^3} F_{\mathbf{h}} \exp[-2\pi i \mathbf{h} \cdot (\mathbf{N}^{-1} \mathbf{m})].$$

In the presence of symmetry, the unique data are
 – $\{\rho_{\mathbf{x}}\}_{\mathbf{x} \in D}$ or $\{\rho_{\mathbf{m}}\}_{\mathbf{m} \in D}$ in real space (by abuse of notation, D will denote an asymmetric unit for \mathbf{x} or for \mathbf{m} indifferently);
 – $\{F_{\mathbf{h}}\}_{\mathbf{h} \in D^*}$ in reciprocal space.

The previous summations may then be subjected to orbital decomposition, to yield the following 'crystallographic DFT' (CDFT) defining relations:

$$\begin{aligned} F_{\mathbf{h}} &= \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{x} \in D} \rho_{\mathbf{x}} \left[\sum_{\gamma \in G/G_{\mathbf{x}}} \exp\{2\pi i \mathbf{h} \cdot [S_{\gamma}(\mathbf{x})]\} \right] \\ &= \frac{1}{|\det \mathbf{N}|} \sum_{\mathbf{x} \in D} \rho_{\mathbf{x}} \left[\frac{1}{|G_{\mathbf{x}}|} \sum_{g \in G} \exp\{2\pi i \mathbf{h} \cdot [S_g(\mathbf{x})]\} \right], \\ \rho_{\mathbf{x}} &= \sum_{\mathbf{h} \in D^*} F_{\mathbf{h}} \left[\sum_{\gamma \in G/G_{\mathbf{x}}} \exp\{-2\pi i \mathbf{h} \cdot [S_{\gamma}(\mathbf{x})]\} \right] \\ &= \sum_{\mathbf{h} \in D^*} F_{\mathbf{h}} \left[\frac{1}{|G_{\mathbf{h}}|} \sum_{g \in G} \exp\{-2\pi i \mathbf{h} \cdot [S_g(\mathbf{x})]\} \right], \end{aligned}$$

with the obvious alternatives in terms of $\rho_{\mathbf{m}}$, $\mathbf{m} = \mathbf{N}\mathbf{x}$. Our problem is to evaluate the CDFT for a given space group as efficiently as possible, in spite of the fact that the group action has spoiled the simple tensor-product structure of the ordinary three-dimensional DFT (Section 1.3.3.3.1).

Two procedures are available to carry out the 3D summations involved as a succession of smaller summations:

(1) *decomposition* into successive transforms of *fewer dimensions* but on the *same number of points* along these dimensions. This possibility depends on the reducibility of the space group, as defined in Section 1.3.4.2.2.4, and simply invokes the tensor product property of the DFT;

(2) *factorization* of the transform into transforms of the *same number of dimensions* as the original one, but on *fewer points* along each dimension. This possibility depends on the arithmetic factorability of the decimation matrix \mathbf{N} , as described in Section 1.3.3.3.2.

Clearly, a symmetry expansion to the largest fully reducible subgroup of the space group will give maximal decomposability, but will require computing more than the unique results from more than the unique data. Economy will follow from factoring the transforms in the subspaces within which the space group acts irreducibly.

For irreducible subspaces of dimension 1, the group action is readily incorporated into the factorization of the transform, as first shown by Ten Eyck (1973).

For irreducible subspaces of dimension 2 or 3, the ease of incorporation of symmetry into the factorization depends on the type of factorization method used. The multidimensional Cooley–Tukey method (Section 1.3.3.3.1) is rather complicated; the multidimensional Good method (Section 1.3.3.3.2.2) is somewhat simpler; and the Rader/Winograd factorization admits a generalization, based on the arithmetic of certain rings of *algebraic*

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

integers, which accommodates 2D crystallographic symmetries in a most powerful and pleasing fashion.

At each stage of the calculation, it is necessary to keep track of the definition of the asymmetric unit and of the symmetry properties of the numbers being manipulated. This requirement applies not only to the initial data and to the final results, where these are familiar; but also to all the intermediate quantities produced by partial transforms (on subsets of factors, or subsets of dimensions, or both), where they are less familiar. Here, the general formalism of transposition (or 'orbit exchange') described in Section 1.3.4.2.2.2 plays a central role.

1.3.4.3.3. Interaction between symmetry and decomposition

Suppose that the space-group action is reducible, *i.e.* that for each $g \in G$

$$\mathbf{R}_g = \begin{pmatrix} \mathbf{R}'_g & \mathbf{0} \\ \mathbf{0} & \mathbf{R}''_g \end{pmatrix}, \quad \mathbf{t}_g = \begin{pmatrix} \mathbf{t}'_g \\ \mathbf{t}''_g \end{pmatrix};$$

by Schur's lemma, the decimation matrix must then be of the form

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}' & \mathbf{0} \\ \mathbf{0} & \mathbf{N}'' \end{pmatrix} \text{ if it is to commute with all the } \mathbf{R}_g.$$

Putting $\mathbf{x} = \begin{pmatrix} \mathbf{x}' \\ \mathbf{x}'' \end{pmatrix}$ and $\mathbf{h} = \begin{pmatrix} \mathbf{h}' \\ \mathbf{h}'' \end{pmatrix}$, we may define

$$\begin{aligned} S'_g(\mathbf{x}') &= \mathbf{R}'_g \mathbf{x}' + \mathbf{t}'_g, \\ S''_g(\mathbf{x}'') &= \mathbf{R}''_g \mathbf{x}'' + \mathbf{t}''_g, \end{aligned}$$

and write $S_g = S'_g \oplus S''_g$ (direct sum) as a shorthand for $S_g(\mathbf{x}) = \begin{pmatrix} S'_g(\mathbf{x}') \\ S''_g(\mathbf{x}'') \end{pmatrix}$.

We may also define the representation operators $S_g^{\#}$ and $S_g^{*\#}$ acting on functions of \mathbf{x}' and \mathbf{x}'' , respectively (as in Section 1.3.4.2.2.4), and the operators S_g^* and $S_g^{*\#}$ acting on functions of \mathbf{h}' and \mathbf{h}'' , respectively (as in Section 1.3.4.2.2.5). Then we may write

$$S_g^{\#} = (S'_g)^{\#} \oplus (S''_g)^{\#}$$

and

$$S_g^* = (S'_g)^* \oplus (S''_g)^*$$

in the sense that g acts on $f(\mathbf{x}) \equiv f(\mathbf{x}', \mathbf{x}'')$ by

$$(S_g^{\#} f)(\mathbf{x}', \mathbf{x}'') = f[(S'_g)^{-1}(\mathbf{x}'), (S''_g)^{-1}(\mathbf{x}'')]$$

and on $\Phi(\mathbf{h}) \equiv \Phi(\mathbf{h}', \mathbf{h}'')$ by

$$\begin{aligned} (S_g^* \Phi)(\mathbf{h}', \mathbf{h}'') &= \exp(2\pi i \mathbf{h}' \cdot \mathbf{t}'_g) \exp(2\pi i \mathbf{h}'' \cdot \mathbf{t}''_g) \\ &\quad \times \Phi[\mathbf{R}_g^T \mathbf{h}', \mathbf{R}_g^T \mathbf{h}'']. \end{aligned}$$

Thus equipped we may now derive concisely a general identity describing the symmetry properties of intermediate quantities of the form

$$\begin{aligned} T(\mathbf{x}', \mathbf{h}'') &= \sum_{\mathbf{h}'} F(\mathbf{h}', \mathbf{h}'') \exp(-2\pi i \mathbf{h}' \cdot \mathbf{x}') \\ &= \frac{1}{|\det \mathbf{N}'|} \sum_{\mathbf{x}''} \rho(\mathbf{x}', \mathbf{x}'') \exp(+2\pi i \mathbf{h}'' \cdot \mathbf{x}''), \end{aligned}$$

which arise through partial transformation of F on \mathbf{h}' or of ρ on \mathbf{x}'' . The action of $g \in G$ on these quantities will be

- (i) through $(S'_g)^{\#}$ on the function $\mathbf{x}' \mapsto T(\mathbf{x}', \mathbf{h}'')$,
- (ii) through $(S''_g)^*$ on the function $\mathbf{h}'' \mapsto T(\mathbf{x}', \mathbf{h}'')$,

and hence the symmetry properties of T are expressed by the identity

$$T = [(S'_g)^{\#} \oplus (S''_g)^*] T.$$

Applying this relation not to T but to $[(S'_{g-1})^{\#} \oplus (S''_e)^*] T$ gives

$$[(S'_{g-1})^{\#} \oplus (S''_e)^*] T = [(S'_e)^{\#} \oplus (S''_g)^*] T,$$

i.e.

$$T(S'_g(\mathbf{x}'), \mathbf{h}'') = \exp(2\pi i \mathbf{h}'' \cdot \mathbf{t}'_g) T(\mathbf{x}', \mathbf{R}_g^T \mathbf{h}'').$$

If the unique $F(\mathbf{h}) \equiv F(\mathbf{h}', \mathbf{h}'')$ were initially indexed by

$$(\text{all } \mathbf{h}') \times (\text{unique } \mathbf{h}'')$$

(see Section 1.3.4.2.2.2), this formula allows the reindexing of the intermediate results $T(\mathbf{x}', \mathbf{h}'')$ from the initial form

$$(\text{all } \mathbf{x}') \times (\text{unique } \mathbf{h}'')$$

to the final form

$$(\text{unique } \mathbf{x}') \times (\text{all } \mathbf{h}''),$$

on which the second transform (on \mathbf{h}'') may now be performed, giving the final results $\rho(\mathbf{x}', \mathbf{x}'')$ indexed by

$$(\text{unique } \mathbf{x}') \times (\text{all } \mathbf{x}''),$$

which is an asymmetric unit. An analogous interpretation holds if one is going from ρ to F .

The above formula solves the general problem of transposing from one invariant subspace to another, and is the main device for decomposing the CDFT. Particular instances of this formula were derived and used by Ten Eyck (1973); it is useful for orthorhombic groups, and for dihedral groups containing screw axes n_m with g.c.d. $(m, n) = 1$. For comparison with later uses of orbit exchange, it should be noted that the type of intermediate results just dealt with is obtained after transforming on *all* factors in *one* summand.

A central piece of information for driving such a decomposition is the definition of the full asymmetric unit in terms of the asymmetric units in the invariant subspaces. As indicated at the end of Section 1.3.4.2.2.2, this is straightforward when G acts without fixed points, but becomes more involved if fixed points do exist. To this day, no systematic 'calculus of asymmetric units' exists which can automatically generate a complete description of the asymmetric unit of an arbitrary space group in a form suitable for directing the orbit exchange process, although Shenefelt (1988) has outlined a procedure for dealing with space group $P622$ and its subgroups. The asymmetric unit definitions given in Volume A of *International Tables* are incomplete in this respect, in that they do not specify the possible residual symmetries which may exist on the boundaries of the domains.

1.3.4.3.4. Interaction between symmetry and factorization

Methods for factoring the DFT in the absence of symmetry were examined in Sections 1.3.3.2 and 1.3.3.3. They are based on the observation that the finite sets which index both data and results are endowed with certain algebraic structures (*e.g.* are Abelian groups, or rings), and that subsets of indices may be found which are not merely subsets but *substructures* (*e.g.* subgroups or subrings). Summation over these substructures leads to partial transforms, and the way in which substructures fit into the global structure indicates how to reassemble the partial results into the final results. As a rule, the richer the algebraic structure which is identified in the indexing set, the more powerful the factoring method.

1. GENERAL RELATIONSHIPS AND TECHNIQUES

The ability of a given factoring method to accommodate crystallographic symmetry will thus be determined by the extent to which the crystallographic group action respects (or fails to respect) the partitioning of the index set into the substructures pertaining to that method. This remark justifies trying to gain an overall view of the algebraic structures involved, and of the possibilities of a crystallographic group acting ‘naturally’ on them.

The index sets $\{\mathbf{m} | \mathbf{m} \in \mathbb{Z}^3 / \mathbf{N}\mathbb{Z}^3\}$ and $\{\mathbf{h} | \mathbf{h} \in \mathbb{Z}^3 / \mathbf{N}^T\mathbb{Z}^3\}$ are finite Abelian groups under component-wise addition. If an iterated addition is viewed as an action of an integer scalar $n \in \mathbb{Z}$ via

$$\begin{aligned} n\mathbf{h} &= \mathbf{h} + \mathbf{h} + \dots + \mathbf{h} && (n \text{ times}) && \text{for } n > 0, \\ &= \mathbf{0} && && \text{for } n = 0, \\ &= -(\mathbf{h} + \mathbf{h} + \dots + \mathbf{h}) && (|n| \text{ times}) && \text{for } n < 0, \end{aligned}$$

then an Abelian group becomes a *module* over the ring \mathbb{Z} (or, for short, a \mathbb{Z} -module), a module being analogous to a vector space but with scalars drawn from a ring rather than a field. The left actions of a crystallographic group G by

$$g : \mathbf{m} \mapsto \mathbf{R}_g \mathbf{m} + \mathbf{N} \mathbf{t}_g \pmod{\mathbf{N}\mathbb{Z}^3}$$

and by

$$g : \mathbf{h} \mapsto (\mathbf{R}_g^{-1})^T \mathbf{h} \pmod{\mathbf{N}^T\mathbb{Z}^3}$$

can be combined with this \mathbb{Z} action as follows:

$$\begin{aligned} \sum_{g \in G} n_g g : \mathbf{m} &\mapsto \sum_{g \in G} n_g (\mathbf{R}_g \mathbf{m} + \mathbf{N} \mathbf{t}_g) \pmod{\mathbf{N}\mathbb{Z}^3}, \\ \sum_{g \in G} n_g g : \mathbf{h} &\mapsto \sum_{g \in G} n_g [(\mathbf{R}_g^{-1})^T \mathbf{h}] \pmod{\mathbf{N}^T\mathbb{Z}^3}. \end{aligned}$$

This provides a left action, on the indexing sets, of the set

$$\mathbb{Z}G = \left\{ \sum_{g \in G} n_g g \mid n_g \in \mathbb{Z} \text{ for each } g \in G \right\}$$

of symbolic linear combinations of elements of G with integral coefficients. If addition and multiplication are defined in $\mathbb{Z}G$ by

$$\left(\sum_{g_1 \in G} a_{g_1} g_1 \right) + \left(\sum_{g_2 \in G} b_{g_2} g_2 \right) = \sum_{g \in G} (a_g + b_g) g$$

and

$$\left(\sum_{g_1 \in G} a_{g_1} g_1 \right) \times \left(\sum_{g_2 \in G} b_{g_2} g_2 \right) = \sum_{g \in G} c_g g,$$

with

$$c_g = \sum_{g' \in G} a_{g'} b_{(g')^{-1}g},$$

then $\mathbb{Z}G$ is a *ring*, and the action defined above makes the indexing sets into $\mathbb{Z}G$ -modules. The ring $\mathbb{Z}G$ is called the *integral group ring* of G (Curtis & Reiner, 1962, p. 44).

From the algebraic standpoint, therefore, the interaction between symmetry and factorization can be expected to be favourable whenever the indexing sets of partial transforms are $\mathbb{Z}G$ -submodules of the main $\mathbb{Z}G$ -modules.

1.3.4.3.4.1. Multidimensional Cooley–Tukey factorization

Suppose, as in Section 1.3.3.3.2.1, that the decimation matrix \mathbf{N} may be factored as $\mathbf{N}_1 \mathbf{N}_2$. Then any grid point index $\mathbf{m} \in \mathbb{Z}^3 / \mathbf{N}\mathbb{Z}^3$ in real space may be written

$$\mathbf{m} = \mathbf{m}_1 + \mathbf{N}_1 \mathbf{m}_2$$

with $\mathbf{m}_1 \in \mathbb{Z}^3 / \mathbf{N}_1 \mathbb{Z}^3$ and $\mathbf{m}_2 \in \mathbb{Z}^3 / \mathbf{N}_2 \mathbb{Z}^3$ determined by

$$\begin{aligned} \mathbf{m}_1 &= \mathbf{m} \pmod{\mathbf{N}_1 \mathbb{Z}^3}, \\ \mathbf{m}_2 &= \mathbf{N}_1^{-1} (\mathbf{m} - \mathbf{m}_1) \pmod{\mathbf{N}_2 \mathbb{Z}^3}. \end{aligned}$$

These relations establish a one-to-one correspondence $\mathbf{m} \leftrightarrow (\mathbf{m}_1, \mathbf{m}_2)$ between $I = \mathbb{Z}^3 / \mathbf{N}\mathbb{Z}^3$ and the Cartesian product $I_1 \times I_2$ of $I_1 = \mathbb{Z}^3 / \mathbf{N}_1 \mathbb{Z}^3$ and $I_2 = \mathbb{Z}^3 / \mathbf{N}_2 \mathbb{Z}^3$, and hence $I \cong I_1 \times I_2$ as a set. However $I \not\cong I_1 \times I_2$ as an Abelian group, since in general $\mathbf{m} + \mathbf{m}' \not\leftrightarrow (\mathbf{m}_1 + \mathbf{m}'_1, \mathbf{m}_2 + \mathbf{m}'_2)$ because there can be a ‘carry’ from the addition of the first components into the second components; therefore, $I \not\cong I_1 \times I_2$ as a $\mathbb{Z}G$ -module, which shows that the incorporation of symmetry into the Cooley–Tukey algorithm is not a trivial matter.

Let $g \in G$ act on I through

$$g : \mathbf{m} \mapsto S_g(\mathbf{m}) = \mathbf{R}_g \mathbf{m} + \mathbf{N} \mathbf{t}_g \pmod{\mathbf{N}\mathbb{Z}^3}$$

and suppose that \mathbf{N} ‘integerizes’ all the non-primitive translations \mathbf{t}_g so that we may write

$$\mathbf{N} \mathbf{t}_g = \mathbf{t}_g^{(1)} + \mathbf{N}_1 \mathbf{t}_g^{(2)},$$

with $\mathbf{t}_g^{(1)} \in I_1$ and $\mathbf{t}_g^{(2)} \in I_2$ determined as above. Suppose further that \mathbf{N} , \mathbf{N}_1 and \mathbf{N}_2 commute with \mathbf{R}_g for all $g \in G$, i.e. (by Schur’s lemma, Section 1.3.4.2.2.4) that these matrices are integer multiples of the identity in each G -invariant subspace. The action of g on $\mathbf{m} = \mathbf{N} \mathbf{x} \pmod{\mathbf{N}\mathbb{Z}^3}$ leads to

$$\begin{aligned} S_g(\mathbf{m}) &= \mathbf{N}[\mathbf{R}_g(\mathbf{N}^{-1} \mathbf{m}) + \mathbf{N} \mathbf{t}_g] \pmod{\mathbf{N}\mathbb{Z}^3} \\ &= \mathbf{N} \mathbf{R}_g \mathbf{N}^{-1} (\mathbf{m}_1 + \mathbf{N}_1 \mathbf{m}_2) + \mathbf{t}_g^{(1)} + \mathbf{N}_1 \mathbf{t}_g^{(2)} \pmod{\mathbf{N}\mathbb{Z}^3} \\ &= \mathbf{R}_g \mathbf{m}_1 + \mathbf{t}_g^{(1)} + \mathbf{N}_1 (\mathbf{R}_g \mathbf{m}_2 + \mathbf{t}_g^{(2)}) \pmod{\mathbf{N}\mathbb{Z}^3}, \end{aligned}$$

which we may decompose as

$$S_g(\mathbf{m}) = [S_g(\mathbf{m})]_1 + \mathbf{N}_1 [S_g(\mathbf{m})]_2$$

with

$$[S_g(\mathbf{m})]_1 \equiv S_g(\mathbf{m}) \pmod{\mathbf{N}_1 \mathbb{Z}^3}$$

and

$$[S_g(\mathbf{m})]_2 \equiv \mathbf{N}_1^{-1} \{S_g(\mathbf{m}) - [S_g(\mathbf{m})]_1\} \pmod{\mathbf{N}_2 \mathbb{Z}^3}.$$

Introducing the notation

$$\begin{aligned} S_g^{(1)}(\mathbf{m}_1) &= \mathbf{R}_g \mathbf{m}_1 + \mathbf{t}_g^{(1)} \pmod{\mathbf{N}_1 \mathbb{Z}^3}, \\ S_g^{(2)}(\mathbf{m}_2) &= \mathbf{R}_g \mathbf{m}_2 + \mathbf{t}_g^{(2)} \pmod{\mathbf{N}_2 \mathbb{Z}^3}, \end{aligned}$$

the two components of $S_g(\mathbf{m})$ may be written

$$\begin{aligned} [S_g(\mathbf{m})]_1 &= S_g^{(1)}(\mathbf{m}_1), \\ [S_g(\mathbf{m})]_2 &= S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \pmod{\mathbf{N}_2 \mathbb{Z}^3}, \end{aligned}$$

with

$$\boldsymbol{\mu}_2(g, \mathbf{m}_1) = \mathbf{N}_1^{-1} \{(\mathbf{R}_g \mathbf{m}_1 + \mathbf{t}_g^{(1)}) - [S_g(\mathbf{m}_1)]_1\} \pmod{\mathbf{N}_2 \mathbb{Z}^3}.$$

The term $\boldsymbol{\mu}_2$ is the geometric equivalent of a *carry* or *borrow*: it arises because $\mathbf{R}_g \mathbf{m}_1 + \mathbf{t}_g^{(1)}$, calculated as a vector in $\mathbb{Z}^3 / \mathbf{N}_1 \mathbb{Z}^3$, may be outside the unit cell $\mathbf{N}_1[0, 1]^3$, and may need to be brought back into it by a ‘large’ translation with a non-zero component in the \mathbf{m}_2 space; equivalently, the action of g may need to be applied around different permissible origins for different values of \mathbf{m}_1 , so as to map the unit cell into itself without any recourse to lattice translations. [Readers familiar with the cohomology of groups (see e.g. Hall, 1959; MacLane, 1963) will recognize $\boldsymbol{\mu}_2$ as the cocycle of the extension of $\mathbb{Z}G$ -modules described by the exact sequence $0 \rightarrow I_2 \rightarrow I \rightarrow I_1 \rightarrow 0$.]

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

Thus G acts on I in a rather complicated fashion: although $g \mapsto S_g^{(1)}$ does define a left action in I_1 alone, no action can be defined in I_2 alone because $\boldsymbol{\mu}_2$ depends on \mathbf{m}_1 . However, because $S_g, S_g^{(1)}$ and $S_g^{(2)}$ are left actions, it follows that $\boldsymbol{\mu}_2$ satisfies the identity

$$\boldsymbol{\mu}_2(gg', \mathbf{m}_1) = S_g^{(2)}[\boldsymbol{\mu}_2(g', \mathbf{m}_1)] + \boldsymbol{\mu}_2[g, S_g^{(1)}(\mathbf{m}_1)] \quad \text{mod } \mathbf{N}_2\mathbb{Z}^3$$

for all g, g' in G and all \mathbf{m}_1 in I_1 . In particular, $\boldsymbol{\mu}_2(\mathbf{e}, \mathbf{m}_1) = \mathbf{0}$ for all \mathbf{m}_1 , and

$$\boldsymbol{\mu}_2(g^{-1}, \mathbf{m}_1) = -S_{g^{-1}}^{(2)}\{\boldsymbol{\mu}_2[g, S_g^{(1)}(\mathbf{m}_1)]\} \quad \text{mod } \mathbf{N}_2\mathbb{Z}^3.$$

This action will now be used to achieve optimal use of symmetry in the multidimensional Cooley–Tukey algorithm of Section 1.3.3.2.1. Let us form an array Y according to

$$Y(\mathbf{m}_1, \mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$$

for all $\mathbf{m}_2 \in I_2$ but only for the *unique* \mathbf{m}_1 under the action $S_g^{(1)}$ of G in I_1 . Except in special cases which will be examined later, these vectors contain essentially an asymmetric unit of electron-density data, up to some redundancies on boundaries. We may then compute the partial transform on \mathbf{m}_2 :

$$Y^*(\mathbf{m}_1, \mathbf{h}_2) = \frac{1}{|\det \mathbf{N}_2|} \sum_{\mathbf{m}_2 \in I_2} Y(\mathbf{m}_1, \mathbf{m}_2) e[\mathbf{h}_2 \cdot (\mathbf{N}_2^{-1}\mathbf{m}_2)].$$

Using the symmetry of ρ in the form $\rho = S_g^\# \rho$ yields by the procedure of Section 1.3.3.2 the transposition formula

$$Y^*(S_g^{(1)}(\mathbf{m}_1), \mathbf{h}_2) = e\{\mathbf{h}_2 \cdot [\mathbf{N}_2^{-1}(\mathbf{t}_g^{(2)} + \boldsymbol{\mu}_2(g, \mathbf{m}_1))]\} \\ \times Y^*(\mathbf{m}_1, [\mathbf{R}_g^{(2)}]^T \mathbf{h}_2).$$

By means of this identity we can transpose intermediate results Y^* initially indexed by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{h}_2),$$

so as to have them indexed by

$$(\text{all } \mathbf{m}_1) \times (\text{unique } \mathbf{h}_2).$$

We may then apply twiddle factors to get

$$Z(\mathbf{m}_1, \mathbf{h}_2) = e[\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] Y^*(\mathbf{m}_1, \mathbf{h}_2)$$

and carry out the second transform

$$Z^*(\mathbf{h}_1, \mathbf{h}_2) = \frac{1}{|\det \mathbf{N}_1|} \sum_{\mathbf{m}_1 \in I_1} Z(\mathbf{m}_1, \mathbf{h}_2) e[\mathbf{h}_1 \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)].$$

The final results are indexed by

$$(\text{all } \mathbf{h}_1) \times (\text{unique } \mathbf{h}_2),$$

which yield essentially an asymmetric unit of structure factors after unscrambling by:

$$F(\mathbf{h}_2 + \mathbf{N}_2^T \mathbf{h}_1) = Z^*(\mathbf{h}_1, \mathbf{h}_2).$$

The transposition formula above applies to intermediate results when going backwards from F to ρ , provided these results are considered *after* the twiddle-factor stage. A transposition formula applicable *before* that stage can be obtained by characterizing the action of G on \mathbf{h} (including the effects of periodization by $\mathbf{N}^T\mathbb{Z}^3$) in a manner similar to that used for \mathbf{m} .

Let

$$\mathbf{h} = \mathbf{h}_2 + \mathbf{N}_2^T \mathbf{h}_1,$$

with

$$\mathbf{h}_2 = \mathbf{h} \quad \text{mod } \mathbf{N}_2^T \mathbb{Z}^3,$$

$$\mathbf{h}_1 = (\mathbf{N}_2^{-1})^T (\mathbf{h} - \mathbf{h}_2) \quad \text{mod } \mathbf{N}_1^T \mathbb{Z}^3.$$

We may then write

$$\mathbf{R}_g^T \mathbf{h} = [\mathbf{R}_g^T \mathbf{h}]_2 + \mathbf{N}_2^T [\mathbf{R}_g^T \mathbf{h}]_1,$$

with

$$[\mathbf{R}_g^T \mathbf{h}]_2 = [\mathbf{R}_g^{(2)}]^T \mathbf{h}_2 \quad \text{mod } \mathbf{N}_2^T \mathbb{Z}^3,$$

$$[\mathbf{R}_g^T \mathbf{h}]_1 = [\mathbf{R}_g^{(1)}]^T \mathbf{h}_1 + \boldsymbol{\eta}_1(g, \mathbf{h}_2) \quad \text{mod } \mathbf{N}_1^T \mathbb{Z}^3.$$

Here $[\mathbf{R}_g^{(2)}]^T, [\mathbf{R}_g^{(1)}]^T$ and $\boldsymbol{\eta}_1$ are defined by

$$[\mathbf{R}_g^{(2)}]^T \mathbf{h}_2 = \mathbf{R}_g^T \mathbf{h} \quad \text{mod } \mathbf{N}_2^T \mathbb{Z}^3,$$

$$[\mathbf{R}_g^{(1)}]^T \mathbf{h}_1 = \mathbf{R}_g^T \mathbf{h} \quad \text{mod } \mathbf{N}_1^T \mathbb{Z}^3$$

and

$$\boldsymbol{\eta}_1(g, \mathbf{h}_2) = (\mathbf{N}_2^{-1})^T (\mathbf{R}_g^T \mathbf{h}_2 - [\mathbf{R}_g^{(2)}]^T \mathbf{h}_2) \quad \text{mod } \mathbf{N}_1^T \mathbb{Z}^3.$$

Let us then form an array Z^* according to

$$Z^*(\mathbf{h}'_1, \mathbf{h}'_2) = F(\mathbf{h}'_2 + \mathbf{N}_2^T \mathbf{h}'_1)$$

for all \mathbf{h}'_1 but only for the *unique* \mathbf{h}'_2 under the action of G in $\mathbb{Z}^3/\mathbf{N}_2^T \mathbb{Z}^3$, and transform on \mathbf{h}'_1 to obtain

$$Z(\mathbf{m}_1, \mathbf{h}_2) = \sum_{\mathbf{h}'_1 \in \mathbb{Z}^3/\mathbf{N}_1^T \mathbb{Z}^3} Z^*(\mathbf{h}'_1, \mathbf{h}'_2) e[-\mathbf{h}'_1 \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)].$$

Putting $\mathbf{h}' = \mathbf{R}_g^T \mathbf{h}$ and using the symmetry of F in the form

$$F(\mathbf{h}') = F(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{t}_g),$$

where

$$\mathbf{h} \cdot \mathbf{t}_g = (\mathbf{h}_2^T + \mathbf{h}_1^T \mathbf{N}_2)(\mathbf{N}_2^{-1} \mathbf{N}_1^{-1})(\mathbf{t}_g^{(1)} + \mathbf{N}_1 \mathbf{t}_g^{(2)}) \\ \equiv \mathbf{h}_2 \cdot \mathbf{t}_g + \mathbf{h}_2 \cdot (\mathbf{N}_1^{-1} \mathbf{t}_g^{(1)}) \quad \text{mod } 1$$

yields by a straightforward rearrangement

$$Z(\mathbf{m}_1, [\mathbf{R}_g^{(2)}]^T \mathbf{h}_2) = e[-\{\mathbf{h}_2 \cdot \mathbf{t}_g + \boldsymbol{\eta}_1(g, \mathbf{h}_2) \cdot (\mathbf{N}_1^{-1}\mathbf{m}_1)\}] \\ \times Z\{S_g^{(1)}(\mathbf{m}_1), \mathbf{h}_2\}.$$

This formula allows the transposition of intermediate results Z from an indexing by

$$(\text{all } \mathbf{m}_1) \times (\text{unique } \mathbf{h}_2)$$

to an indexing by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{h}_2).$$

We may then apply the twiddle factors to obtain

$$Y^*(\mathbf{m}_1, \mathbf{h}_2) = e[-\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] Z(\mathbf{m}_1, \mathbf{h}_2)$$

and carry out the second transform on \mathbf{h}_2

$$Y(\mathbf{m}_1, \mathbf{m}_2) = \sum_{\mathbf{h}_2 \in \mathbb{Z}^3/\mathbf{N}_2^T \mathbb{Z}^3} Y^*(\mathbf{m}_1, \mathbf{h}_2) e[-\mathbf{h}_2 \cdot (\mathbf{N}_2^{-1}\mathbf{m}_2)].$$

The results, indexed by

$$(\text{unique } \mathbf{m}_1) \times (\text{all } \mathbf{m}_2)$$

yield essentially an asymmetric unit of electron densities by the rearrangement

$$\rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) = Y(\mathbf{m}_1, \mathbf{m}_2).$$

1. GENERAL RELATIONSHIPS AND TECHNIQUES

The equivalence of the two transposition formulae up to the intervening twiddle factors is readily established, using the relation

$$\mathbf{h}_2 \cdot [\mathbf{N}_2^{-1} \boldsymbol{\mu}_2(g, \mathbf{m}_1)] = \boldsymbol{\eta}_1(g, \mathbf{h}_2) \cdot (\mathbf{N}_1^{-1} \mathbf{m}_1) \bmod 1$$

which is itself a straightforward consequence of the identity

$$\mathbf{h} \cdot [\mathbf{N}^{-1} S_g(\mathbf{m})] = \mathbf{h} \cdot \mathbf{t}_g + (\mathbf{R}_g^T \mathbf{h}) \cdot (\mathbf{N}^{-1} \mathbf{m}).$$

To complete the characterization of the effect of symmetry on the Cooley–Tukey factorization, and of the economy of computation it allows, it remains to consider the possibility that some values of \mathbf{m}_1 may be invariant under some transformations $g \in G$ under the action $\mathbf{m}_1 \mapsto S_g^{(1)}(\mathbf{m}_1)$.

Suppose that \mathbf{m}_1 has a non-trivial isotropy subgroup $G_{\mathbf{m}_1}$, and let $g \in G_{\mathbf{m}_1}$. Then each subarray $Y_{\mathbf{m}_1}$ defined by

$$Y_{\mathbf{m}_1}(\mathbf{m}_2) = Y(\mathbf{m}_1, \mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1 \mathbf{m}_2)$$

satisfies the identity

$$\begin{aligned} Y_{\mathbf{m}_1}(\mathbf{m}_2) &= Y_{S_g^{(1)}(\mathbf{m}_1)} [S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1)] \\ &= Y_{\mathbf{m}_1} [S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1)] \end{aligned}$$

so that the data for the transform on \mathbf{m}_2 have residual symmetry properties. In this case the identity satisfied by $\boldsymbol{\mu}_2$ simplifies to

$$\boldsymbol{\mu}_2(gg', \mathbf{m}_1) = S_g^{(2)}[\boldsymbol{\mu}_2(g', \mathbf{m}_1)] + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \bmod \mathbf{N}_2 \mathbb{Z}^3,$$

which shows that the mapping $g \mapsto \boldsymbol{\mu}_2(g, \mathbf{m}_1)$ satisfies the Frobenius congruences (Section 1.3.4.2.2.3). Thus the internal symmetry of subarray $Y_{\mathbf{m}_1}$ with respect to the action of G on \mathbf{m}_2 is given by $G_{\mathbf{m}_1}$ acting on $\mathbb{Z}^3/\mathbf{N}_2 \mathbb{Z}^3$ via

$$\mathbf{m}_2 \mapsto S_g^{(2)}(\mathbf{m}_2) + \boldsymbol{\mu}_2(g, \mathbf{m}_1) \bmod \mathbf{N}_2 \mathbb{Z}^3.$$

The transform on \mathbf{m}_2 needs only be performed for one out of $[G : G_{\mathbf{m}_1}]$ distinct arrays $Y_{\mathbf{m}_1}$ (results for the others being obtainable by the transposition formula), and this transform is $G_{\mathbf{m}_1}$ -symmetric. In other words, the following cases occur:

- (i) $G_{\mathbf{m}_1} = \{e\}$ maximum saving in computation (by $|G|$);
 \mathbf{m}_2 -transform has no symmetry.
- (ii) $G_{\mathbf{m}_1} = G' < G$ saving in computation by a factor of $[G : G']$;
 \mathbf{m}_2 -transform is G' -symmetric.
- (iii) $G_{\mathbf{m}_1} = G$ no saving in computation;
 \mathbf{m}_2 -transform is G -symmetric.

The symmetry properties of the \mathbf{m}_2 -transform may themselves be exploited in a similar way if \mathbf{N}_2 can be factored as a product of smaller decimation matrices; otherwise, an appropriate symmetrized DFT routine may be provided, using for instance the idea of ‘multiplexing/demultiplexing’ (Section 1.3.4.3.5). We thus have a recursive *descent procedure*, in which the deeper stages of the recursion deal with transforms on *fewer points*, or of *lower symmetry* (usually both).

The same analysis applies to the \mathbf{h}_1 -transforms on the subarrays $Z_{\mathbf{h}_2}^*$, and leads to a similar descent procedure.

In conclusion, crystallographic symmetry can be fully exploited to reduce the amount of computation to the minimum required to obtain the unique results from the unique data. No such analysis was so far available in cases where the asymmetric units in real and reciprocal space are not parallelepipeds. An example of this procedure will be given in Section 1.3.4.3.6.5.

1.3.4.3.4.2. Multidimensional Good factorization

This procedure was described in Section 1.3.3.3.2.2. The main difference with the Cooley–Tukey factorization is that if $\mathbf{N} = \mathbf{N}_1 \mathbf{N}_2 \dots \mathbf{N}_{d-1} \mathbf{N}_d$, where the different factors are pairwise coprime, then the Chinese remainder theorem reindexing makes $\mathbb{Z}^3/\mathbf{N} \mathbb{Z}^3$ isomorphic to a direct sum.

$$\mathbb{Z}^3/\mathbf{N} \mathbb{Z}^3 \cong (\mathbb{Z}^3/\mathbf{N}_1 \mathbb{Z}^3) \oplus \dots \oplus (\mathbb{Z}^3/\mathbf{N}_d \mathbb{Z}^3),$$

where each p -primary piece is endowed with an induced $\mathbb{Z}G$ -module structure by letting G operate in the usual way but with the corresponding modular arithmetic. The situation is thus more favourable than with the Cooley–Tukey method, since there is no interference between the factors (no ‘carry’). In the terminology of Section 1.3.4.2.2.2, G acts *diagonally* on this direct sum, and results of a partial transform may be transposed by orbit exchange as in Section 1.3.4.3.4.1 but without the extra terms $\boldsymbol{\mu}$ or $\boldsymbol{\eta}$. The analysis of the symmetry properties of partial transforms also carries over, again without the extra terms. Further simplification occurs for all p -primary pieces with p other than 2 or 3, since all non-primitive translations (including those associated to lattice centring) disappear modulo p .

Thus the cost of the CRT reindexing is compensated by the computational savings due to the absence of twiddle factors and of other phase shifts associated with non-primitive translations and with geometric ‘carries’.

Within each p -primary piece, however, higher powers of p may need to be split up by a Cooley–Tukey factorization, or carried out directly by a suitably adapted Winograd algorithm.

1.3.4.3.4.3. Crystallographic extension of the Rader/Winograd factorization

As was the case in the absence of symmetry, the two previous classes of algorithms can only factor the global transform into partial transforms on prime numbers of points, but cannot break the latter down any further. Rader’s idea of using the action of the group of units $U(p)$ to obtain further factorization of a p -primary transform has been used in ‘scalar’ form by Auslander & Shenefelt (1987), Shenefelt (1988), and Auslander *et al.* (1988). It will be shown here that it can be adapted to the crystallographic case so as to take advantage also of the possible existence of n -fold cyclic symmetry elements ($n = 3, 4, 6$) in a two-dimensional transform (Bricogne & Tolimieri, 1990). This adaptation entails the use of certain rings of *algebraic* integers rather than ordinary integers, whose connection with the handling of cyclic symmetry will now be examined.

Let G be the group associated with a threefold axis of symmetry: $G = \{e, g, g^2\}$ with $g^3 = e$. In a standard trigonal basis, G has matrix representation

$$\mathbf{R}_e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \mathbf{R}_g = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{R}_{g^2} = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

in real space,

$$\mathbf{R}_e^* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}, \quad \mathbf{R}_g^* = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{R}_{g^2}^* = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$$

in reciprocal space. Note that

$$\mathbf{R}_{g^2}^* = [\mathbf{R}_{g^2}^{-1}]^T = \mathbf{R}_g^T,$$

and that

$$\mathbf{R}_g^T = \mathbf{J}^{-1} \mathbf{R}_g \mathbf{J}, \quad \text{where } \mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

so that \mathbf{R}_g and \mathbf{R}_g^T are conjugate in the group of 2×2 unimodular

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

integer matrices. The group ring $\mathbb{Z}G$ is commutative, and has the structure of the polynomial ring $\mathbb{Z}[X]$ with the single relation $X^2 + X + 1 = 0$ corresponding to the minimal polynomial of \mathbf{R}_g . In the terminology of Section 1.3.3.2.4, the ring structure of $\mathbb{Z}G$ is obtained from that of $\mathbb{Z}[X]$ by carrying out polynomial addition and multiplication modulo $X^2 + X + 1$, then replacing X by any generator of G . This type of construction forms the very basis of algebraic number theory [see Artin (1944, Section IIc) for an illustration of this viewpoint], and $\mathbb{Z}G$ as just defined is isomorphic to the ring $\mathbb{Z}[\omega]$ of algebraic integers of the form $a + b\omega$ [$a, b \in \mathbb{Z}, \omega = \exp(2\pi i/3)$] under the identification $X \leftrightarrow \omega$. Addition in this ring is defined component-wise, while multiplication is defined by

$$(a_1 + b_1\omega) \times (a_2 + b_2\omega) = (a_1a_2 - b_1b_2) + [(a_1 - b_1)b_2 + b_1a_2]\omega.$$

In the case of a fourfold axis, $G = \{e, g, g^2, g^3\}$ with $g^4 = e$, and

$$\mathbf{R}_g = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \mathbf{R}_g^*, \quad \text{with again } \mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}.$$

$\mathbb{Z}G$ is obtained from $\mathbb{Z}[X]$ by carrying out polynomial arithmetic modulo $X^2 + 1$. This identifies $\mathbb{Z}G$ with the ring $\mathbb{Z}[i]$ of Gaussian integers of the form $a + bi$, in which addition takes place component-wise while multiplication is defined by

$$(a_1 + b_1i) \times (a_2 + b_2i) = (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i.$$

In the case of a sixfold axis, $G = \{e, g, g^2, g^3, g^4, g^5\}$ with $g^6 = e$, and

$$\mathbf{R}_g = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{R}_g^* = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{R}_g^T = \mathbf{J}^{-1}\mathbf{R}_g\mathbf{J}.$$

$\mathbb{Z}G$ is isomorphic to $\mathbb{Z}[\omega]$ under the mapping $g \leftrightarrow 1 + \omega$ since $(1 + \omega)^6 = 1$.

Thus in all cases $\mathbb{Z}G \cong \mathbb{Z}[X]/P(X)$ where $P(X)$ is an irreducible quadratic polynomial with integer coefficients.

The actions of G on lattices in real and reciprocal space (Sections 1.3.4.2.2.4, 1.3.4.2.2.5) extend naturally to actions of $\mathbb{Z}G$ on \mathbb{Z}^2 in which an element $z = a + b g$ of $\mathbb{Z}G$ acts via

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \mapsto z\mathbf{m} = (a\mathbf{I} + b\mathbf{R}_g) \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$$

in real space, and via

$$\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto z\mathbf{h} = (a\mathbf{I} + b\mathbf{R}_g^T) \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

in reciprocal space. These two actions are related by conjugation, since

$$(a\mathbf{I} + b\mathbf{R}_g^T) = \mathbf{J}^{-1}(a\mathbf{I} + b\mathbf{R}_g)\mathbf{J}$$

and the following identity (which is fundamental in the sequel) holds:

$$(z\mathbf{h}) \cdot \mathbf{m} = \mathbf{h} \cdot (z\mathbf{m}) \quad \text{for all } \mathbf{m}, \mathbf{h} \in \mathbb{Z}^2.$$

Let us now consider the calculation of a $p \times p$ two-dimensional DFT with n -fold cyclic symmetry ($n = 3, 4, 6$) for an odd prime $p \geq 5$. Denote $\mathbb{Z}/p\mathbb{Z}$ by \mathbb{Z}_p . Both the data and the results of the DFT are indexed by $\mathbb{Z}_p \times \mathbb{Z}_p$; hence the action of $\mathbb{Z}G$ on these indices is in fact an action of $\mathbb{Z}_p G$, the latter being obtained from $\mathbb{Z}G$ by carrying out all integer arithmetic in $\mathbb{Z}G$ modulo p . The algebraic structure of $\mathbb{Z}_p G$ combines the symmetry-carrying ring structure of $\mathbb{Z}G$ with the finite field structure of \mathbb{Z}_p used in Section 1.3.3.2.3.1, and holds the key to a symmetry-adapted factorization of the DFT at hand.

The structure of $\mathbb{Z}_p G$ depends on whether $P(X)$ remains irreducible when considered as a polynomial over \mathbb{Z}_p . Thus two cases arise:

- (1) $P(X)$ remains irreducible mod p , i.e. there is no n th root of unity in \mathbb{Z}_p ;
- (2) $P(X)$ factors as $(X - u)(X - v)$, i.e. there are n th roots of unity in \mathbb{Z}_p .

These two cases require different developments.

Case 1. $\mathbb{Z}_p G$ is a finite field with p^2 elements. There is essentially (i.e. up to isomorphism) only one such field, denoted $GF(p^2)$, and its group of units is a cyclic group with $p^2 - 1$ elements. If γ is a generator of this group of units, the input data $\rho_{\mathbf{m}}$ with $\mathbf{m} \neq \mathbf{0}$ may be reordered as

$$\mathbf{m}_0, \gamma\mathbf{m}_0, \gamma^2\mathbf{m}_0, \gamma^3\mathbf{m}_0, \dots, \gamma^{p^2-2}\mathbf{m}_0$$

by the *real-space action* of γ ; while the results $F_{\mathbf{h}}$ with $\mathbf{h} \neq \mathbf{0}$ may be reordered as

$$\mathbf{h}_0, \gamma\mathbf{h}_0, \gamma^2\mathbf{h}_0, \gamma^3\mathbf{h}_0, \dots, \gamma^{p^2-2}\mathbf{h}_0$$

by the *reciprocal-space action* of γ , where \mathbf{m}_0 and \mathbf{h}_0 are arbitrary non-zero indices.

The core $\mathbf{C}_{p \times p}$ of the DFT matrix, defined by

$$\mathbf{F}_{p \times p} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & & & \\ \vdots & & \mathbf{C}_{p \times p} & \\ 1 & & & \end{pmatrix},$$

will then have a skew-circulant structure (Section 1.3.3.2.3.1) since

$$(\mathbf{C}_{p \times p})_{jk} = e \left[\frac{(\gamma^j \mathbf{h}_0) \cdot (\gamma^k \mathbf{m}_0)}{p} \right] = e \left[\frac{\mathbf{h}_0 \cdot (\gamma^{j+k} \mathbf{m}_0)}{p} \right]$$

depends only on $j + k$. Multiplication by $\mathbf{C}_{p \times p}$ may then be turned into a cyclic convolution of length $p^2 - 1$, which may be factored by two DFTs (Section 1.3.3.2.3.1) or by Winograd's techniques (Section 1.3.3.2.4). The latter factorization is always favourable, as it is easily shown that $p^2 - 1$ is divisible by 24 for any odd prime $p \geq 5$. This procedure is applicable even if no symmetry is present in the data.

Assume now that cyclic symmetry of order $n = 3, 4$ or 6 is present. Since n divides 24 hence divides $p^2 - 1$, the generator g of this symmetry is representable as $\gamma^{(p^2-1)/n}$ for a suitable generator γ of the group of units. The reordered data will then be $(p^2 - 1)/n$ -periodic rather than simply $(p^2 - 1)$ -periodic; hence the reindexed results will be n -decimated (Section 1.3.2.7.2), and the $(p^2 - 1)/n$ non-zero results can be calculated by applying the DFT to the $(p^2 - 1)/n$ unique input data. In this way, the n -fold symmetry can be used in full to calculate the core contributions from the unique data to the unique results by a DFT of length $(p^2 - 1)/n$.

It is a simple matter to incorporate non-primitive translations into this scheme. For example, when going from structure factors to electron densities, reordered data items separated by $(p^2 - 1)/n$ are not equal but differ by a phase shift proportional to their index mod p , whose effect is simply to shift the origin of the n -decimated transformed sequence. The same economy of computation can therefore be achieved as in the purely cyclic case.

Dihedral symmetry elements, which map g to g^{-1} (Section 1.3.4.2.2.3), induce extra one-dimensional symmetries of order 2 in the reordered data which can also be fully exploited to reduce computation.

Case 2. If $p \geq 5$, it can be shown that the two roots u and v are always distinct. Then, by the Chinese remainder theorem (CRT) for polynomials (Section 1.3.3.2.4) we have a ring isomorphism

$$\mathbb{Z}_p[X]/P(X) \cong \{\mathbb{Z}_p[X]/(X - u)\} \times \{\mathbb{Z}_p[X]/(X - v)\}$$

1. GENERAL RELATIONSHIPS AND TECHNIQUES

defined by sending a polynomial $Q(X)$ from the left-hand-side ring to its two residue classes modulo $X - u$ and $X - v$, respectively. Since the latter are simply the constants $Q(u)$ and $Q(v)$, the CRT reindexing has the particularly simple form

$$a + bX \mapsto (a + bu, a + bv) = (\alpha, \beta)$$

or equivalently

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{M} \begin{pmatrix} a \\ b \end{pmatrix} \pmod{p}, \quad \text{with } \mathbf{M} = \begin{pmatrix} 1 & u \\ 1 & v \end{pmatrix}.$$

The CRT reconstruction formula similarly simplifies to

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mapsto \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{M}^{-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \pmod{p},$$

$$\text{with } \mathbf{M}^{-1} = \frac{1}{v - u} \begin{pmatrix} v & -u \\ -1 & 1 \end{pmatrix}.$$

The use of the CRT therefore amounts to the *simultaneous diagonalization* (by \mathbf{M}) of all the matrices representing the elements of $\mathbb{Z}_p G$ in the basis $(1, X)$.

A first consequence of this diagonalization is that the internal structure of $\mathbb{Z}_p G$ becomes clearly visible. Indeed, $\mathbb{Z}_p G$ is mapped isomorphically to a direct product of two copies of \mathbb{Z}_p , in which arithmetic is carried out *component-wise* between eigenvalues α and β . Thus if

$$z = a + bX \xleftrightarrow{\text{CRT}} (\alpha, \beta),$$

$$z' = a' + b'X \xleftrightarrow{\text{CRT}} (\alpha', \beta'),$$

then

$$z + z' \xleftrightarrow{\text{CRT}} (\alpha + \alpha', \beta + \beta'),$$

$$zz' \xleftrightarrow{\text{CRT}} (\alpha\alpha', \beta\beta').$$

Taking in particular

$$z \xleftrightarrow{\text{CRT}} (\alpha, 0) \neq (0, 0),$$

$$z' \xleftrightarrow{\text{CRT}} (0, \beta) \neq (0, 0),$$

we have $zz' = 0$, so that $\mathbb{Z}_p G$ contains zero divisors; therefore $\mathbb{Z}_p G$ is not a field. On the other hand, if $z \xleftrightarrow{\text{CRT}} (\alpha, \beta)$ with $\alpha \neq 0$ and $\beta \neq 0$, then α and β belong to the group of units $U(p)$ (Section 1.3.3.2.3.1) and hence have inverses α^{-1} and β^{-1} ; it follows that z is a unit in $\mathbb{Z}_p G$, with inverse $z^{-1} \xleftrightarrow{\text{CRT}} (\alpha^{-1}, \beta^{-1})$. Therefore, $\mathbb{Z}_p G$ consists of four distinct pieces:

$$0 \xleftrightarrow{\text{CRT}} \{(0, 0)\},$$

$$D_1 \xleftrightarrow{\text{CRT}} \{(\alpha, 0) | \alpha \in U(p)\} \cong U(p),$$

$$D_2 \xleftrightarrow{\text{CRT}} \{(0, \beta) | \beta \in U(p)\} \cong U(p),$$

$$U \xleftrightarrow{\text{CRT}} \{(\alpha, \beta) | \alpha \in U(p), \beta \in U(p)\} \cong U(p) \times U(p).$$

A second consequence of this diagonalization is that the actions of $\mathbb{Z}_p G$ on indices \mathbf{m} and \mathbf{h} can themselves be brought to diagonal form by basis changes:

$$\mathbf{m} \mapsto (a\mathbf{I} + b\mathbf{R}_g)\mathbf{m}$$

$$\text{becomes } \boldsymbol{\mu} \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \boldsymbol{\mu} \quad \text{with } \boldsymbol{\mu} = \mathbf{M}\mathbf{m},$$

$$\mathbf{h} \mapsto (a\mathbf{I} + b\mathbf{R}_g^T)\mathbf{h}$$

$$\text{becomes } \boldsymbol{\eta} \mapsto \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \boldsymbol{\eta} \quad \text{with } \boldsymbol{\eta} = \mathbf{M}\mathbf{J}\mathbf{h}.$$

Thus the sets of indices $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ can be split into four pieces as $\mathbb{Z}_p G$ itself, according as these indices have none, one or two of their coordinates in $U(p)$. These pieces will be labelled by the same symbols – 0, D_1 , D_2 and U – as those of $\mathbb{Z}_p G$.

The scalar product $\mathbf{h} \cdot \mathbf{m}$ may be written in terms of $\boldsymbol{\eta}$ and $\boldsymbol{\mu}$ as

$$\mathbf{h} \cdot \mathbf{m} = [\boldsymbol{\eta} \cdot ((\mathbf{M}^{-1})^T \mathbf{J} \mathbf{M}^{-1}) \boldsymbol{\mu}],$$

and an elementary calculation shows that the matrix $= (\mathbf{M}^{-1})^T \mathbf{J} \mathbf{M}^{-1}$ is *diagonal* by virtue of the relation

$$uv = \text{constant term in } P(X) = 1.$$

Therefore, $\mathbf{h} \cdot \mathbf{m} = 0$ if $\mathbf{h} \in D_1$ and $\boldsymbol{\mu} \in D_2$ or *vice versa*.

We are now in a position to rearrange the DFT matrix $\mathbf{F}_{p \times p}$. Clearly, the structure of $\mathbf{F}_{p \times p}$ is more complex than in case 1, as there are three types of ‘core’ matrices:

- type 1: $D \times D$ (with $D = D_1$ or D_2);
- type 2: $D \times U$ or $U \times D$;
- type 3: $U \times U$.

(Submatrices of type $D_1 \times D_2$ and $D_2 \times D_1$ have all their elements equal to 1 by the previous remark.)

Let γ be a generator of $U(p)$. We may reorder the elements in D_1 , D_2 and U – and hence the data and results indexed by these elements – according to powers of γ . This requires one exponent in each of D_1 and D_2 , and two exponents in U . For instance, in the \mathbf{h} -index space:

$$D_1 = \left\{ \begin{pmatrix} \gamma & 0 \\ 0 & 0 \end{pmatrix}^j \begin{pmatrix} \eta_1 \\ 0 \end{pmatrix} \Big| j = 1, \dots, p-1 \right\}$$

$$D_2 = \left\{ \begin{pmatrix} 0 & 0 \\ 0 & \gamma \end{pmatrix}^j \begin{pmatrix} 0 \\ \eta_2 \end{pmatrix} \Big| j = 1, \dots, p-1 \right\}$$

$$U = \left\{ \begin{pmatrix} \gamma & 0 \\ 0 & 1 \end{pmatrix}^{j_1} \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}^{j_2} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \Big| j_1 = 1, \dots, p-1; \right. \\ \left. j_2 = 1, \dots, p-1 \right\}$$

and similarly for the $\boldsymbol{\mu}$ index.

Since the diagonal matrix $\mathbf{\Delta}$ commutes with all the matrices representing the action of γ , this rearrangement will induce skew-circulant structures in all the core matrices. The corresponding cyclic convolutions may be carried out by Rader’s method, *i.e.* by diagonalizing them by means of two $(p-1)$ -point one-dimensional DFTs in the $D \times D$ pieces and of two $(p-1) \times (p-1)$ -point two-dimensional DFTs in the $U \times U$ piece (the $U \times D$ and $D \times U$ pieces involve extra section and projection operations).

In the absence of symmetry, no computational saving is achieved, since the same reordering could have been applied to the initial $\mathbb{Z}_p \times \mathbb{Z}_p$ indexing, without the CRT reindexing.

In the presence of n -fold cyclic symmetry, however, the rearranged $\mathbf{F}_{p \times p}$ lends itself to an n -fold reduction in size. The basic fact is that whenever case 2 occurs, $p-1$ is divisible by n (*i.e.* $p-1$ is divisible by 6 when $n=3$ or 6, and by 4 when $n=4$), say

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

$p - 1 = nq$. If g is a generator of the cyclic symmetry, the generator γ of $U(p)$ may be chosen in such a way that $g = \gamma^q$. The action of g is then to increment the j index in D_1 and D_2 by q , and the (j_1, j_2) index in U by (q, q) . Since the data items whose indices are related in this way have identical values, the DFTs used to diagonalize the Rader cyclic convolutions will operate on *periodized data*, hence yield *decimated results*; and the non-zero results will be obtained from the unique data by DFTs n times smaller than their counterparts in the absence of symmetry.

A more thorough analysis is needed to obtain a Winograd factorization into the normal from CBA in the presence of symmetry (see Bricogne & Tolimieri, 1990).

Non-primitive translations and dihedral symmetry may also be accommodated within this framework, as in case 1.

This reindexing by means of algebraic integers yields larger orbits, hence more efficient algorithms, than that of Auslander *et al.* (1988) which only uses ordinary integers acting by scalar dilation.

1.3.4.3.5. Treatment of conjugate and parity-related symmetry properties

Most crystallographic Fourier syntheses are real-valued and originate from Hermitian-symmetric collections of Fourier coefficients. Hermitian symmetry is closely related to the action of a centre of inversion in reciprocal space, and thus interacts strongly with all other genuinely crystallographic symmetry elements of order 2. All these symmetry properties are best treated by factoring by 2 and reducing the computation of the initial transform to that of a collection of smaller transforms with less symmetry or none at all.

1.3.4.3.5.1. Hermitian-symmetric or real-valued transforms

The computation of a DFT with Hermitian-symmetric or real-valued data can be carried out at a cost of half that of an ordinary transform, essentially by ‘multiplexing’ pairs of special partial transforms into general complex transforms, and then ‘demultiplexing’ the results on the basis of their symmetry properties. The treatment given below is for general dimension n ; a subset of cases for $n = 1$ was treated by Ten Eyck (1973).

(a) Underlying group action

Hermitian symmetry is not a geometric symmetry, but it is defined in terms of the action in reciprocal space of point group $G = \bar{1}$, *i.e.* $G = \{e, -e\}$, where e acts as \mathbf{I} (the $n \times n$ identity matrix) and $-e$ acts as $-\mathbf{I}$.

This group action on $\mathbb{Z}^n / N\mathbb{Z}^n$ with $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2$ will now be characterized by the calculation of the cocycle $\boldsymbol{\eta}_1$ (Section 1.3.4.3.4.1) under the assumption that \mathbf{N}_1 and \mathbf{N}_2 are both *diagonal*. For this purpose it is convenient to associate to any integer vector

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \text{ in } \mathbb{Z}^n \text{ the vector } \boldsymbol{\zeta}(\mathbf{v}) \text{ whose } j\text{th component is } \begin{cases} 0 & \text{if } v_j = 0 \\ 1 & \text{if } v_j \neq 0. \end{cases}$$

Let $\mathbf{m} = \mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2$, and hence $\mathbf{h} = \mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1$. Then

$$\begin{aligned} -\mathbf{h}_2 \bmod N\mathbb{Z}^n &= \mathbf{N}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2, \\ -\mathbf{h}_2 \bmod N_2\mathbb{Z}^n &= \mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2, \end{aligned}$$

hence

$$\begin{aligned} \boldsymbol{\eta}_1(-e, \mathbf{h}_2) &= \mathbf{N}_2^{-1} \{ [\mathbf{N}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2] - [\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2] \} \bmod \mathbf{N}_1\mathbb{Z}^n \\ &= -\boldsymbol{\zeta}(\mathbf{h}_2) \bmod \mathbf{N}_1\mathbb{Z}^n. \end{aligned}$$

Therefore $-e$ acts by

$$(\mathbf{h}_2, \mathbf{h}_1) \mapsto [\mathbf{N}_2\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2, \mathbf{N}_1\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1 - \boldsymbol{\zeta}(\mathbf{h}_2)].$$

Hermitian symmetry is traditionally dealt with by factoring by 2, *i.e.* by assuming $\mathbf{N} = 2\mathbf{M}$. If $\mathbf{N}_2 = 2\mathbf{I}$, then each \mathbf{h}_2 is invariant under G , so that each partial vector $\mathbf{Z}_{\mathbf{h}_2}^*$ (Section 1.3.4.3.4.1) inherits the symmetry internally, with a ‘modulation’ by $\boldsymbol{\eta}_1(g, \mathbf{h}_2)$. The ‘multiplexing–demultiplexing’ technique provides an efficient treatment of this singular case.

(b) Calculation of structure factors

The computation may be summarized as follows:

$$\rho \xrightarrow{\text{dec}(\mathbf{N}_1)} \mathbf{Y} \xrightarrow{\bar{F}(\mathbf{N}_2)} \mathbf{Y}^* \xrightarrow{\text{TW}} \mathbf{Z} \xrightarrow{\bar{F}(\mathbf{N}_1)} \mathbf{Z}^* \xrightarrow{\text{rev}(\mathbf{N}_2)} \mathbf{F}$$

where $\text{dec}(\mathbf{N}_1)$ is the initial decimation given by $\mathbf{Y}_{\mathbf{m}_1}(\mathbf{m}_2) = \rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2)$, TW is the transposition and twiddle-factor stage, and $\text{rev}(\mathbf{N}_2)$ is the final unscrambling by coset reversal given by $F(\mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1) = \mathbf{Z}_{\mathbf{h}_2}^*(\mathbf{h}_1)$.

(i) Decimation in time ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

The decimated vectors $\mathbf{Y}_{\mathbf{m}_1}$ are real and hence have Hermitian transforms $\mathbf{Y}_{\mathbf{m}_1}^*$. The 2^n values of \mathbf{m}_1 may be grouped into 2^{n-1} pairs $(\mathbf{m}'_1, \mathbf{m}''_1)$ and the vectors corresponding to each pair may be multiplexed into a general complex vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + i\mathbf{Y}_{\mathbf{m}''_1}.$$

The transform $\mathbf{Y}^* = \bar{F}(\mathbf{M})[\mathbf{Y}]$ can then be resolved into the separate transforms $\mathbf{Y}_{\mathbf{m}'_1}^*$ and $\mathbf{Y}_{\mathbf{m}''_1}^*$ by using the Hermitian symmetry of the latter, which yields the demultiplexing formulae

$$\begin{aligned} Y_{\mathbf{m}'_1}^*(\mathbf{h}_2) + iY_{\mathbf{m}''_1}^*(\mathbf{h}_2) &= Y^*(\mathbf{h}_2) \\ \overline{Y_{\mathbf{m}'_1}^*(\mathbf{h}_2) + iY_{\mathbf{m}''_1}^*(\mathbf{h}_2)} &= Y^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_2) - \mathbf{h}_2]. \end{aligned}$$

The number of partial transforms $\bar{F}(\mathbf{M})$ is thus reduced from 2^n to 2^{n-1} . Once this separation has been achieved, the remaining steps need only be carried out for a unique half of the values of \mathbf{h}_2 .

(ii) Decimation in frequency ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

Since $\mathbf{h}_2 \in \mathbb{Z}^n / 2\mathbb{Z}^n$ we have $-\mathbf{h}_2 = \mathbf{h}_2 + \boldsymbol{\zeta}(\mathbf{h}_2) = \mathbf{h}_2 \bmod 2\mathbb{Z}^n$. The vectors of decimated and scrambled results $\mathbf{Z}_{\mathbf{h}_2}^*$ then obey the symmetry relations

$$\mathbf{Z}_{\mathbf{h}_2}^*(\mathbf{h}_1 - \mathbf{h}_2) = \overline{\mathbf{Z}_{\mathbf{h}_2}^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1]}$$

which can be used to halve the number of $\bar{F}(\mathbf{M})$ necessary to compute them, as follows.

Having formed the vectors $\mathbf{Z}_{\mathbf{h}_2}$ given by

$$\mathbf{Z}_{\mathbf{h}_2}(\mathbf{m}_1) = \left[\sum_{\mathbf{m}_2 \in \mathbb{Z}^n / 2\mathbb{Z}^n} \frac{(-1)^{\mathbf{h}_2 \cdot \mathbf{m}_2}}{2^n} \rho(\mathbf{m}_1 + \mathbf{M}\mathbf{m}_2) \right] e[\mathbf{h}_2 \cdot (\mathbf{N}^{-1}\mathbf{m}_1)],$$

we may group the 2^n values of \mathbf{h}_2 into 2^{n-1} pairs $(\mathbf{h}'_2, \mathbf{h}''_2)$ and for each pair form the multiplexed vector:

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}'_2} + i\mathbf{Z}_{\mathbf{h}''_2}.$$

After calculating the 2^{n-1} transforms $\mathbf{Z}^* = \bar{F}(\mathbf{M})[\mathbf{Z}]$, the 2^n individual transforms $\mathbf{Z}_{\mathbf{h}'_2}^*$ and $\mathbf{Z}_{\mathbf{h}''_2}^*$ can be separated by using for each pair the demultiplexing formulae

$$\begin{aligned} \mathbf{Z}_{\mathbf{h}'_2}^*(\mathbf{h}_1) + i\mathbf{Z}_{\mathbf{h}''_2}^*(\mathbf{h}_1) &= \mathbf{Z}^*(\mathbf{h}_1) \\ \mathbf{Z}_{\mathbf{h}'_2}^*(\mathbf{h}_1 - \mathbf{h}'_2) + i\mathbf{Z}_{\mathbf{h}''_2}^*(\mathbf{h}_1 - \mathbf{h}''_2) &= \overline{\mathbf{Z}^*[\mathbf{M}\boldsymbol{\zeta}(\mathbf{h}_1) - \mathbf{h}_1]} \end{aligned}$$

which can be solved recursively. If all pairs are chosen so that they differ only in the j th coordinate $(\mathbf{h}_2)_j$, the recursion is along $(\mathbf{h}_1)_j$ and can be initiated by introducing the (real) values of $\mathbf{Z}_{\mathbf{h}'_2}^*$ and $\mathbf{Z}_{\mathbf{h}''_2}^*$ at $(\mathbf{h}_1)_j = 0$ and $(\mathbf{h}_1)_j = M_j$, accumulated *e.g.* while forming \mathbf{Z} for that pair. Only points with $(\mathbf{h}_1)_j$ going from 0 to $\frac{1}{2}M_j$ need be resolved,

1. GENERAL RELATIONSHIPS AND TECHNIQUES

and they contain the unique half of the Hermitian-symmetric transform \mathbf{F} .

(c) Calculation of electron densities

The computation may be summarized as follows:

$$\mathbf{F} \xrightarrow{\text{scr}(\mathbf{N}_2)} \mathbf{Z}^* \xrightarrow{F(\mathbf{N}_1)} \mathbf{Z} \xrightarrow{\text{TW}} \mathbf{Y}^* \xrightarrow{F(\mathbf{N}_2)} \mathbf{Y} \xrightarrow{\text{nat}(\mathbf{N}_1)} \rho$$

where $\text{scr}(\mathbf{N}_2)$ is the decimation with coset reversal given by $\mathbf{Z}_{\mathbf{h}_2}^*(\mathbf{h}_1) = F(\mathbf{h}_2 + \mathbf{N}_2\mathbf{h}_1)$, TW is the transposition and twiddle-factor stage, and $\text{nat}(\mathbf{N}_1)$ is the recovery in natural order given by $\rho(\mathbf{m}_1 + \mathbf{N}_1\mathbf{m}_2) = Y_{\mathbf{m}_1}(\mathbf{m}_2)$.

(i) Decimation in time ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

The last transformation $F(2\mathbf{I})$ has a real-valued matrix, and the final result ρ is real-valued. It follows that the vectors $\mathbf{Y}_{\mathbf{m}_1}^*$ of intermediate results after the twiddle-factor stage are real-valued, hence lend themselves to multiplexing along the real and imaginary components of half as many general complex vectors.

Let the 2^n initial vectors $\mathbf{Z}_{\mathbf{h}_2}^*$ be multiplexed into 2^{n-1} vectors

$$\mathbf{Z}^* = \mathbf{Z}_{\mathbf{h}_2'}^* + i\mathbf{Z}_{\mathbf{h}_2''}^*$$

[one for each pair $(\mathbf{h}_2', \mathbf{h}_2'')$], each of which yields by $F(\mathbf{M})$ a vector

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}_2'} + i\mathbf{Z}_{\mathbf{h}_2''}.$$

The real-valuedness of the $\mathbf{Y}_{\mathbf{m}_1}^*$ may be used to recover the separate result vectors for \mathbf{h}_2' and \mathbf{h}_2'' . For this purpose, introduce the abbreviated notation

$$\begin{aligned} e[-\mathbf{h}_2' \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] &= (c' + is')(\mathbf{m}_1) \\ e[-\mathbf{h}_2'' \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] &= (c'' + is'')(\mathbf{m}_1) \\ R_{\mathbf{h}_2}(\mathbf{m}_1) &= Y_{\mathbf{m}_1}^*(\mathbf{h}_2) \\ \mathbf{R}' = \mathbf{R}_{\mathbf{h}_2'}, \quad \mathbf{R}'' = \mathbf{R}_{\mathbf{h}_2''}. \end{aligned}$$

Then we may write

$$\begin{aligned} \mathbf{Z} &= (c' + is')\mathbf{R}' + i(c'' + is'')\mathbf{R}'' \\ &= (c'\mathbf{R}' + s''\mathbf{R}'') + i(-s'\mathbf{R}' + c''\mathbf{R}'') \end{aligned}$$

or, equivalently, for each \mathbf{m}_1 ,

$$\begin{pmatrix} \mathcal{R}e \mathbf{Z} \\ \mathcal{I}m \mathbf{Z} \end{pmatrix} = \begin{pmatrix} c' & s'' \\ -s' & c'' \end{pmatrix} \begin{pmatrix} \mathbf{R}' \\ \mathbf{R}'' \end{pmatrix}.$$

Therefore \mathbf{R}' and \mathbf{R}'' may be retrieved from \mathbf{Z} by the 'demultiplexing' formula:

$$\begin{pmatrix} \mathbf{R}' \\ \mathbf{R}'' \end{pmatrix} = \frac{1}{c'c'' + s's''} \begin{pmatrix} c'' & -s'' \\ s' & c' \end{pmatrix} \begin{pmatrix} \mathcal{R}e \mathbf{Z} \\ \mathcal{I}m \mathbf{Z} \end{pmatrix}$$

which is valid at all points \mathbf{m}_1 where $c'c'' + s's'' \neq 0$, *i.e.* where

$$\cos[2\pi(\mathbf{h}_2' - \mathbf{h}_2'') \cdot (\mathbf{N}^{-1}\mathbf{m}_1)] \neq 0.$$

Demultiplexing fails when

$$(\mathbf{h}_2' - \mathbf{h}_2'') \cdot (\mathbf{N}^{-1}\mathbf{m}_1) = \frac{1}{2} \pmod{1}.$$

If the pairs $(\mathbf{h}_2', \mathbf{h}_2'')$ are chosen so that their members differ only in one coordinate (the j th, say), then the exceptional points are at $(\mathbf{m}_1)_j = \frac{1}{2}M_j$ and the missing transform values are easily obtained *e.g.* by accumulation while forming \mathbf{Z}^* .

The final stage of the calculation is then

$$\rho(\mathbf{m}_1 + \mathbf{M}\mathbf{m}_2) = \sum_{\mathbf{h}_2 \in \mathbf{Z}^n/2\mathbf{Z}^n} (-1)^{\mathbf{h}_2 \cdot \mathbf{m}_2} R_{\mathbf{h}_2}(\mathbf{m}_1).$$

(ii) Decimation in frequency ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

The last transformation $F(\mathbf{M})$ gives the real-valued results ρ , therefore the vectors $\mathbf{Y}_{\mathbf{m}_1}^*$ after the twiddle-factor stage each have Hermitian symmetry.

A first consequence is that the intermediate vectors $\mathbf{Z}_{\mathbf{h}_2}$ need only be computed for the unique half of the values of \mathbf{h}_2 , the other half being related by the Hermitian symmetry of $\mathbf{Y}_{\mathbf{m}_1}^*$.

A second consequence is that the 2^n vectors $\mathbf{Y}_{\mathbf{m}_1}^*$ may be condensed into 2^{n-1} general complex vectors

$$\mathbf{Y}^* = \mathbf{Y}_{\mathbf{m}_1'}^* + i\mathbf{Y}_{\mathbf{m}_1''}^*$$

[one for each pair $(\mathbf{m}_1', \mathbf{m}_1'')$] to which a general complex $F(\mathbf{M})$ may be applied to yield

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}_1'} + i\mathbf{Y}_{\mathbf{m}_1''}$$

with $\mathbf{Y}_{\mathbf{m}_1'}$ and $\mathbf{Y}_{\mathbf{m}_1''}$ *real-valued*. The final results can therefore be retrieved by the particularly simple demultiplexing formulae:

$$\begin{aligned} \rho(\mathbf{m}_1' + 2\mathbf{m}_2) &= \mathcal{R}e Y(\mathbf{m}_2), \\ \rho(\mathbf{m}_1'' + 2\mathbf{m}_2) &= \mathcal{I}m Y(\mathbf{m}_2). \end{aligned}$$

1.3.4.3.5.2. Hermitian-antisymmetric or pure imaginary transforms

A vector $\mathbf{X} = \{X(\mathbf{k}) | \mathbf{k} \in \mathbb{Z}^n/\mathbf{N}\mathbb{Z}^n\}$ is said to be Hermitian-antisymmetric if

$$X(\mathbf{k}) = -\overline{X(-\mathbf{k})} \text{ for all } \mathbf{k}.$$

Its transform \mathbf{X}^* then satisfies

$$X^*(\mathbf{k}^*) = -\overline{X^*(\mathbf{k}^*)} \text{ for all } \mathbf{k}^*,$$

i.e. is purely imaginary.

If \mathbf{X} is Hermitian-antisymmetric, then $\mathbf{F} = \pm i\mathbf{X}$ is Hermitian-symmetric, with $\rho = \pm i\mathbf{X}^*$ real-valued. The treatment of Section 1.3.4.3.5.1 may therefore be adapted, with trivial factors of i or -1 , or used as such in conjunction with changes of variable by multiplication by $\pm i$.

1.3.4.3.5.3. Complex symmetric and antisymmetric transforms

The matrix $-\mathbf{I}$ is its own contragredient, and hence (Section 1.3.2.4.2.2) the transform of a symmetric (respectively antisymmetric) function is symmetric (respectively antisymmetric). In this case the group $G = \{e, -e\}$ acts in both real and reciprocal space as $\{\mathbf{I}, -\mathbf{I}\}$. If $\mathbf{N} = \mathbf{N}_1\mathbf{N}_2$ with both factors diagonal, then $-e$ acts by

$$\begin{aligned} (\mathbf{m}_1, \mathbf{m}_2) &\mapsto [\mathbf{N}_1\zeta(\mathbf{m}_1) - \mathbf{m}_1, \mathbf{N}_2\zeta(\mathbf{m}_2) - \mathbf{m}_2 - \zeta(\mathbf{m}_1)], \\ (\mathbf{h}_2, \mathbf{h}_1) &\mapsto [\mathbf{N}_2\zeta(\mathbf{h}_2) - \mathbf{h}_2, \mathbf{N}_1\zeta(\mathbf{h}_1) - \mathbf{h}_1 - \zeta(\mathbf{h}_2)], \end{aligned}$$

i.e.

$$\begin{aligned} \boldsymbol{\mu}_2(-e, \mathbf{m}_1) &= -\zeta(\mathbf{m}_1) \pmod{\mathbf{N}_2\mathbb{Z}^n}, \\ \boldsymbol{\eta}_1(-e, \mathbf{h}_2) &= -\zeta(\mathbf{h}_2) \pmod{\mathbf{N}_1\mathbb{Z}^n}. \end{aligned}$$

The symmetry or antisymmetry properties of \mathbf{X} may be written

$$X(-\mathbf{m}) = -\varepsilon X(\mathbf{m}) \text{ for all } \mathbf{m},$$

with $\varepsilon = +1$ for symmetry and $\varepsilon = -1$ for antisymmetry.

The computation will be summarized as

$$\mathbf{X} \xrightarrow{\text{dec}(\mathbf{N}_1)} \mathbf{Y} \xrightarrow{\bar{F}(\mathbf{N}_2)} \mathbf{Y}^* \xrightarrow{\text{TW}} \mathbf{Z} \xrightarrow{\bar{F}(\mathbf{N}_1)} \mathbf{Z}^* \xrightarrow{\text{rev}(\mathbf{N}_2)} \mathbf{X}^*$$

with the same indexing as that used for structure-factor calculation. In both cases it will be shown that a transform $F(\mathbf{N})$ with $\mathbf{N} = 2\mathbf{M}$ and \mathbf{M} diagonal can be computed using only 2^{n-1} partial transforms $F(\mathbf{M})$ instead of 2^n .

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

(i) *Decimation in time* ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

Since $\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{m}_1 = \mathbf{m}_1$ and $\zeta(\mathbf{m}_1) = \mathbf{m}_1 \bmod 2\mathbb{Z}^n$, so that the symmetry relations for each parity class of data $\mathbf{Y}_{\mathbf{m}_1}$ read

$$Y_{\mathbf{m}_1}[\mathbf{M}\zeta(\mathbf{m}_2) - \mathbf{m}_2 - \mathbf{m}_1] = \varepsilon Y_{\mathbf{m}_1}(\mathbf{m}_2)$$

or equivalently

$$\tau_{\mathbf{m}_1} \mathbf{Y}_{\mathbf{m}_1} = \varepsilon \check{\mathbf{Y}}_{\mathbf{m}_1}.$$

Transforming by $F(\mathbf{M})$, this relation becomes

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}_1)] \mathbf{Y}_{\mathbf{m}_1}^* = \varepsilon \mathbf{Y}_{\mathbf{m}_1}^*.$$

Each parity class thus obeys a different symmetry relation, so that we may multiplex them in pairs by forming for each pair $(\mathbf{m}'_1, \mathbf{m}''_1)$ the vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + \mathbf{Y}_{\mathbf{m}''_1}.$$

Putting

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}'_1)] = (c' + is')(\mathbf{h}_2)$$

$$e[-\mathbf{h}_2 \cdot (\mathbf{M}^{-1}\mathbf{m}''_1)] = (c'' + is'')(\mathbf{h}_2)$$

we then have the demultiplexing relations for each \mathbf{h}_2 :

$$\begin{aligned} Y_{\mathbf{m}'_1}^*(\mathbf{h}_2) + Y_{\mathbf{m}''_1}^*(\mathbf{h}_2) &= Y^*(\mathbf{h}_2) \\ (c' + is')(\mathbf{h}_2) Y_{\mathbf{m}'_1}^*(\mathbf{h}_2) + (c'' + is'')(\mathbf{h}_2) Y_{\mathbf{m}''_1}^*(\mathbf{h}_2) \\ &= \varepsilon Y^*[\mathbf{M}\zeta(\mathbf{h}_2) - \mathbf{h}_2] \end{aligned}$$

which can be solved recursively. Transform values at the exceptional points \mathbf{h}_2 where demultiplexing fails (*i.e.* where $c' + is' = c'' + is''$) can be accumulated while forming \mathbf{Y} .

Only the unique half of the values of \mathbf{h}_2 need to be considered at the demultiplexing stage and at the subsequent TW and $F(2\mathbf{I})$ stages.

(ii) *Decimation in frequency* ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

The vectors of final results $\mathbf{Z}_{\mathbf{h}_2}^*$ for each parity class \mathbf{h}_2 obey the symmetry relations

$$\tau_{\mathbf{h}_2} \mathbf{Z}_{\mathbf{h}_2}^* = \varepsilon \check{\mathbf{Z}}_{\mathbf{h}_2}^*,$$

which are different for each \mathbf{h}_2 . The vectors $\mathbf{Z}_{\mathbf{h}_2}$ of intermediate results after the twiddle-factor stage may then be multiplexed in pairs as

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}'_2} + \mathbf{Z}_{\mathbf{h}''_2}.$$

After transforming by $F(\mathbf{M})$, the results \mathbf{Z}^* may be demultiplexed by using the relations

$$\begin{aligned} Z_{\mathbf{h}'_2}^*(\mathbf{h}_1) + Z_{\mathbf{h}''_2}^*(\mathbf{h}_1) &= Z^*(\mathbf{h}_1) \\ Z_{\mathbf{h}'_2}^*(\mathbf{h}_1 - \mathbf{h}'_2) + Z_{\mathbf{h}''_2}^*(\mathbf{h}_1 - \mathbf{h}''_2) &= \varepsilon Z^*[\mathbf{M}\zeta(\mathbf{h}_1) - \mathbf{h}_1] \end{aligned}$$

which can be solved recursively as in Section 1.3.4.3.5.1(b)(ii).

1.3.4.3.5.4. Real symmetric transforms

Conjugate symmetric (Section 1.3.2.4.2.3) implies that if the data \mathbf{X} are real and symmetric [*i.e.* $X(\mathbf{k}) = \bar{X}(\mathbf{k})$ and $X(-\mathbf{k}) = X(\mathbf{k})$], then so are the results \mathbf{X}^* . Thus if ρ contains a centre of symmetry, \mathbf{F} is real symmetric. There is no distinction (other than notation) between structure-factor and electron-density calculation; the algorithms will be described in terms of the former. It will be shown that if $\mathbf{N} = 2\mathbf{M}$, a real symmetric transform can be computed with only 2^{n-2} partial transforms $F(\mathbf{M})$ instead of 2^n .

(i) *Decimation in time* ($\mathbf{N}_1 = 2\mathbf{I}, \mathbf{N}_2 = \mathbf{M}$)

Since $\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n$ we have $-\mathbf{m}_1 = \mathbf{m}_1$ and $\zeta(\mathbf{m}_1) = \mathbf{m}_1 \bmod 2\mathbb{Z}^n$. The decimated vectors $\mathbf{Y}_{\mathbf{m}_1}$ are not only real, but

have an internal symmetry expressed by

$$\mathbf{Y}_{\mathbf{m}_1}[\mathbf{M}\zeta(\mathbf{m}_2) - \mathbf{m}_2 - \mathbf{m}_1] = \varepsilon \mathbf{Y}_{\mathbf{m}_1}(\mathbf{m}_2).$$

This symmetry, however, is different for each \mathbf{m}_1 so that we may multiplex two such vectors $\mathbf{Y}_{\mathbf{m}'_1}$ and $\mathbf{Y}_{\mathbf{m}''_1}$ into a general *real* vector

$$\mathbf{Y} = \mathbf{Y}_{\mathbf{m}'_1} + \mathbf{Y}_{\mathbf{m}''_1},$$

for each of the 2^{n-1} pairs $(\mathbf{m}'_1, \mathbf{m}''_1)$. The 2^{n-1} Hermitian-symmetric transform vectors

$$\mathbf{Y}^* = \mathbf{Y}_{\mathbf{m}'_1}^* + \mathbf{Y}_{\mathbf{m}''_1}^*$$

can then be evaluated by the methods of Section 1.3.4.3.5.1(b) at the cost of only 2^{n-2} general complex $F(\mathbf{M})$.

The demultiplexing relations by which the separate vectors $\mathbf{Y}_{\mathbf{m}'_1}^*$ and $\mathbf{Y}_{\mathbf{m}''_1}^*$ may be recovered are most simply obtained by observing that the vectors \mathbf{Z} after the twiddle-factor stage are real-valued since $F(2\mathbf{I})$ has a real matrix. Thus, as in Section 1.3.4.3.5.1(c)(i),

$$\mathbf{Y}_{\mathbf{m}'_1}^* = (c' - is')\mathbf{R}'$$

$$\mathbf{Y}_{\mathbf{m}''_1}^* = (c'' - is'')\mathbf{R}'',$$

where \mathbf{R}' and \mathbf{R}'' are real vectors and where the multipliers $(c' - is')$ and $(c'' - is'')$ are the inverse twiddle factors. Therefore,

$$\begin{aligned} \mathbf{Y}^* &= (c' - is')\mathbf{R}' + (c'' - is'')\mathbf{R}'' \\ &= (c'\mathbf{R}' + c''\mathbf{R}'') - i(s'\mathbf{R}' + s''\mathbf{R}'') \end{aligned}$$

and hence the demultiplexing relation for each \mathbf{h}_2 :

$$\begin{pmatrix} R' \\ R'' \end{pmatrix} = \frac{1}{c's'' - s'c''} \begin{pmatrix} s'' & -c'' \\ -s' & c' \end{pmatrix} \begin{pmatrix} \text{Re } Y^* \\ -\text{Im } Y^* \end{pmatrix}.$$

The values of $R'_{\mathbf{h}_2}$ and $R''_{\mathbf{h}_2}$ at those points \mathbf{h}_2 where $c's'' - s'c'' = 0$ can be evaluated directly while forming \mathbf{Y} . This demultiplexing and the final stage of the calculation, namely

$$F(\mathbf{h}_2 + \mathbf{M}\mathbf{h}_1) = \frac{1}{2^n} \sum_{\mathbf{m}_1 \in \mathbb{Z}^n/2\mathbb{Z}^n} (-1)^{\mathbf{h}_1 \cdot \mathbf{m}_1} R_{\mathbf{m}_1}(\mathbf{h}_2)$$

need only be carried out for the unique half of the range of \mathbf{h}_2 .

(ii) *Decimation in frequency* ($\mathbf{N}_1 = \mathbf{M}, \mathbf{N}_2 = 2\mathbf{I}$)

Similarly, the vectors $\mathbf{Z}_{\mathbf{h}_2}^*$ of decimated and scrambled results are real and obey internal symmetries

$$\tau_{\mathbf{h}_2} \mathbf{Z}_{\mathbf{h}_2}^* = \varepsilon \check{\mathbf{Z}}_{\mathbf{h}_2}^*$$

which are different for each \mathbf{h}_2 . For each of the 2^{n-1} pairs $(\mathbf{h}'_2, \mathbf{h}''_2)$ the multiplexed vector

$$\mathbf{Z} = \mathbf{Z}_{\mathbf{h}'_2} + \mathbf{Z}_{\mathbf{h}''_2}$$

is a Hermitian-symmetric vector without internal symmetry, and the 2^{n-1} real vectors

$$\mathbf{Z}^* = \mathbf{Z}_{\mathbf{h}'_2}^* + \mathbf{Z}_{\mathbf{h}''_2}^*$$

may be evaluated at the cost of only 2^{n-2} general complex $F(\mathbf{M})$ by the methods of Section 1.3.4.3.5.1(c). The individual transforms $\mathbf{Z}_{\mathbf{h}'_2}$ and $\mathbf{Z}_{\mathbf{h}''_2}$ may then be retrieved *via* the demultiplexing relations

$$\begin{aligned} Z_{\mathbf{h}'_2}^*(\mathbf{h}_1) + Z_{\mathbf{h}''_2}^*(\mathbf{h}_1) &= Z^*(\mathbf{h}_1) \\ Z_{\mathbf{h}'_2}^*(\mathbf{h}_1 - \mathbf{h}'_2) + Z_{\mathbf{h}''_2}^*(\mathbf{h}_1 - \mathbf{h}''_2) &= Z^*[\mathbf{M}\zeta(\mathbf{h}_1) - \mathbf{h}_1] \end{aligned}$$

which can be solved recursively as described in Section 1.3.4.3.5.1(b)(ii). This yields the unique half of the real symmetric results \mathbf{F} .

1. GENERAL RELATIONSHIPS AND TECHNIQUES

1.3.4.3.5.5. Real antisymmetric transforms

If \mathbf{X} is real antisymmetric, then its transform \mathbf{X}^* is purely imaginary and antisymmetric. The double-multiplexing techniques used for real symmetric transforms may therefore be adapted with only minor changes involving signs and factors of i .

1.3.4.3.5.6. Generalized multiplexing

So far the multiplexing technique has been applied to pairs of vectors with similar types of parity-related and/or conjugate symmetry properties, in particular the same value of ε .

It can be generalized so as to accommodate mixtures of vectors with different symmetry characteristics. For example if \mathbf{X}_1 is Hermitian-symmetric and \mathbf{X}_2 is Hermitian-antisymmetric, so that \mathbf{X}_1^* is real-valued while \mathbf{X}_2^* has purely imaginary values, the multiplexing process should obviously form $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ (instead of $\mathbf{X} = \mathbf{X}_1 + i\mathbf{X}_2$ if both had the same type of symmetry), and demultiplexing consists in separating

$$\begin{aligned}\mathbf{X}_1^* &= \mathcal{R}e \mathbf{X}^* \\ \mathbf{X}_2^* &= i \mathcal{I}m \mathbf{X}^*.\end{aligned}$$

The general multiplexing formula for pairs of vectors may therefore be written

$$\mathbf{X} = \mathbf{X}_1 + \omega \mathbf{X}_2,$$

where ω is a phase factor (*e.g.* 1 or i) chosen in such a way that all non-exceptional components of \mathbf{X}_1 and \mathbf{X}_2 (or \mathbf{X}_1^* and \mathbf{X}_2^*) be embedded in the complex plane \mathbb{C} along linearly independent directions, thus making multiplexing possible.

It is possible to develop a more general form of multiplexing/demultiplexing for more than two vectors, which can be used to deal with symmetry elements of order 3, 4 or 6. It is based on the theory of group characters (Ledermann, 1987).

1.3.4.3.6. Global crystallographic algorithms

All the necessary ingredients are now available for calculating the CDFT for any given space group.

1.3.4.3.6.1. Triclinic groups

Space group $P1$ is dealt with by the methods of Section 1.3.4.3.5.1 and $P\bar{1}$ by those of Section 1.3.4.3.5.4.

1.3.4.3.6.2. Monoclinic groups

A general monoclinic transformation is of the form

$$S_g : \mathbf{x} \mapsto \mathbf{R}_g \mathbf{x} + \mathbf{t}_g$$

with \mathbf{R}_g a diagonal matrix whose entries are $+1$ or -1 , and \mathbf{t}_g a vector whose entries are 0 or $\frac{1}{2}$. We may thus decompose both real and reciprocal space into a direct sum of a subspace \mathbb{Z}^{n_+} where \mathbf{R}_g acts as the identity, and a subspace \mathbb{Z}^{n_-} where \mathbf{R}_g acts as minus the identity, with $n_+ + n_- = n = 3$. All usual entities may be correspondingly written as direct sums, for instance:

$$\begin{aligned}\mathbf{R}_g &= \mathbf{R}_g^+ \oplus \mathbf{R}_g^-, & \mathbf{N} &= \mathbf{N}^+ \oplus \mathbf{N}^-, & \mathbf{M} &= \mathbf{M}^+ \oplus \mathbf{M}^-, \\ \mathbf{t}_g &= \mathbf{t}_g^+ \oplus \mathbf{t}_g^-, & \mathbf{t}_g^{(1)} &= \mathbf{t}_g^{(1)+} \oplus \mathbf{t}_g^{(1)-}, & \mathbf{t}_g^{(2)} &= \mathbf{t}_g^{(2)+} \oplus \mathbf{t}_g^{(2)-}, \\ \mathbf{m} &= \mathbf{m}^+ \oplus \mathbf{m}^-, & \mathbf{m}_1 &= \mathbf{m}_1^+ \oplus \mathbf{m}_1^-, & \mathbf{m}_2 &= \mathbf{m}_2^+ \oplus \mathbf{m}_2^-, \\ \mathbf{h} &= \mathbf{h}^+ \oplus \mathbf{h}^-, & \mathbf{h}_1 &= \mathbf{h}_1^+ \oplus \mathbf{h}_1^-, & \mathbf{h}_2 &= \mathbf{h}_2^+ \oplus \mathbf{h}_2^-.\end{aligned}$$

We will use factoring by 2, with decimation in frequency when computing structure factors, and decimation in time when computing electron densities; this corresponds to $\mathbf{N} = \mathbf{N}_1 \mathbf{N}_2$ with $\mathbf{N}_1 = \mathbf{M}$, $\mathbf{N}_2 = 2\mathbf{I}$. The non-primitive translation vector \mathbf{Nt}_g then belongs to $\mathbf{M}\mathbb{Z}^n$, and thus

$$\mathbf{t}_g^{(1)} = \mathbf{0} \bmod \mathbf{M}\mathbb{Z}^n, \quad \mathbf{t}_g^{(2)} \in \mathbb{Z}^n / 2\mathbb{Z}^n.$$

The symmetry relations obeyed by ρ and F are as follows: for electron densities

$$\rho(\mathbf{m}^+, \mathbf{m}^-) = \rho(\mathbf{m}^+ + \mathbf{N}^+ \mathbf{t}_g^+, -\mathbf{m}^- - \mathbf{N}^- \mathbf{t}_g^-)$$

or, after factoring by 2,

$$\begin{aligned}\rho(\mathbf{m}_1^+, \mathbf{m}_2^+, \mathbf{m}_1^-, \mathbf{m}_2^-) \\ = \rho(\mathbf{m}_1^+, \mathbf{m}_2^+ + \mathbf{t}_g^{(2)+}, \mathbf{M}^- \zeta(\mathbf{m}_1^-) - \mathbf{m}_1^- - \mathbf{m}_2^-, \mathbf{m}_2^- + \mathbf{t}_g^{(2)-});\end{aligned}$$

while for structure factors

$$F(\mathbf{h}^+, \mathbf{h}^-) = \exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)] F(\mathbf{h}^+, -\mathbf{h}^-)$$

with its Friedel counterpart

$$F(\mathbf{h}^+, \mathbf{h}^-) = \exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)] \overline{F(-\mathbf{h}^+, \mathbf{h}^-)}$$

or, after factoring by 2,

$$\begin{aligned}F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-) = (-1)^{\mathbf{h}_2^+ \cdot \mathbf{t}_g^{(2)+} + \mathbf{h}_2^- \cdot \mathbf{t}_g^{(2)-}} \\ \times F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{M}^- \zeta(\mathbf{h}_1^-) - \mathbf{h}_1^- - \mathbf{h}_2^-, \mathbf{h}_2^-)\end{aligned}$$

with Friedel counterpart

$$\begin{aligned}F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-) \\ = (-1)^{\mathbf{h}_2^+ \cdot \mathbf{t}_g^{(2)+} + \mathbf{h}_2^- \cdot \mathbf{t}_g^{(2)-}} \overline{F[\mathbf{M}^+ \zeta(\mathbf{h}_1^+) - \mathbf{h}_1^+ - \mathbf{h}_2^+, \mathbf{h}_1^-, \mathbf{h}_2^-]}.\end{aligned}$$

When calculating electron densities, two methods may be used.

(i) Transform on \mathbf{h}^- first.

The partial vectors defined by $X_{\mathbf{h}^+, \mathbf{h}_2^-} = F(\mathbf{h}^+, \mathbf{h}_1^-, \mathbf{h}_2^-)$ obey symmetry relations of the form

$$X(\mathbf{h}_1^- - \mathbf{h}_2^-) = \varepsilon X[\mathbf{M}^- \zeta(\mathbf{h}_1^-) - \mathbf{h}_1^-]$$

with $\varepsilon = \pm 1$ independent of \mathbf{h}_1^- . This is the same relation as for the same parity class of data for a (complex or real) symmetric ($\varepsilon = +1$) or antisymmetric ($\varepsilon = -1$) transform. The same techniques can be used to decrease the number of $F(\mathbf{M}^-)$ by multiplexing pairs of such vectors and demultiplexing their transforms. Partial vectors with different values of ε may be mixed in this way (Section 1.3.4.3.5.6).

Once $F(\mathbf{N}^-)$ is completed, its results have Hermitian symmetry with respect to \mathbf{h}^+ , and the methods of Section 1.3.4.3.5.1 may be used to obtain the unique electron densities.

(ii) Transform on \mathbf{h}^+ first.

The partial vectors defined by $X_{\mathbf{h}^-, \mathbf{h}_2^+} = F(\mathbf{h}_1^+, \mathbf{h}_2^+, \mathbf{h}^-)$ obey symmetry relations of the form

$$X(\mathbf{h}_1^+ - \mathbf{h}_2^+) = \varepsilon X[\mathbf{M}^+ \zeta(\mathbf{h}_1^+) - \mathbf{h}_1^+]$$

with $\varepsilon = \pm 1$ independent of \mathbf{h}_1^+ . This is the same relation as for the same parity class of data for a Hermitian symmetric ($\varepsilon = +1$) or antisymmetric ($\varepsilon = -1$) transform. The same techniques may be used to decrease the number of $F(\mathbf{M}^+)$. This generalizes the procedure described by Ten Eyck (1973) for treating dyad axes, *i.e.* for the case $n_+ = 1$, $\mathbf{t}_g^{(2)-} = \mathbf{0}$, and $\mathbf{t}_g^{(2)+} = \mathbf{0}$ (simple dyad) or $\mathbf{t}_g^{(2)+} \neq \mathbf{0}$ (screw dyad).

Once $F(\mathbf{N}^+)$ is completed, its results have Hermitian symmetry properties with respect to \mathbf{h}^- which can be used to obtain the unique electron densities.

Structure factors may be computed by applying the reverse procedures in the reverse order.

1.3.4.3.6.3. Orthorhombic groups

Almost all orthorhombic space groups are generated by two monoclinic transformations g_1 and g_2 of the type described in Section 1.3.4.3.6.2, with the addition of a centre of inversion $-e$ for centrosymmetric groups. The only exceptions are $Fdd2$ and $Fddd$

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

which contain diamond glides, in which some non-primitive translations are ‘square roots’ not of primitive lattice translations, but of centring translations. The generic case will be examined first.

To calculate electron densities, the unique octant of data may first be transformed on \mathbf{h}^+ (respectively \mathbf{h}^-) as in Section 1.3.4.3.6.2 using the symmetry pertaining to generator g_1 . These intermediate results may then be expanded by generator g_2 by the formula of Section 1.3.4.3.3 prior to the final transform on \mathbf{h}^- (respectively \mathbf{h}^+). To calculate structure factors, the reverse operations are applied in the reverse order.

The two exceptional groups $Fdd2$ and $Fddd$ only require a small modification. The F -centring causes the systematic absence of parity classes with mixed parities, leaving only (000) and (111). For the former, the phase factors $\exp[2\pi i(\mathbf{h}^+ \cdot \mathbf{t}_g^+ + \mathbf{h}^- \cdot \mathbf{t}_g^-)]$ in the symmetry relations of Section 1.3.4.3.6.2 become powers of (-1) so that one is back to the generic case. For the latter, these phase factors are odd powers of i which it is a simple matter to incorporate into a modified multiplexing/demultiplexing procedure.

1.3.4.3.6.4. Trigonal, tetragonal and hexagonal groups

All the symmetries in this class of groups can be handled by the generalized Rader/Winograd algorithms of Section 1.3.4.3.4.3, but no implementation of these is yet available.

In groups containing axes of the form n_m with g.c.d. $(m, n) = 1$ (i.e. 3₁, 3₂, 4₁, 4₃, 6₁, 6₅) along the c direction, the following procedure may be used (Ten Eyck, 1973):

(i) to calculate electron densities, the unique structure factors indexed by

$$[\text{unique } (h, k)] \times (\text{all } l)$$

are transformed on l ; the results are rearranged by the transposition formula of Section 1.3.4.3.3 so as to be indexed by

$$[\text{all } (h, k)] \times \left[\text{unique } \left(\frac{1}{n} \right) \text{th of } z \right]$$

and are finally transformed on (h, k) to produce an asymmetric unit. For a dihedral group, the extra twofold axis may be used in the transposition to produce a unique $(1/2n)$ th of z .

(ii) to calculate structure factors, the unique densities in $(1/n)$ th of z [or $(1/2n)$ th for a dihedral group] are first transformed on x and y , then transposed by the formula of Section 1.3.4.3.3 to reindex the intermediate results by

$$[\text{unique } (h, k)] \times (\text{all } z);$$

the last transform on z is then carried out.

1.3.4.3.6.5. Cubic groups

These are usually treated as their orthorhombic or tetragonal subgroups, as the body-diagonal threefold axis cannot be handled by ordinary methods of decomposition.

The three-dimensional factorization technique of Section 1.3.4.3.4.1 allows a complete treatment of cubic symmetry. Factoring by 2 along all three dimensions gives four types (i.e. orbits) of parity classes:

(000)	with residual threefold symmetry,
(100), (010), (001)	related by threefold axis,
(110), (101), (011)	related by threefold axis,
(111)	with residual threefold symmetry.

Orbit exchange using the threefold axis thus allows one to reduce the number of partial transforms from 8 to 4 (one per orbit). Factoring by 3 leads to a reduction from 27 to 11 (in this case, further reduction to 9 can be gained by multiplexing the three

diagonal classes with residual threefold symmetry into a single class; see Section 1.3.4.3.5.6). More generally, factoring by q leads to a reduction from q^3 to $\frac{1}{3}(q^3 - q) - q$. Each of the remaining transforms then has a symmetry induced from the orthorhombic or tetragonal subgroup, which can be treated as above.

No implementation of this procedure is yet available.

1.3.4.3.6.6. Treatment of centred lattices

Lattice centring is an instance of the duality between periodization and decimation: the extra translational periodicity of ρ induces a decimation of $\mathbf{F} = \{F_{\mathbf{h}}\}$ described by the ‘reflection conditions’ on \mathbf{h} . As was pointed out in Section 1.3.4.2.2.3, non-primitive lattices are introduced in order to retain the same matrix representation for a given geometric symmetry operation in all the arithmetic classes in which it occurs. From the computational point of view, therefore, the main advantage in using centred lattices is that it *maximizes decomposability* (Section 1.3.4.2.2.4); reindexing to a primitive lattice would for instance often destroy the diagonal character of the matrix representing a dyad.

In the usual procedure involving three successive one-dimensional transforms, the loss of efficiency caused by the duplication of densities or the systematic vanishing of certain classes of structure factors may be avoided by using a multiplexing/demultiplexing technique (Ten Eyck, 1973):

(i) for base-centred or body-centred lattices, two successive planes of structure factors may be overlaid into a single plane; after transformation, the results belonging to each plane may be separated by parity considerations;

(ii) for face-centred lattices the same method applies, using four successive planes (the third and fourth with an origin translation);

(iii) for rhombohedral lattices in hexagonal coordinates, three successive planes may be overlaid, and the results may be separated by linear combinations involving cube roots of unity.

The three-dimensional factorization technique of Section 1.3.4.3.4.1 is particularly well suited to the treatment of centred lattices: if the decimation matrix of \mathbf{N} contains as a factor \mathbf{N}_1 a matrix which ‘integerizes’ all the non-primitive lattice vectors, then centring is reflected by the systematic vanishing of certain classes of vectors of decimated data or results, which can simply be omitted from the calculation. An alternative possibly is to reindex on a primitive lattice and use different representative matrices for the symmetry operations: the loss of decomposability is of little consequence in this three-dimensional scheme, although it substantially complicates the definition of the cocycles μ_2 and η_1 .

1.3.4.3.6.7. Programming considerations

The preceding sections have been devoted to showing how the raw computational efficiency of a crystallographic Fourier transform *algorithm* can be maximized. This section will briefly discuss another characteristic (besides speed) which a crystallographic Fourier transform *program* may be required to possess if it is to be useful in various applications: a convenient and versatile mode of presentation of input data or output results.

The standard crystallographic FFT programs (Ten Eyck, 1973, 1985) are rather rigid in this respect, and use rather rudimentary data structures (lists of structure-factor values, and two-dimensional arrays containing successive sections of electron-density maps). It is frequently the case that considerable reformatting of these data or results must be carried out before they can be used in other computations; for instance, maps have to be converted from 2D sections to 3D ‘bricks’ before they can be inspected on a computer graphics display.

The explicitly three-dimensional approach to the factorization of the DFT and the use of symmetry offers the possibility of richer and more versatile data structures. For instance, the use of ‘decimation in frequency’ in real space and of ‘decimation in time’ in reciprocal

1. GENERAL RELATIONSHIPS AND TECHNIQUES

space leads to data structures in which real-space coordinates are handled by blocks (thus preserving, at least locally, the three-dimensional topological connectivity of the maps) while reciprocal-space indices are handled by parity classes or their generalizations for factors other than 2 (thus making the treatment of centred lattices extremely easy). This global three-dimensional indexing also makes it possible to carry symmetry and multiplicity characteristics for each subvector of intermediate results for the purpose of automating the use of the orbit exchange mechanism.

Brünger (1989) has described the use of a similar three-dimensional factoring technique in the context of structure-factor calculations for the refinement of macromolecular structures.

1.3.4.4. Basic crystallographic computations

1.3.4.4.1. Introduction

Fourier transform (FT) calculations play an indispensable role in crystallography, because the Fourier transformation is inherent in the diffraction phenomenon itself.

Besides this obligatory use, the FT has numerous other applications, motivated more often by its mathematical properties than by direct physical reasoning (although the latter can be supplied after the fact). Typically, many crystallographic computations turn out to be convolutions in disguise, which can be speeded up by orders of magnitude through a judicious use of the FT. Several recent advances in crystallographic computation have been based on this kind of observation.

1.3.4.4.2. Fourier synthesis of electron-density maps

Bragg (1929) was the first to use this type of calculation to assist structure determination. Progress in computing techniques since that time was reviewed in Section 1.3.4.3.1.

The usefulness of the maps thus obtained can be adversely affected by three main factors:

- (i) limited resolution;
- (ii) errors in the data;
- (iii) computational errors.

Limited resolution causes 'series-termination errors' first investigated by Bragg & West (1930), who used an optical analogy with the numerical aperture of a microscope. James (1948*b*) gave a quantitative description of this phenomenon as a convolution with the 'spherical Dirichlet kernel' (Section 1.3.4.2.1.3), which reflects the truncation of the Fourier spectrum by multiplication with the indicator function of the limiting resolution sphere. Bragg & West (1930) suggested that the resulting ripples might be diminished by applying an artificial temperature factor to the data, which performs a further convolution with a Gaussian point-spread function. When the electron-density map is to be used for model refinement, van Reijen (1942) suggested using Fourier coefficients calculated from the model when no observation is available, as a means of combating series-termination effects.

Errors in the data introduce errors in the electron-density maps, with the same mean-square value by virtue of Parseval's theorem. Special positions accrue larger errors (Cruickshank & Rollett, 1953; Cruickshank, 1965*a*). To minimize the mean-square electron-density error due to large phase uncertainties, Blow & Crick (1959) introduced the 'best Fourier' which uses centroid Fourier coefficients; the associated error level in the electron-density map was evaluated by Blow & Crick (1959) and Dickerson *et al.* (1961*a,b*).

Computational errors used to be a serious concern when Beevers-Lipson strips were used, and Cochran (1948*a*) carried out a critical evaluation of the accuracy limitations imposed by strip methods. Nowadays, the FFT algorithm implemented on digital computers with a word size of at least 32 bits gives results accurate

to six decimal places or better in most applications (see Gentleman & Sande, 1966).

1.3.4.4.3. Fourier analysis of modified electron-density maps

Various approaches to the phase problem are based on certain modifications of the electron-density map, followed by Fourier analysis of the modified map and extraction of phase information from the resulting Fourier coefficients.

1.3.4.4.3.1. Squaring

Sayre (1952*a*) derived his 'squaring method equation' for structures consisting of equal, resolved and spherically symmetric atoms by observing that squaring such an electron density is equivalent merely to sharpening each atom into its square. Thus

$$F_{\mathbf{h}} = \theta_{\mathbf{h}} \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}},$$

where $\theta_{\mathbf{h}} = f(\mathbf{h})/f^{\text{sq}}(\mathbf{h})$ is the ratio between the form factor $f(\mathbf{h})$ common to all the atoms and the form factor $f^{\text{sq}}(\mathbf{h})$ for the squared version of that atom.

Most of the central results of direct methods, such as the tangent formula, are an immediate consequence of Sayre's equation. Phase refinement for a macromolecule by enforcement of the squaring method equation was demonstrated by Sayre (1972, 1974).

1.3.4.4.3.2. Other non-linear operations

A category of phase improvement procedures known as 'density modification' is based on the pointwise application of various quadratic or cubic 'filters' to electron-density maps after removal of negative regions (Hoppe & Gassmann, 1968; Hoppe *et al.*, 1970; Barrett & Zwick, 1971; Gassmann & Zechmeister, 1972; Collins, 1975; Collins *et al.*, 1976; Gassmann, 1976). These operations are claimed to be equivalent to reciprocal-space phase-refinement techniques such as those based on the tangent formula. Indeed the replacement of

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

by $P[\rho(\mathbf{x})]$, where P is a polynomial

$$P(\rho) = a_0 + a_1\rho + a_2\rho^2 + a_3\rho^3 + \dots$$

yields

$$P[\rho(\mathbf{x})] = a_0 + \sum_{\mathbf{h}} \left[a_1 F_{\mathbf{h}} + a_2 \sum_{\mathbf{k}} F_{\mathbf{k}} F_{\mathbf{h}-\mathbf{k}} + a_3 \sum_{\mathbf{k}} \sum_{\mathbf{l}} F_{\mathbf{k}} F_{\mathbf{l}} F_{\mathbf{h}-\mathbf{k}-\mathbf{l}} + \dots \right] \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

and hence gives rise to the convolution-like families of terms encountered in direct methods. This equivalence, however, has been shown to be rather superficial (Bricogne, 1982) because the 'uncertainty principle' embodied in Heisenberg's inequality (Section 1.3.2.4.4.3) imposes severe limitations on the effectiveness of any procedure which operates *pointwise* in *both* real and reciprocal space.

In applying such methods, sampling considerations must be given close attention. If the spectrum of ρ extends to resolution Δ and if the pointwise non-linear filter involves a polynomial P of 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,

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

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):

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

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 ρ , 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 ρ with χ_S , it can be speeded up considerably (Leslie, 1987) by computing the moving average ρ_{mov} as

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

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 χ_U in real space, whereas they involve a *convolution* with $\tilde{\mathcal{F}}[\chi_U]$ 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 non-integral 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 ρ 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 ρ 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 ρ . 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

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

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 $\tilde{\mathcal{F}}[\rho](\boldsymbol{\xi})$ attenuated by multiplication by the central region of $\tilde{\mathcal{F}}[K][(\mathbf{N}_1^{-1})^T \boldsymbol{\xi}]$; in the case of linear interpolation, for example,

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

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

Thus $I_{\mathbf{N}_1}\rho$ is not band-limited even if ρ is. Supposing, however, that ρ is band-limited and that grid \mathcal{G}_1 satisfies the Shannon sampling criterion, we see that there will be no overlap between the different bands: $\tilde{\mathcal{F}}[\rho]$ 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 \mathcal{G}_2 , which causes its transform to become periodized into

$$|\det \mathbf{N}_2| \sum_{\mathbf{k}_2 \in \mathbb{Z}^3} \tau_{\mathbf{N}_2^T \mathbf{k}_2} \left\{ \left[\sum_{\mathbf{k}_1 \in \mathbb{Z}^3} \tau_{\mathbf{N}_1^T \mathbf{k}_1} \tilde{\mathcal{F}}[\rho] \right] (\mathbf{N}_1^T)^\# \tilde{\mathcal{F}}[K] \right\}.$$

1. GENERAL RELATIONSHIPS AND TECHNIQUES

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 $I_{N_1, \rho}$.

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

The sampling rate in grid \mathcal{G}_2 needs only exceed the Shannon minimum to the extent required to accommodate the increase in bandwidth due to convolution with $\tilde{\mathcal{F}}[\chi_U]$, 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 $\tilde{\mathcal{F}}[\chi_U]$ 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:

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

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:

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

i.e.

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

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:

$$\frac{1}{\text{vol}(U)} \tilde{\mathcal{F}}[\chi_U](\mathbf{H}) = \frac{3}{X^3} [\sin X - X \cos X]$$

$$\text{with } X = 2\pi \|\mathbf{H}\| R.$$

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 $\propto N \mathcal{N}$ for N atoms and \mathcal{N} reflections. Since N and \mathcal{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 (1936*b*) 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 form factor was indicated in Section 1.3.4.4.3.4 and gives rise to aliasing effects (see below) not considered by Beevers & Lipson.

The correct formulation of this idea came with the work of Sayre (1951), who showed that structure factors could be calculated by Fourier analysis of a sampled electron-density map previously generated on a subdivision $\mathbf{N}^{-1}\Lambda$ of the crystal lattice Λ . When generating such a map, care must be taken to distribute onto the sample grid not only the electron densities of all the atoms in the asymmetric motif, but also those of their images under space-group symmetries and lattice translations. Considerable savings in computation occur, especially for large structures, because atoms are *localized*: each atom sends contributions to only a few grid points in real space, rather than to all reciprocal-lattice points. The generation of the sampled electron-density map is still of complexity $\propto N \mathcal{N}$ for N atoms and \mathcal{N} reflections, but the proportionality constant is smaller than that in Section 1.3.4.4.4 by orders of magnitude; the extra cost of Fourier analysis, proportional to $\mathcal{N} \log \mathcal{N}$, is negligible.

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

The idea of approximating a Fourier transform by a discrete transform on sampled values had already been used by Whittaker (1948), who tested it on the first three odd Hermite functions and did not consider the problem of aliasing errors. By contrast, Sayre gave a lucid analysis of the sampling problems associated to this technique. If the periodic sampled map is written in the form of a weighted lattice distribution (as in Section 1.3.2.7.3) as

$$\rho^s = \sum_{\mathbf{m} \in \mathbb{Z}^3} \rho(\mathbf{N}^{-1}\mathbf{m}) \delta_{(\mathbf{N}^{-1}\mathbf{m})},$$

then its discrete Fourier transform yields

$$F^s(\mathbf{h}) = \sum_{\boldsymbol{\eta} \in \mathbb{Z}^3} F(\mathbf{h} + \mathbf{N}^T \boldsymbol{\eta})$$

so that each correct value $F(\mathbf{h})$ is corrupted by its aliases $F(\mathbf{h} + \mathbf{N}^T \boldsymbol{\eta})$ for $\boldsymbol{\eta} \neq \mathbf{0}$.

To cure this aliasing problem, Sayre used ‘hypothetical atoms’ with form factors equal to those of standard atoms within the resolution range of interest, but set to zero outside that range. This amounts to using atomic densities with built-in series-termination errors, which has the detrimental effect of introducing slowly decaying ripples around the atom which require incrementing sample densities at many more grid points per atom.

Sayre considered another cure in the form of an artificial temperature factor B (Bragg & West, 1930) applied to all atoms. This spreads each atom on more grid points in real space but speeds up the decay of its transform in reciprocal space, thus allowing the use of a coarser sampling grid in real space. He discounted it as spoiling the agreement with observed data, but Ten Eyck (1977) pointed out that this agreement could be restored by applying the negative of the artificial temperature factor to the results. This idea cannot be carried to extremes: if B is chosen too large, the atoms will be so spread out in real space as each to occupy a sizeable fraction of the unit cell and the advantage of atom localization will be lost; furthermore, the form factors will fall off so rapidly that round-off error amplification will occur when the results are sharpened back. Clearly, there exists an optimal combination of B and sampling rate yielding the most economical computation for a given accuracy at a given resolution, and a formula will now be given to calculate it.

Let us make the simplifying assumption that all atoms are roughly equal and that their common form factor can be represented by an equivalent temperature factor B_{eq} . Let $\Delta = 1/d_{\text{max}}^*$ be the resolution to which structure factors are wanted. The Shannon sampling interval is $\Delta/2 = 1/2d_{\text{max}}^*$. Let σ be the oversampling rate, so that the actual sampling interval in the map is $\Delta/2\sigma = 1/2\sigma d_{\text{max}}^*$; then consecutive copies of the transform are separated by a distance $2\sigma d_{\text{max}}^*$ in reciprocal space. Let the artificial temperature factor B_{extra} be added, and let

$$B = B_{\text{eq}} + B_{\text{extra}}.$$

The worst aliasing occurs at the outer resolution limit d_{max}^* , where the ‘signal’ due to an atom is proportional to

$$\exp[(-B/4)(d_{\text{max}}^*)^2],$$

while the ‘noise’ due to the closest alias is proportional to

$$\exp\{(-B/4)[(2\sigma - 1)d_{\text{max}}^*]^2\}.$$

Thus the signal-to-noise ratio, or quality factor, Q is

$$\exp[B\sigma(\sigma - 1)(d_{\text{max}}^*)^2].$$

If a certain value of Q is desired (e.g. $Q = 100$ for 1% accuracy), then the equation

$$B = \frac{\log Q}{\sigma(\sigma - 1)(d_{\text{max}}^*)^2}$$

defines B in terms of σ , d_{max}^* and Q .

The overall cost of the structure-factor calculation from N atoms is then

- (i) $C_1 \times B^{2/3} \times N$ for density generation,
- (ii) $C_2 \times (2\sigma d_{\text{max}}^*)^3 \times \log[(2\sigma d_{\text{max}}^*)^3]$ for Fourier analysis,

where C_1 and C_2 are constant depending on the speed of the computer used. This overall cost may be minimized with respect to σ for given d_{max}^* and Q , determining the optimal B (and hence B_{extra}) in passing by the above relation.

Sayre (1951) did observe that applying an artificial temperature factor in real space would not create series-termination ripples: the resulting atoms would have a smaller effective radius than his hypothetical atoms, so that step (i) would be faster. This optimality of Gaussian smearing is ultimately a consequence of Hardy’s theorem (Section 1.3.2.4.4.3).

1.3.4.4.6. Derivatives for variational phasing techniques

Some methods of phase determination rely on maximizing a certain global criterion $S[\rho]$ involving the electron density, of the form $\int_{\mathbb{R}^3/\mathbb{Z}^3} K[\rho(\mathbf{x})] d^3\mathbf{x}$, under constraint of agreement with the observed structure-factor amplitudes, typically measured by a χ^2 residual C . Several recently proposed methods use for $S[\rho]$ various measures of entropy defined by taking $K(\rho) = -\rho \log(\rho/\mu)$ or $K(\rho) = \log \rho$ (Bricogne, 1982; Britten & Collins, 1982; Narayan & Nityananda, 1982; Bryan *et al.*, 1983; Wilkins *et al.*, 1983; Bricogne, 1984; Navaza, 1985; Livesey & Skilling, 1985). Sayre’s use of the squaring method to improve protein phases (Sayre, 1974) also belongs to this category, and is amenable to the same computational strategies (Sayre, 1980).

These methods differ from the density-modification procedures of Section 1.3.4.4.3.2 in that they seek an optimal solution by moving electron densities (or structure factors) *jointly* rather than pointwise, *i.e.* by moving along suitably chosen *search directions* $v_i(\mathbf{x})$ [or $V_i(\mathbf{h})$].

For computational purposes, these search directions may be handled either as column vectors of sample values $\{v_i(\mathbf{N}^{-1}\mathbf{m})\}_{\mathbf{m} \in \mathbb{Z}^3/\mathbb{N}\mathbb{Z}^3}$ on a grid in real space, or as column vectors of Fourier coefficients $\{V_i(\mathbf{h})\}_{\mathbf{h} \in \mathbb{Z}^3/\mathbb{N}^T\mathbb{Z}^3}$ in reciprocal space. These column vectors are the coordinates of the same vector \mathbf{V}_i in an abstract vector space $\mathcal{V} \cong L(\mathbb{Z}^3/\mathbb{N}\mathbb{Z}^3)$ of dimension $\mathcal{N} = |\det \mathbf{N}|$ over \mathbb{R} , but referred to two different bases which are related by the DFT and its inverse (Section 1.3.2.7.3).

The problem of finding the optimum of S for a given value of C amounts to achieving collinearity between the gradients ∇S and ∇C of S and of C in \mathcal{V} , the scalar ratio between them being a Lagrange multiplier. In order to move towards such a solution from a trial position, the dependence of ∇S and ∇C on position in \mathcal{V} must be represented. This involves the $\mathcal{N} \times \mathcal{N}$ Hessian matrices $\mathbf{H}(S)$ and $\mathbf{H}(C)$, whose size precludes their use in the whole of \mathcal{V} . Restricting the search to a smaller *search subspace* of dimension n spanned by $\{\mathbf{V}_i\}_{i=1, \dots, n}$ we may build local quadratic models of S and C (Bryan & Skilling, 1980; Burch *et al.*, 1983) with respect to n coordinates \mathbf{X} in that subspace:

$$\begin{aligned} S(\mathbf{X}) &= S(\mathbf{X}_0) + \mathbf{S}_0^T(\mathbf{X} - \mathbf{X}_0) \\ &\quad + \frac{1}{2}(\mathbf{X} - \mathbf{X}_0)^T \mathbf{H}_0(S)(\mathbf{X} - \mathbf{X}_0) \\ C(\mathbf{X}) &= C(\mathbf{X}_0) + \mathbf{C}_0^T(\mathbf{X} - \mathbf{X}_0) \\ &\quad + \frac{1}{2}(\mathbf{X} - \mathbf{X}_0)^T \mathbf{H}_0(C)(\mathbf{X} - \mathbf{X}_0). \end{aligned}$$

The coefficients of these linear models are given by scalar products:

1. GENERAL RELATIONSHIPS AND TECHNIQUES

$$\begin{aligned} [\mathbf{S}_0]_i &= (\mathbf{V}_i, \nabla S) \\ [\mathbf{C}_0]_i &= (\mathbf{V}_i, \nabla C) \\ [\mathbf{H}_0(S)]_{ij} &= [\mathbf{V}_i, \mathbf{H}(S)\mathbf{V}_j] \\ [\mathbf{H}_0(C)]_{ij} &= [\mathbf{V}_i, \mathbf{H}(C)\mathbf{V}_j] \end{aligned}$$

which, by virtue of Parseval's theorem, may be evaluated either in real space or in reciprocal space (Bricogne, 1984). In doing so, special positions and reflections must be taken into account, as in Section 1.3.4.2.2.8. Scalar products involving S are best evaluated by real-space grid summation, because $\mathbf{H}(S)$ is *diagonal* in this representation; those involving C are best calculated by reciprocal-space summation, because $\mathbf{H}(C)$ is at worst 2×2 *block-diagonal* in this representation. Using these Hessian matrices in the wrong space would lead to prohibitively expensive convolutions instead of scalar (or at worst 2×2 matrix) multiplications.

1.3.4.4.7. Derivatives for model refinement

Since the origins of X-ray crystal structure analysis, the calculation of crystallographic Fourier series has been closely associated with the process of refinement. Fourier coefficients *with phases* were obtained for all or part of the measured reflections in the basis of some trial model for all or part of the structure, and Fourier syntheses were then used to complete and improve this initial model. This approach is clearly described in the classic paper by Bragg & West (1929), and was put into practice in the determination of the structures of topaz (Alston & West, 1929) and diopside (Warren & Bragg, 1929). Later, more systematic methods of arriving at a trial model were provided by the Patterson synthesis (Patterson, 1934, 1935*a,b*; Harker, 1936) and by isomorphous replacement (Robertson, 1935, 1936*c*). The role of Fourier syntheses, however, remained essentially unchanged [see Robertson (1937) for a review] until more systematic methods of structure refinement were introduced in the 1940s. A particularly good account of the processes of structure completion and refinement may be found in Chapters 15 and 16 of Stout & Jensen (1968).

It is beyond the scope of this section to review the vast topic of refinement methods: rather, it will give an account of those aspects of their development which have sought improved power by exploiting properties of the Fourier transformation. It is of more than historical interest that some recent advances in the crystallographic refinement of macromolecular structures had been anticipated by Cochran and Cruickshank in the early 1950s.

1.3.4.4.7.1. The method of least squares

Hughes (1941) was the first to use the already well established multivariate least-squares method (Whittaker & Robinson, 1944) to refine initial estimates of the parameters describing a model structure. The method gained general acceptance through the programming efforts of Friedlander *et al.* (1955), Sparks *et al.* (1956), Busing & Levy (1961), and others.

The Fourier relations between ρ and F (Section 1.3.4.2.2.6) are used to derive the 'observational equations' connecting the structure parameters $\{u_p\}_{p=1, \dots, n}$ to the observations $\{|F_{\mathbf{h}}|^{\text{obs}}, (\sigma_{\mathbf{h}}^2)^{\text{obs}}\}_{\mathbf{h} \in \mathcal{H}}$ comprising the amplitudes and their experimental variances for a set \mathcal{H} of unique reflections.

The normal equations giving the corrections $\delta \mathbf{u}$ to the parameters are then

$$(\mathbf{A}^T \mathbf{W} \mathbf{A}) \delta \mathbf{u} = -\mathbf{A}^T \mathbf{W} \Delta,$$

where

$$\begin{aligned} A_{hp} &= \frac{\partial |F_{\mathbf{h}}^{\text{calc}}|}{\partial u_p} \\ \Delta_{\mathbf{h}} &= |F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}} \\ \mathbf{W} &= \text{diag}(W_{\mathbf{h}}) \quad \text{with} \quad W_{\mathbf{h}} = \frac{1}{(\sigma_{\mathbf{h}}^2)^{\text{obs}}}. \end{aligned}$$

To calculate the elements of \mathbf{A} , write:

$$F = |F| \exp(i\varphi) = \alpha + i\beta;$$

hence

$$\begin{aligned} \frac{\partial |F|}{\partial u} &= \frac{\partial \alpha}{\partial u} \cos \varphi + \frac{\partial \beta}{\partial u} \sin \varphi \\ &= \Re \left[\frac{\partial F}{\partial u} \overline{\exp(i\varphi)} \right] = \Re \left[\frac{\partial F}{\partial u} \exp(i\varphi) \right]. \end{aligned}$$

In the simple case of atoms with real-valued form factors and isotropic thermal agitation in space group $P1$,

$$F_{\mathbf{h}}^{\text{calc}} = \sum_{j \in J} g_j(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_j),$$

where

$$g_j(\mathbf{h}) = Z_j f_j(\mathbf{h}) \exp[-\frac{1}{4} B_j (d_{\mathbf{h}}^*)^2],$$

Z_j being a fractional occupancy.

Positional derivatives with respect to \mathbf{x}_j are given by

$$\begin{aligned} \frac{\partial F_{\mathbf{h}}^{\text{calc}}}{\partial \mathbf{x}_j} &= (2\pi i \mathbf{h}) g_j(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_j) \\ \frac{\partial |F_{\mathbf{h}}^{\text{calc}}|}{\partial \mathbf{x}_j} &= \Re \left[(-2\pi i \mathbf{h}) g_j(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) \exp(i\varphi_{\mathbf{h}}^{\text{calc}}) \right] \end{aligned}$$

so that the corresponding 3×1 subvector of the right-hand side of the normal equations reads:

$$\begin{aligned} & - \sum_{\mathbf{h} \in \mathcal{H}} W_{\mathbf{h}} \frac{\partial |F_{\mathbf{h}}^{\text{calc}}|}{\partial \mathbf{x}_j} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}}) \\ & = -\Re \left[\sum_{\mathbf{h} \in \mathcal{H}} g_j(\mathbf{h}) (-2\pi i \mathbf{h}) W_{\mathbf{h}} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}}) \right. \\ & \quad \left. \times \exp(i\varphi_{\mathbf{h}}^{\text{calc}}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) \right]. \end{aligned}$$

The setting up and solution of the normal equations lends itself well to computer programming and has the advantage of providing a thorough analysis of the accuracy of its results (Cruickshank, 1965*b*, 1970; Rollett, 1970). It is, however, an expensive task, of complexity $\propto n \times |\mathcal{H}|^2$, which is unaffordable for macromolecules.

1.3.4.4.7.2. Booth's differential Fourier syntheses

It was the use of Fourier syntheses in the completion of trial structures which provided the incentive to find methods for computing 2D and 3D syntheses efficiently, and led to the Beever-Lipson strips. The limited accuracy of the latter caused the estimated positions of atoms (identified as peaks in the maps) to be somewhat in error. Methods were therefore sought to improve the accuracy with which the coordinates of the electron-density maxima could be determined. The naive method of peak-shape analysis from densities recalculated on a $3 \times 3 \times 3$ grid using high-accuracy trigonometric tables entailed 27 summations per atom.

Booth (1946*a*) suggested examining the rapidly varying derivatives of the electron density rather than its slowly varying values. If

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

$$\rho(\mathbf{x}) = \sum_{\mathbf{h}} F_{\mathbf{h}} \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

then the gradient vector $\nabla_{\mathbf{x}}\rho$ of ρ at \mathbf{x}^0

$$(\nabla_{\mathbf{x}}\rho)(\mathbf{x}^0) = \sum_{\mathbf{h}} F_{\mathbf{h}}(-2\pi i \mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}^0)$$

can be calculated by means of three Fourier summations from the 3×1 vector of Fourier coefficients

$$(-2\pi i \mathbf{h}) F_{\mathbf{h}}.$$

Similarly, the Hessian matrix of ρ at \mathbf{x}^0

$$[(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T)\rho](\mathbf{x}^0) = \sum_{\mathbf{h}} F_{\mathbf{h}}(-4\pi^2 \mathbf{h}\mathbf{h}^T) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}^0)$$

can be calculated by six Fourier summations from the unique elements of the symmetric matrix of Fourier coefficients:

$$-4\pi^2 \begin{pmatrix} h^2 & hk & hl \\ hk & k^2 & kl \\ hl & kl & l^2 \end{pmatrix} F_{\mathbf{h}}.$$

The scalar maps giving the components of the gradient and Hessian matrix of ρ will be called differential syntheses of 1st order and 2nd order respectively. If \mathbf{x}^0 is approximately but not exactly a maximum of ρ , then the Newton–Raphson estimate of the true maximum \mathbf{x}^* is given by:

$$\mathbf{x}^* = \mathbf{x}^0 - [(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^T)\rho](\mathbf{x}^0)^{-1}[\nabla_{\mathbf{x}}\rho(\mathbf{x}^0)].$$

This calculation requires only nine accurate Fourier summations (instead of 27), and this number is further reduced to four if the peak is assumed to be spherically symmetrical.

The resulting positions are affected by series-termination errors in the differential syntheses. Booth (1945c, 1946c) proposed a ‘back-shift correction’ to eliminate them, and extended this treatment to the acentric case (Booth, 1946b). He cautioned against the use of an artificial temperature factor to fight series-termination errors (Brill *et al.*, 1939), as this could be shown to introduce coordinate errors by causing overlap between atoms (Booth, 1946c, 1947a,b).

Cruickshank was able to derive estimates for the standard uncertainties of the atomic coordinates obtained in this way (Cox & Cruickshank, 1948; Cruickshank, 1949a,b) and to show that they agreed with those provided by the least-squares method.

The calculation of differential Fourier syntheses was incorporated into the crystallographic programs of Ahmed & Cruickshank (1953b) and of Sparks *et al.* (1956).

1.3.4.4.7.3. Booth’s method of steepest descents

Having defined the now universally adopted R factors (Booth, 1945b) as criteria of agreement between observed and calculated amplitudes or intensities, Booth proposed that R should be minimized with respect to the set of atomic coordinates $\{\mathbf{x}_j\}_{j \in J}$ by descending along the gradient of R in parameter space (Booth, 1947c,d). This ‘steepest descents’ procedure was compared with Patterson methods by Cochran (1948d).

When calculating the necessary derivatives, Booth (1948a, 1949) used the formulae given above in connection with least squares. This method was implemented by Qurashi (1949) and by Vand (1948, 1951) with parameter-rescaling modifications which made it very close to the least-squares method (Cruickshank, 1950; Qurashi & Vand, 1953; Qurashi, 1953).

1.3.4.4.7.4. Cochran’s Fourier method

Cochran (1948b,c, 1951a) undertook to exploit an algebraic similarity between the right-hand side of the normal equations in the least-squares method on the one hand, and the expression for the

coefficients used in Booth’s differential syntheses on the other hand (see also Booth, 1948a). In doing so he initiated a remarkable sequence of formal and computational developments which are still actively pursued today.

Let $\rho_C(\mathbf{x})$ be the electron-density map corresponding to the current atomic model, with structure factors $|F_{\mathbf{h}}^{\text{calc}}| \exp(i\varphi_{\mathbf{h}}^{\text{calc}})$; and let $\rho_O(\mathbf{x})$ be the map calculated from observed moduli and calculated phases, *i.e.* with coefficients $\{|F_{\mathbf{h}}|^{\text{obs}} \exp(i\varphi_{\mathbf{h}}^{\text{calc}})\}_{\mathbf{h} \in \mathcal{H}}$. If there are enough data for ρ_C to have a resolved peak at each model atomic position \mathbf{x}_j , then

$$(\nabla_{\mathbf{x}}\rho_C)(\mathbf{x}_j) = \mathbf{0} \quad \text{for each } j \in J;$$

while if the calculated phases $\varphi_{\mathbf{h}}^{\text{calc}}$ are good enough, ρ_O will also have peaks at each \mathbf{x}_j :

$$(\nabla_{\mathbf{x}}\rho_O)(\mathbf{x}_j) = \mathbf{0} \quad \text{for each } j \in J.$$

It follows that

$$\begin{aligned} [\nabla_{\mathbf{x}}(\rho_C - \rho_O)](\mathbf{x}_j) &= \sum_{\mathbf{h}} (-2\pi i \mathbf{h}) [(|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}}) \exp(i\varphi_{\mathbf{h}}^{\text{calc}})] \\ &\quad \times \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) \\ &= \mathbf{0} \quad \text{for each } j \in J, \end{aligned}$$

where the summation is over all reflections in \mathcal{H} or related to \mathcal{H} by space-group and Friedel symmetry (overlooking multiplicity factors!). This relation is less sensitive to series-termination errors than either of the previous two, since the spectrum of ρ_O could have been extrapolated beyond the data in \mathcal{H} by using that of ρ_C [as in van Reijen (1942)] without changing its right-hand side.

Cochran then used the identity

$$\frac{\partial F_{\mathbf{h}}^{\text{calc}}}{\partial \mathbf{x}_j} = (2\pi i \mathbf{h}) g_j(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_j)$$

in the form

$$(-2\pi i \mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) = \frac{1}{g_j(\mathbf{h})} \frac{\partial F_{\mathbf{h}}^{\text{calc}}}{\partial \mathbf{x}_j}$$

to rewrite the previous relation as

$$\begin{aligned} [\nabla_{\mathbf{x}}(\rho_C - \rho_O)](\mathbf{x}_j) &= \sum_{\mathbf{h}} \frac{1}{g_j(\mathbf{h})} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}}) \mathcal{R}e \left[\frac{\partial F_{\mathbf{h}}^{\text{calc}}}{\partial \mathbf{x}_j} \exp(i\varphi_{\mathbf{h}}^{\text{calc}}) \right] \\ &= \sum_{\mathbf{h}} \frac{1}{g_j(\mathbf{h})} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}}) \frac{\partial |F_{\mathbf{h}}^{\text{calc}}|}{\partial \mathbf{x}_j} \\ &= \mathbf{0} \quad \text{for each } j \in J \end{aligned}$$

(the operation $\mathcal{R}e$ [] on the first line being neutral because of Friedel symmetry). This is equivalent to the vanishing of the 3×1 subvector of the right-hand side of the normal equations associated to a least-squares refinement in which the weights would be

$$W_{\mathbf{h}} = \frac{1}{g_j(\mathbf{h})}.$$

Cochran concluded that, for equal-atom structures with $g_j(\mathbf{h}) = g(\mathbf{h})$ for all j , the positions \mathbf{x}_j obtained by Booth’s method applied to the difference map $\rho_O - \rho_C$ are such that they minimize the residual

$$\frac{1}{2} \sum_{\mathbf{h}} \frac{1}{g(\mathbf{h})} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}|^{\text{obs}})^2$$

with respect to the atomic positions. If it is desired to minimize the residual of the ordinary least-squares method, then the differential

1. GENERAL RELATIONSHIPS AND TECHNIQUES

synthesis method should be applied to the *weighted* difference map

$$\sum_{\mathbf{h}} \frac{W_{\mathbf{h}}}{g(\mathbf{h})} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}^{\text{obs}}|) \exp(i\varphi_{\mathbf{h}}^{\text{calc}}).$$

He went on to show (Cochran, 1951*b*) that the refinement of temperature factors could also be carried out by inspecting appropriate derivatives of the weighted difference map.

This Fourier method was used by Freer *et al.* (1976) in conjunction with a stereochemical regularization procedure to refine protein structures.

1.3.4.4.7.5. Cruickshank's modified Fourier method

Cruickshank consolidated and extended Cochran's derivations in a series of classic papers (Cruickshank, 1949*b*, 1950, 1952, 1956). He was able to show that all the coefficients involved in the right-hand side and normal matrix of the least-squares method could be calculated by means of suitable differential Fourier syntheses *even when the atoms overlap*. This remarkable achievement lay essentially dormant until its independent rediscovery by Agarwal in 1978 (Section 1.3.4.4.7.6).

To ensure rigorous equivalence between the summations over $\mathbf{h} \in \mathcal{H}$ (in the expressions of least-squares right-hand side and normal matrix elements) and genuine Fourier summations, multiplicity-corrected weights were introduced by:

$$w_{\mathbf{h}} = \frac{1}{|G_{\mathbf{h}}|} W_{\mathbf{h}} \quad \text{if } \mathbf{h} \in G_{\mathbf{h}} \quad \text{with } \mathbf{h} \in \mathcal{H},$$

$$w_{\mathbf{h}} = 0 \quad \text{otherwise,}$$

where $G_{\mathbf{h}}$ denotes the orbit of \mathbf{h} and $G_{\mathbf{h}}$ its isotropy subgroup (Section 1.3.4.2.2.5). Similarly, derivatives with respect to parameters of symmetry-unique atoms were expressed, *via* the chain rule, as sums over the orbits of these atoms.

Let $p = 1, \dots, n$ be the label of a parameter u_p belonging to atoms with label j . Then Cruickshank showed that the p th element of the right-hand side of the normal equations can be obtained as $D_{p,j}(\mathbf{x}_j)$, where $D_{p,j}$ is a differential synthesis of the form

$$D_{p,j}(\mathbf{x}) = \sum_{\mathbf{h}} P_p(\mathbf{h}) g_j(\mathbf{h}) w_{\mathbf{h}} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}^{\text{obs}}|) \times \exp(i\varphi_{\mathbf{h}}^{\text{calc}}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

with $P_p(\mathbf{h})$ a polynomial in (h, k, l) depending on the type of parameter p . The correspondence between parameter type and the associated polynomial extends Booth's original range of differential syntheses, and is recapitulated in the following table.

Parameter type	$P(h, k, l)$
x coordinate	$-2\pi ih$
y coordinate	$-2\pi ik$
z coordinate	$-2\pi il$
B isotropic	$-\frac{1}{4}(d_{\mathbf{h}}^*)^2$
B^{11} anisotropic	$-h^2$
B^{12} anisotropic	$-hk$
B^{13} anisotropic	$-hl$
B^{22} anisotropic	$-k^2$
B^{23} anisotropic	$-kl$
B^{33} anisotropic	$-l^2$

Unlike Cochran's original heuristic argument, this result does not depend on the atoms being resolved.

Cruickshank (1952) also considered the elements of the normal matrix, of the form

$$\sum_{\mathbf{h}} w_{\mathbf{h}} \frac{\partial |F_{\mathbf{h}}^{\text{calc}}|}{\partial u_p} \frac{\partial |F_{\mathbf{h}}^{\text{calc}}|}{\partial u_q}$$

associated with positional parameters. The 3×3 block for parameters \mathbf{x}_j and \mathbf{x}_k may be written

$$\sum_{\mathbf{h}} w_{\mathbf{h}} (\mathbf{h}\mathbf{h}^T) \Re[(-2\pi i) g_j(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_j) \exp(i\varphi_{\mathbf{h}}^{\text{calc}})] \times \Re[(-2\pi i) g_k(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}_k) \exp(i\varphi_{\mathbf{h}}^{\text{calc}})]$$

which, using the identity

$$\Re(z_1) \Re(z_2) = \frac{1}{2} [\Re(z_1 z_2) + \Re(z_1 \bar{z}_2)],$$

becomes

$$2\pi^2 \sum_{\mathbf{h}} w_{\mathbf{h}} (\mathbf{h}\mathbf{h}^T) g_j(\mathbf{h}) g_k(\mathbf{h}) \times \{ \exp[-2\pi i \mathbf{h} \cdot (\mathbf{x}_j - \mathbf{x}_k)] - \exp(2i\varphi_{\mathbf{h}}^{\text{calc}}) \exp[-2\pi i \mathbf{h} \cdot (\mathbf{x}_j + \mathbf{x}_k)] \}$$

(Friedel's symmetry makes \Re redundant on the last line). Cruickshank argued that the first term would give a good approximation to the diagonal blocks of the normal matrix and to those off-diagonal blocks for which \mathbf{x}_j and \mathbf{x}_k are close. On this basis he was able to justify the '*n*-shift rule' of Shoemaker *et al.* (1950). Cruickshank gave this derivation in a general space group, but using a very terse notation which somewhat obscures it. Using the symmetrized trigonometric structure-factor kernel Ξ^- of Section 1.3.4.2.2.9 and its multiplication formula, the above expression is seen to involve the values of a Fourier synthesis at points of the form $\mathbf{x}_j \pm S_g(\mathbf{x}_k)$.

Cruickshank (1956) showed that this analysis could also be applied to the refinement of temperature factors.

These two results made it possible to obtain all coefficients involved in the normal equations by looking up the values of certain differential Fourier syntheses at \mathbf{x}_j or at $\mathbf{x}_j \pm S_g(\mathbf{x}_k)$. At the time this did not confer any superiority over the standard form of the least-squares procedure, because the accurate computation of Fourier syntheses was an expensive operation. The modified Fourier method was used by Truter (1954) and by Ahmed & Cruickshank (1953*a*), and was incorporated into the program system described by Cruickshank *et al.* (1961). A more recent comparison with the least-squares method was made by Dietrich (1972).

There persisted, however, some confusion about the nature of the relationship between Fourier and least-squares methods, caused by the extra factors $g_j(\mathbf{h})$ which make it necessary to compute a differential synthesis for each type of atom. This led Cruickshank to conclude that 'in spite of their remarkable similarities the least-squares and modified-Fourier methods are fundamentally distinct'.

1.3.4.4.7.6. Agarwal's FFT implementation of the Fourier method

Agarwal (1978) rederived and completed Cruickshank's results at a time when the availability of the FFT algorithm made the Fourier method of calculating the coefficients of the normal equations much more economical than the standard method, especially for macromolecules.

As obtained by Cruickshank, the modified Fourier method required a full 3D Fourier synthesis

– for each type of parameter, since this determines [*via* the polynomial $P_p(\mathbf{h})$] the type of differential synthesis to be computed;

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

– for each type of atom $j \in J$, since the coefficients of the differential synthesis must be multiplied by $g_j(\mathbf{h})$.

Agarwal disposed of the latter dependence by pointing out that the multiplication involved is equivalent to a real-space *convolution* between the differential synthesis and $\sigma_j(\mathbf{x})$, the standard electron density ρ_j for atom type j (Section 1.3.4.2.1.2) smeared by the isotropic thermal agitation of that atom. Since σ_j is *localized*, this convolution involves only a small number of grid points. The requirement of a distinct differential synthesis for each parameter type, however, continued to hold, and created some difficulties at the FFT level because the symmetries of differential syntheses are more complex than ordinary space-group symmetries. Jack & Levitt (1978) sought to avoid the calculation of difference syntheses by using instead finite differences calculated from ordinary Fourier or difference Fourier maps.

In spite of its complication, this return to the Fourier implementation of the least-squares method led to spectacular increases in speed (Isaacs & Agarwal, 1978; Agarwal, 1980; Baker & Dodson, 1980) and quickly gained general acceptance (Dodson, 1981; Isaacs, 1982a,b, 1984).

1.3.4.4.7.7. Lifchitz's reformulation

Lifchitz [see Agarwal *et al.* (1981), Agarwal (1981)] proposed that the idea of treating certain multipliers in Cruickshank's modified differential Fourier syntheses by means of a convolution in real space should be applied not only to $g_j(\mathbf{h})$, but also to the polynomials $P_p(\mathbf{h})$ which determine the type of differential synthesis being calculated. This leads to convoluting $\partial\sigma_j/\partial u_p$ with the *same* ordinary weighted difference Fourier synthesis, rather than σ_j with the differential synthesis of type p . In this way, a *single* Fourier synthesis, with *ordinary* (scalar) symmetry properties, needs be computed; the parameter type and atom type both intervene through the function $\partial\sigma_j/\partial u_p$ with which it is convoluted. This approach has been used as the basis of an efficient general-purpose least-squares refinement program for macromolecular structures (Tronrud *et al.*, 1987).

This rearrangement amounts to using the fact (Section 1.3.2.3.9.7) that convolution commutes with differentiation. Let

$$D(\mathbf{x}) = \sum_{\mathbf{h}} w_{\mathbf{h}} (|F_{\mathbf{h}}^{\text{calc}}| - |F_{\mathbf{h}}^{\text{obs}}|) \exp(i\varphi_{\mathbf{h}}^{\text{calc}}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$$

be the inverse-variance weighted difference map, and let us assume that parameter u_p belongs to atom j . Then the Agarwal form for the p th component of the right-hand side of the normal equations is

$$\left(\frac{\partial D}{\partial u_p} * \sigma_j \right) (x_j),$$

while the Lifchitz form is

$$\left(D * \frac{\partial \sigma_j}{\partial u_p} \right) (x_j).$$

1.3.4.4.7.8. A simplified derivation

A very simple derivation of the previous results will now be given, which suggests the possibility of many generalizations.

The weighted difference map $D(\mathbf{x})$ has coefficients $D_{\mathbf{h}}$ which are the gradients of the global residual with respect to each $F_{\mathbf{h}}^{\text{calc}}$:

$$D_{\mathbf{h}} = \frac{\partial R}{\partial A_{\mathbf{h}}^{\text{calc}}} + i \frac{\partial R}{\partial B_{\mathbf{h}}^{\text{calc}}}.$$

By the chain rule, a variation of each $F_{\mathbf{h}}^{\text{calc}}$ by $\delta F_{\mathbf{h}}^{\text{calc}}$ will result in a variation of R by δR with

$$\delta R = \sum_{\mathbf{h}} \left[\frac{\partial R}{\partial A_{\mathbf{h}}^{\text{calc}}} \delta A_{\mathbf{h}}^{\text{calc}} + \frac{\partial R}{\partial B_{\mathbf{h}}^{\text{calc}}} \delta B_{\mathbf{h}}^{\text{calc}} \right] = \mathcal{R}e \sum_{\mathbf{h}} [D_{\mathbf{h}} \delta F_{\mathbf{h}}^{\text{calc}}].$$

The $\mathcal{R}e$ operation is superfluous because of Friedel symmetry, so that δR may be simply written in terms of the Hermitian scalar product in $\ell^2(\mathbb{Z}^3)$:

$$\delta R = (\mathbf{D}, \delta \mathbf{F}^{\text{calc}}).$$

If ρ^{calc} is the transform of $\delta \mathbf{F}^{\text{calc}}$, we have also by Parseval's theorem

$$\delta R = (D, \delta \rho^{\text{calc}}).$$

We may therefore write

$$D(\mathbf{x}) = \frac{\partial R}{\partial \rho^{\text{calc}}(\mathbf{x})},$$

which states that $D(\mathbf{x})$ is the functional derivative of R with respect to ρ^{calc} .

The right-hand side of the normal equations has $\partial R/\partial u_p$ for its p th element, and this may be written

$$\frac{\partial R}{\partial u_p} = \int_{\mathbb{R}^3/\mathbb{Z}^3} \frac{\partial R}{\partial \rho^{\text{calc}}(\mathbf{x})} \frac{\partial \rho^{\text{calc}}(\mathbf{x})}{\partial u_p} d^2\mathbf{x} = \left(D, \frac{\partial \rho^{\text{calc}}}{\partial u_p} \right).$$

If u_p belongs to atom j , then

$$\frac{\partial \rho^{\text{calc}}}{\partial u_p} = \frac{\partial(\tau_{\mathbf{x}_j} \sigma_j)}{\partial u_p} = \tau_{\mathbf{x}_j} \left(\frac{\partial \sigma_j}{\partial u_p} \right);$$

hence

$$\frac{\partial R}{\partial u_p} = \left(D, \tau_{\mathbf{x}_j} \left(\frac{\partial \sigma_j}{\partial u_p} \right) \right).$$

By the identity of Section 1.3.2.4.3.5, this is identical to Lifchitz's expression $(D * \partial\sigma_j/\partial u_p)(\mathbf{x}_j)$. The present derivation in terms of scalar products [see Brünger (1989) for another presentation of it] is conceptually simpler, since it invokes only the chain rule [other uses of which have been reviewed by Lunin (1985)] and Parseval's theorem; economy of computation is obviously related to the good localization of $\partial\rho^{\text{calc}}/\partial u_p$ compared to $\partial F^{\text{calc}}/\partial u_p$. Convolutions, whose meaning is less clear, are no longer involved; they were a legacy of having first gone over to reciprocal space *via* differential syntheses in the 1940s.

Cast in this form, the calculation of derivatives by FFT methods appears as a particular instance of the procedure described in connection with variational techniques (Section 1.3.4.4.6) to calculate the coefficients of local quadratic models in a search subspace; this is far from surprising since varying the electron density through a variation of the parameters of an atomic model is a particular case of the 'free' variations considered by the variational approach. The latter procedure would accommodate in a very natural fashion the joint consideration of an energetic (Jack & Levitt, 1978; Brünger *et al.*, 1987; Brünger, 1988; Brünger *et al.*, 1989; Kuriyan *et al.*, 1989) or stereochemical (Konnert, 1976; Sussman *et al.*, 1977; Konnert & Hendrickson, 1980; Hendrickson & Konnert, 1980; Tronrud *et al.*, 1987) restraint function (which would play the role of S) and of the crystallographic residual (which would be C). It would even have over the latter the superiority of affording a genuine *second-order* approximation, albeit only in a subspace, hence the ability of detecting negative curvature and the resulting bifurcation behaviour (Bricogne, 1984). Current methods are unable to do this because they use only *first-order* models, and this is known to degrade severely the overall efficiency of the refinement process.

1. GENERAL RELATIONSHIPS AND TECHNIQUES

1.3.4.4.7.9. Discussion of macromolecular refinement techniques

The impossibility of carrying out a full-matrix least-squares refinement of a macromolecular crystal structure, caused by excessive computational cost and by the paucity of observations, led Diamond (1971) to propose a real-space refinement method in which stereochemical knowledge was used to keep the number of free parameters to a minimum. Refinement took place by a least-squares fit between the ‘observed’ electron-density map and a model density consisting of Gaussian atoms. This procedure, coupled to iterative recalculation of the phases, led to the first highly refined protein structures obtained without using full-matrix least squares (Huber *et al.*, 1974; Bode & Schwager, 1975; Deisenhofer & Steigemann, 1975; Takano, 1977*a,b*).

Real-space refinement takes advantage of the localization of atoms (each parameter interacts only with the density near the atom to which it belongs) and gives the most immediate description of stereochemical constraints. A disadvantage is that fitting the ‘observed’ electron density amounts to treating the phases of the structure factors as observed quantities, and to ignoring the experimental error estimates on their moduli. The method is also much more vulnerable to series-termination errors and accidentally missing data than the least-squares method. These objections led to the progressive disuse of Diamond’s method, and to a switch towards reciprocal-space least squares following Agarwal’s work.

The connection established above between the Cruickshank–Agarwal modified Fourier method and the simple use of the chain rule affords a partial refutation to both the premises of Diamond’s method and to the objections made against it:

(i) it shows that refinement *can* be performed through localized computations in real space *without* having to treat the phases as observed quantities;

(ii) at the same time, it shows that measurement errors on the moduli *can* be fully utilized in real space, *via* the Fourier synthesis of the functional derivative $\partial R/\partial \rho^{\text{calc}}(\mathbf{x})$ or by means of the coefficients of a quadratic model of R in a search subspace.

1.3.4.4.7.10. Sampling considerations

The calculation of the inner products $(D, \partial \rho^{\text{calc}}/\partial u_p)$ from a sampled gradient map D requires even more caution than that of structure factors *via* electron-density maps described in Section 1.3.4.4.5, because the functions $\partial \sigma_j/\partial u_p$ have transforms which extend even further in reciprocal space than the σ_j themselves. Analytically, if the σ_j are Gaussians, the $\partial \sigma_j/\partial u_p$ are finite sums of multivariate Hermite functions (Section 1.3.2.4.4.2) and hence the same is true of their transforms. The difference map D must therefore be finely sampled and the relation between error and sampling rate may be investigated as in Section 1.3.4.4.5. An examination of the sampling rates commonly used (*e.g.* one third of the resolution) shows that they are insufficient. Tronrud *et al.* (1987) propose to relax this requirement by applying an artificial temperature factor to σ_j (*cf.* Section 1.3.4.4.5) and the negative of that temperature factor to D , a procedure of questionable validity because the latter ‘sharpening’ operation is ill defined [the function $\exp(-\|\mathbf{x}\|^2)$ does not define a tempered distribution, so the associativity properties of convolution may be lost]. A more robust procedure would be to compute the scalar product by means of a more sophisticated numerical quadrature formula than a mere grid sum.

1.3.4.4.8. Miscellaneous correlation functions

Certain correlation functions can be useful to detect the presence of multiple copies of the same molecule (known or unknown) in the asymmetric unit of a crystal of unknown structure.

Suppose that a crystal contains one or several copies of a molecule \mathcal{M} in its asymmetric unit. If $\mu(\mathbf{x})$ is the electron density of that molecule in some reference position and orientation, then

$$\rho^0 = \sum_{j \in J} \left[\sum_{g \in G} S_g^\#(T_j^\# \mu) \right],$$

where $T_j : \mathbf{x} \mapsto \mathbf{C}_j \mathbf{x} + \mathbf{d}_j$ describes the placement of the j th copy of the molecule with respect to the reference copy. It is assumed that each such copy is in a general position, so that there is no isotropy subgroup.

The methods of Section 1.3.4.2.2.9 (with ρ_j replaced by $C_j^\# \mu$, and \mathbf{x}_j by \mathbf{d}_j) lead to the following expression for the auto-correlation of ρ^0 :

$$\check{\rho}^0 * \rho^0 = \sum_{j_1} \sum_{j_2} \sum_{g_1} \sum_{g_2} \tau_{S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})} \times [(R_{g_1}^\# C_{j_1}^\# \check{\mu}) * (R_{g_2}^\# C_{j_2}^\# \mu)].$$

If μ is unknown, consider the subfamily σ of terms with $j_1 = j_2 = j$ and $g_1 = g_2 = g$:

$$\sigma = \sum_j \sum_g R_g^\# C_j^\# (\check{\mu} * \mu).$$

The scalar product $(\sigma, R^\# \sigma)$ in which R is a variable rotation will have a peak whenever

$$R = (R_{g_1} C_{j_1})^{-1} (R_{g_2} C_{j_2})$$

since two copies of the ‘self-Patterson’ $\check{\mu} * \mu$ of the molecule will be brought into coincidence. If the interference from terms in the Patterson $\pi = r * \check{\rho}^0 * \rho^0$ other than those present in σ is not too serious, the ‘self-rotation function’ $(\pi, R^\# \pi)$ (Rossmann & Blow, 1962; Crowther, 1972) will show the same peaks, from which the rotations $\{C_j\}_{j \in J}$ may be determined, either individually or jointly if for instance they form a group.

If μ is known, then its self-Patterson $\check{\mu} * \mu$ may be calculated, and the C_j may be found by examining the ‘cross-rotation function’ $[\pi, R^\#(\check{\mu} * \mu)]$ which will have peaks at $R = R_g C_j, g \in G, j \in J$. Once the C_j are known, then the various copies $C_j^\# \mu$ of \mathcal{M} may be Fourier-analysed into structure factors:

$$M_j(\mathbf{h}) = \mathcal{F}[C_j^\# \mu](\mathbf{h}).$$

The cross terms with $j_1 \neq j_2, g_1 \neq g_2$ in $\check{\rho}^0 * \rho^0$ then contain ‘motifs’

$$(R_{g_1}^\# C_{j_1}^\# \check{\mu}) * (R_{g_2}^\# C_{j_2}^\# \mu),$$

with Fourier coefficients

$$\overline{M_{j_1}(\mathbf{R}_{g_1}^T \mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T \mathbf{h}),$$

translated by $S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$. Therefore the ‘translation functions’ (Crowther & Blow, 1967)

$$\mathcal{F}_{j_1 g_1, j_2 g_2}(\mathbf{s}) = \sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \overline{M_{j_1}(\mathbf{R}_{g_1}^T \mathbf{h})} \times M_{j_2}(\mathbf{R}_{g_2}^T \mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{s})$$

will have peaks at $\mathbf{s} = S_{g_2}(\mathbf{d}_{j_2}) - S_{g_1}(\mathbf{d}_{j_1})$ corresponding to the detection of these motifs.

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

1.3.4.5. Related applications

1.3.4.5.1. Helical diffraction

The theory of diffraction by helical structures (Cochran *et al.*, 1952; Klug *et al.*, 1958) has played an important part in the study of polypeptides, of nucleic acids and of tobacco mosaic virus.

1.3.4.5.1.1. Circular harmonic expansions in polar coordinates

Let $f = f(x, y)$ be a reasonably regular function in two-dimensional real space. Going over to polar coordinates

$$x = r \cos \varphi \quad y = r \sin \varphi$$

and writing, by slight misuse of notation, $f(r, \varphi)$ for $f(r \cos \varphi, r \sin \varphi)$ we may use the periodicity of f with respect to φ to expand it as a Fourier series (Byerly, 1893):

$$f(r, \varphi) = \sum_{n \in \mathbb{Z}} f_n(r) \exp(in\varphi)$$

with

$$f_n(r) = \frac{1}{2\pi} \int_0^{2\pi} f(r, \varphi) \exp(-in\varphi) d\varphi.$$

Similarly, in reciprocal space, if $F = F(\xi, \eta)$ and if

$$\xi = R \cos \psi \quad \eta = R \sin \psi$$

then

$$F(R, \psi) = \sum_{n \in \mathbb{Z}} i^n F_n(R) \exp(in\psi)$$

with

$$F_n(R) = \frac{1}{2\pi i^n} \int_0^{2\pi} F(R, \psi) \exp(-in\psi) d\psi,$$

where the phase factor i^n has been introduced for convenience in the forthcoming step.

1.3.4.5.1.2. The Fourier transform in polar coordinates

The Fourier transform relation between f and F may then be written in terms of f_n 's and F_n 's. Observing that $\xi x + \eta y = Rr \cos(\varphi - \psi)$, and that (Watson, 1944)

$$\int_0^{2\pi} \exp(iX \cos \theta + in\theta) d\theta = 2\pi i^n J_n(X),$$

we obtain:

$$\begin{aligned} F(R, \psi) &= \int_0^{2\pi} \int_0^{\infty} \left[\sum_{n \in \mathbb{Z}} f_n(r) \exp(in\varphi) \right] \\ &\quad \times \exp[2\pi i R r \cos(\varphi - \psi)] r dr d\varphi \\ &= \sum_{n \in \mathbb{Z}} i^n \left[\int_0^{\infty} f_n(r) J_n(2\pi R r) 2\pi r dr \right] \exp(in\psi); \end{aligned}$$

hence, by the uniqueness of the Fourier expansion of F :

$$F_n(R) = \int_0^{\infty} f_n(r) J_n(2\pi R r) 2\pi r dr.$$

The inverse Fourier relationship leads to

$$f_n(r) = \int_0^{\infty} F_n(R) J_n(2\pi R r) 2\pi R dR.$$

The integral transform involved in the previous two equations is

called the *Hankel transform* (see *e.g.* Titchmarsh, 1922; Sneddon, 1972) of order n .

1.3.4.5.1.3. The transform of an axially periodic fibre

Let ρ be the electron-density distribution in a fibre, which is assumed to have translational periodicity with period 1 along z , and to have compact support with respect to the (x, y) coordinates. Thus ρ may be written

$$\rho = \left[\delta_x \otimes \delta_y \otimes \left(\sum_{k \in \mathbb{Z}} \delta_{(k)} \right)_z \right] * \rho^0,$$

where $\rho^0 = \rho^0(x, y, z)$ is the motif.

By the tensor product property, the inverse Fourier transform $F = \mathcal{F}_{xyz}[\rho]$ may be written

$$F = \left[1_\xi \otimes 1_\eta \otimes \left(\sum_{l \in \mathbb{Z}} \delta_{(l)} \right)_\zeta \right] \times \mathcal{F}[\rho^0]$$

and hence consists of 'layers' labelled by l :

$$F = \sum_{l \in \mathbb{Z}} F(\xi, \eta, l) (\delta_{(l)})_\zeta$$

with

$$F(\xi, \eta, l) = \int_0^1 \mathcal{F}_{xy}[\rho^0](\xi, \eta, z) \exp(2\pi i l z) dz.$$

Changing to polar coordinates in the (x, y) and (ξ, η) planes decomposes the calculation of F from ρ into the following steps:

$$g_{nl}(r) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^1 \rho(r, \varphi, z) \exp[i(-n\varphi + 2\pi l z)] d\varphi dz$$

$$G_{nl}(R) = \int_0^{\infty} g_{nl}(r) J_n(2\pi R r) 2\pi r dr$$

$$F(R, \psi, l) = \sum_{n \in \mathbb{Z}} i^n G_{nl}(R) \exp(in\psi)$$

and the calculation of ρ from F into:

$$G_{nl}(R) = \frac{1}{2\pi i^n} \int_0^{2\pi} F(R, \psi, l) \exp(-in\psi) d\psi$$

$$g_{nl}(r) = \int_0^{\infty} G_{nl}(R) J_n(2\pi R r) 2\pi R dR$$

$$\rho(r, \varphi, z) = \sum_{n \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} g_{nl}(r) \exp[i(n\varphi - 2\pi l z)].$$

These formulae are seen to involve a 2D Fourier series with respect to the two periodic coordinates φ and z , and Hankel transforms along the radial coordinates. The two periodicities in φ and z are independent, so that all combinations of indices (n, l) occur in the Fourier summations.

1.3.4.5.1.4. Helical symmetry and associated selection rules

Helical symmetry involves a 'clutching' between the two (hitherto independent) periodicities in φ (period 2π) and z (period 1) which causes a subdivision of the period lattice and hence a decimation (governed by 'selection rules') of the Fourier coefficients.

Let \mathbf{i} and \mathbf{j} be the basis vectors along $\varphi/2\pi$ and z . The integer lattice with basis (\mathbf{i}, \mathbf{j}) is a period lattice for the (φ, z) dependence of the electron density ρ of an axially periodic fibre considered in Section 1.3.4.5.1.3:

1. GENERAL RELATIONSHIPS AND TECHNIQUES

$$\rho(r, \varphi + 2\pi k_1, z + k_2) = \rho(r, \varphi, z).$$

Suppose the fibre now has *helical* symmetry, with u copies of the same molecule in t turns, where $\text{g.c.d.}(u, t) = 1$. Using the Euclidean algorithm, write $u = \lambda t + \mu$ with λ and μ positive integers and $\mu < t$. The period lattice for the (φ, z) dependence of ρ may be defined in terms of the new basis vectors:

I, joining subunit 0 to subunit l in the same turn;

J, joining subunit 0 to subunit λ after wrapping around.

In terms of the original basis

$$\mathbf{I} = \frac{t}{u} \mathbf{i} + \frac{1}{u} \mathbf{j}, \quad \mathbf{J} = \frac{-\mu}{u} \mathbf{i} + \frac{\lambda}{u} \mathbf{j}.$$

If α and β are coordinates along **I** and **J**, respectively,

$$\begin{pmatrix} \varphi/2\pi \\ z \end{pmatrix} = \frac{1}{u} \begin{pmatrix} t & -\mu \\ 1 & \lambda \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

or equivalently

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \lambda & \mu \\ -1 & t \end{pmatrix} \begin{pmatrix} \varphi/2\pi \\ z \end{pmatrix}.$$

By Fourier transformation,

$$\begin{aligned} \left(\frac{\varphi}{2\pi}, z\right) &\Leftrightarrow (-n, l) \\ (\alpha, \beta) &\Leftrightarrow (m, p) \end{aligned}$$

with the transformations between indices given by the contragredients of those between coordinates, *i.e.*

$$\begin{pmatrix} n \\ l \end{pmatrix} = \begin{pmatrix} -\lambda & 1 \\ -\mu & t \end{pmatrix} \begin{pmatrix} m \\ p \end{pmatrix}$$

and

$$\begin{pmatrix} m \\ p \end{pmatrix} = \frac{1}{u} \begin{pmatrix} -t & 1 \\ \mu & \lambda \end{pmatrix} \begin{pmatrix} n \\ l \end{pmatrix}.$$

It follows that

$$l = tn + um,$$

or alternatively that

$$\mu n = up - \lambda l,$$

which are two equivalent expressions of the *selection rules* describing the decimation of the transform. These rules imply that only certain orders n contribute to a given layer l .

The 2D Fourier analysis may now be performed by analysing a *single* subunit referred to coordinates α and β to obtain

$$h_{m,p}(r) = \int_0^1 \int_0^1 \rho(r, \alpha, \beta) \exp[2\pi i(m\alpha + p\beta)] \, d\alpha \, d\beta$$

and then reindexing to get only the allowed g_{nl} 's by

$$g_{nl}(r) = uh_{-\lambda m+p, \mu m+tp}(r).$$

This is u times faster than analysing u subunits with respect to the (φ, z) coordinates.

1.3.4.5.2. Application to probability theory and direct methods

The Fourier transformation plays a central role in the branch of probability theory concerned with the limiting behaviour of sums of large numbers of independent and identically distributed random variables or random vectors. This privileged role is a consequence of the convolution theorem and of the 'moment-generating'

properties which follow from the exchange between differentiation and multiplication by monomials. When the limit theorems are applied to the calculation of joint probability distributions of structure factors, which are themselves closely related to the Fourier transformation, a remarkable phenomenon occurs, which leads to the saddlepoint approximation and to the maximum-entropy method.

1.3.4.5.2.1. Analytical methods of probability theory

The material in this section is not intended as an introduction to probability theory [for which the reader is referred to Cramér (1946), Petrov (1975) or Bhattacharya & Rao (1976)], but only as an illustration of the role played by the Fourier transformation in certain specific areas which are used in formulating and implementing direct methods of phase determination.

(a) Convolution of probability densities

The addition of independent random variables or vectors leads to the convolution of their probability distributions: if \mathbf{X}_1 and \mathbf{X}_2 are two n -dimensional random vectors independently distributed with probability densities P_1 and P_2 , respectively, then their sum $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$ has probability density \mathcal{P} given by

$$\begin{aligned} \mathcal{P}(\mathbf{X}) &= \int_{\mathbf{R}^n} P_1(\mathbf{X}_1) P_2(\mathbf{X} - \mathbf{X}_1) \, d^n \mathbf{X}_1 \\ &= \int_{\mathbf{R}^n} P_1(\mathbf{X} - \mathbf{X}_2) P_2(\mathbf{X}_2) \, d^n \mathbf{X}_2 \end{aligned}$$

i.e.

$$\mathcal{P} = P_1 * P_2.$$

This result can be extended to the case where P_1 and P_2 are singular measures (distributions of order zero, Section 1.3.2.3.4) and do not have a density with respect to the Lebesgue measure in \mathbb{R}^n .

(b) Characteristic functions

This convolution can be turned into a simple multiplication by considering the Fourier transforms (called the *characteristic functions*) of P_1 , P_2 and \mathcal{P} , defined with a slightly different normalization in that there is no factor of 2π in the exponent (see Section 1.3.2.4.5), *e.g.*

$$C(\mathbf{t}) = \int_{\mathbf{R}^n} P(\mathbf{X}) \exp(i\mathbf{t} \cdot \mathbf{X}) \, d^n \mathbf{X}.$$

Then by the convolution theorem

$$\mathcal{C}(\mathbf{t}) = C_1(\mathbf{t}) \times C_2(\mathbf{t}),$$

so that $\mathcal{P}(\mathbf{X})$ may be evaluated by Fourier inversion of its characteristic function as

$$\mathcal{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} C_1(\mathbf{t}) C_2(\mathbf{t}) \exp(-i\mathbf{t} \cdot \mathbf{X}) \, d^n \mathbf{t}$$

(see Section 1.3.2.4.5 for the normalization factors).

It follows from the differentiation theorem that the partial derivatives of the characteristic function $C(\mathbf{t})$ at $\mathbf{t} = \mathbf{0}$ are related to the moments of a distribution P by the identities

$$\begin{aligned} \mu_{r_1 r_2 \dots r_n} &\equiv \int_D P(\mathbf{X}) X_1^{r_1} X_2^{r_2} \dots X_n^{r_n} \, d^n \mathbf{X} \\ &= i^{-(r_1 + \dots + r_n)} \left. \frac{\partial^{r_1 + \dots + r_n} C}{\partial t_1^{r_1} \dots \partial t_n^{r_n}} \right|_{\mathbf{t}=\mathbf{0}} \end{aligned}$$

for any n -tuple of non-negative integers (r_1, r_2, \dots, r_n) .

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

(c) Moment-generating functions

The above relation can be freed from powers of i by defining (at least formally) the *moment-generating function*:

$$M(\mathbf{t}) = \int_{\mathbb{R}^n} P(\mathbf{X}) \exp(\mathbf{t} \cdot \mathbf{X}) d^n \mathbf{X}$$

which is related to $C(\mathbf{t})$ by $C(\mathbf{t}) = M(i\mathbf{t})$ so that the inversion formula reads

$$\mathcal{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} M_1(i\mathbf{t}) M_2(i\mathbf{t}) \exp(-i\mathbf{t} \cdot \mathbf{X}) d^n \mathbf{t}.$$

The moment-generating function is well defined, in particular, for any probability distribution with compact support, in which case it may be continued analytically from a function over \mathbb{R}^n into an entire function of n complex variables by virtue of the Paley–Wiener theorem (Section 1.3.2.4.2.10). Its moment-generating properties are summed up in the following relations:

$$\mu_{r_1 r_2 \dots r_n} = \left. \frac{\partial^{r_1 + \dots + r_n} M}{\partial t_1^{r_1} \dots \partial t_n^{r_n}} \right|_{\mathbf{t}=\mathbf{0}}.$$

(d) Cumulant-generating functions

The multiplication of moment-generating functions may be further simplified into the addition of their logarithms:

$$\log \mathcal{M} = \log M_1 + \log M_2,$$

or equivalently of the coefficients of their Taylor series at $\mathbf{t} = \mathbf{0}$, viz:

$$\kappa_{r_1 r_2 \dots r_n} = \left. \frac{\partial^{r_1 + \dots + r_n} (\log M)}{\partial t_1^{r_1} \dots \partial t_n^{r_n}} \right|_{\mathbf{t}=\mathbf{0}}.$$

These coefficients are called *cumulants*, since they add when the independent random vectors to which they belong are added, and $\log M$ is called the *cumulant-generating function*. The inversion formula for \mathcal{P} then reads

$$\mathcal{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[\log M_1(i\mathbf{t}) + \log M_2(i\mathbf{t}) - i\mathbf{t} \cdot \mathbf{X}] d^n \mathbf{t}.$$

(e) Asymptotic expansions and limit theorems

Consider an n -dimensional random vector \mathbf{X} of the form

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_N,$$

where the N summands are independent n -dimensional random vectors identically distributed with probability density P . Then the distribution \mathcal{P} of \mathbf{X} may be written in closed form as a Fourier transform:

$$\begin{aligned} \mathcal{P}(\mathbf{X}) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} M^N(i\mathbf{t}) \exp(-i\mathbf{t} \cdot \mathbf{X}) d^n \mathbf{t} \\ &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[N \log M(i\mathbf{t}) - i\mathbf{t} \cdot \mathbf{X}] d^n \mathbf{t}, \end{aligned}$$

where

$$M(\mathbf{t}) = \int_{\mathbb{R}^n} P(\mathbf{Y}) \exp(\mathbf{t} \cdot \mathbf{Y}) d^n \mathbf{Y}$$

is the moment-generating function common to all the summands.

This an exact expression for \mathcal{P} , which may be exploited analytically or numerically in certain favourable cases. Supposing for instance that P has compact support, then its characteristic function $M(i\mathbf{t})$ can be sampled finely enough to accommodate the bandwidth of the support of $\mathcal{P} = P^{*N}$ (this sampling rate clearly depends on n) so that the above expression for \mathcal{P} can be used for its

numerical evaluation as the *discrete* Fourier transform of $M^N(i\mathbf{t})$. This exact method is practical only for small values of the dimension n .

In all other cases some form of approximation must be used in the Fourier inversion of $M^N(i\mathbf{t})$. For this purpose it is customary (Cramér, 1946) to expand the cumulant-generating function around $\mathbf{t} = \mathbf{0}$ with respect to the carrying variables \mathbf{t} :

$$\log[M^N(i\mathbf{t})] = \sum_{\mathbf{r} \in \mathbb{N}^n} \frac{N \kappa_{\mathbf{r}}}{\mathbf{r}!} (i\mathbf{t})^{\mathbf{r}},$$

where $\mathbf{r} = (r_1, r_2, \dots, r_n)$ is a multi-index (Section 1.3.2.2.3). The first-order terms may be eliminated by recentring \mathcal{P} around its vector of first-order cumulants

$$\langle \mathbf{X} \rangle = \sum_{j=1}^N \langle \mathbf{X}_j \rangle,$$

where $\langle \cdot \rangle$ denotes the mathematical expectation of a random vector. The second-order terms may be grouped separately from the terms of third or higher order to give

$$\begin{aligned} M^N(i\mathbf{t}) &= \exp(-\frac{1}{2} N \mathbf{t}^T \mathbf{Q} \mathbf{t}) \\ &\times \exp \left\{ \sum_{|\mathbf{r}| \geq 3} \frac{N \kappa_{\mathbf{r}}}{\mathbf{r}!} (i\mathbf{t})^{\mathbf{r}} \right\}, \end{aligned}$$

where $\mathbf{Q} = \nabla \nabla^T (\log M)$ is the covariance matrix of the multivariate distribution P . Expanding the exponential gives rise to a series of terms of the form

$$\exp(-\frac{1}{2} N \mathbf{t}^T \mathbf{Q} \mathbf{t}) \times \text{monomial in } t_1, t_2, \dots, t_n,$$

each of which may now be subjected to a Fourier transformation to yield a Hermite function of \mathbf{t} (Section 1.3.2.4.4.2) with coefficients involving the cumulants κ of P . Taking the transformed terms in natural order gives an asymptotic expansion of P for large N called the *Gram–Charlier series* of \mathcal{P} , while grouping the terms according to increasing powers of $1/\sqrt{N}$ gives another asymptotic expansion called the *Edgeworth series* of \mathcal{P} . Both expansions comprise a leading Gaussian term which embodies the *central-limit theorem*:

$$\mathcal{P}(\mathbf{E}) = \frac{1}{\sqrt{\det(2\pi\mathbf{Q})}} \exp(-\frac{1}{2} \mathbf{E}^T \mathbf{Q}^{-1} \mathbf{E}), \quad \text{where } \mathbf{E} = \frac{\mathbf{X} - \langle \mathbf{X} \rangle}{\sqrt{N}}.$$

(f) The saddlepoint approximation

A limitation of the Edgeworth series is that it gives an accurate estimate of $\mathcal{P}(\mathbf{X})$ only in the vicinity of $\mathbf{X} = \langle \mathbf{X} \rangle$, *i.e.* for small values of \mathbf{E} . These convergence difficulties are easily understood: one is substituting a *local* approximation to $\log M$ (*viz* a Taylor-series expansion valid near $\mathbf{t} = \mathbf{0}$) into an integral, whereas integration is a *global* process which consults values of $\log M$ far from $\mathbf{t} = \mathbf{0}$.

It is possible, however, to let the point \mathbf{t} where $\log M$ is expanded as a Taylor series depend on the particular value \mathbf{X}^* of \mathbf{X} for which an accurate evaluation of $\mathcal{P}(\mathbf{X})$ is desired. This is the essence of the *saddlepoint method* (Fowler, 1936; Khinchin 1949; Daniels, 1954; de Bruijn, 1970; Bleistein & Handelsman, 1986), which uses an analytical continuation of $M(\mathbf{t})$ from a function over \mathbb{R}^n to a function over \mathbb{C}^n (see Section 1.3.2.4.2.10). Putting then $\mathbf{t} = \mathbf{s} - i\tau$, the \mathbb{C}^n version of Cauchy's theorem (Hörmander, 1973) gives rise to the identity

1. GENERAL RELATIONSHIPS AND TECHNIQUES

$$\mathcal{P}(\mathbf{X}^*) = \frac{\exp(-\boldsymbol{\tau} \cdot \mathbf{X}^*)}{(2\pi)^n} \times \int_{\mathbb{R}^n} \exp\left\{N \left[\log M(\boldsymbol{\tau} + i\mathbf{s}) - i\mathbf{s} \cdot \frac{\mathbf{X}^*}{N} \right]\right\} d^n \mathbf{s}$$

for any $\boldsymbol{\tau} \in \mathbb{R}^n$. By a convexity argument involving the positive-definiteness of covariance matrix \mathbf{Q} , there is a unique value of $\boldsymbol{\tau}$ such that

$$\nabla(\log M)|_{\mathbf{t}=\mathbf{0}-i\boldsymbol{\tau}} = \frac{\mathbf{X}^*}{N}.$$

At the *saddlepoint* $\mathbf{t}^* = \mathbf{0} - i\boldsymbol{\tau}$, the modulus of the integrand above is a maximum and its phase is stationary with respect to the integration variable \mathbf{s} : as N tends to infinity, all contributions to the integral cancel because of rapid oscillation, except those coming from the immediate vicinity of \mathbf{t}^* where there is no oscillation. A Taylor expansion of $\log M^N$ to second order with respect to \mathbf{s} at \mathbf{t}^* then gives

$$\log M^N(\boldsymbol{\tau} + i\mathbf{s}) \approx \log M^N(\boldsymbol{\tau}) + i\mathbf{s} \cdot \mathbf{X}^* - \frac{N}{2} [\mathbf{s}^T \mathbf{Q} \mathbf{s}]$$

and hence

$$\mathcal{P}(\mathbf{X}^*) \approx \exp[\log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau} \cdot \mathbf{X}^*] \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp(-\frac{1}{2} \mathbf{s}^T \mathbf{Q} \mathbf{s}) d^n \mathbf{s}.$$

The last integral is elementary and gives the ‘saddlepoint approximation’:

$$\mathcal{P}^{\text{SP}}(\mathbf{X}^*) = \frac{\exp(\mathbf{S})}{\sqrt{\det(2\pi\mathcal{Q})}},$$

where

$$\mathbf{S} = \log M^N(\boldsymbol{\tau}) - \boldsymbol{\tau} \cdot \mathbf{X}^*$$

and where

$$\mathcal{Q} = \nabla \nabla^T (\log M^N) = N\mathbf{Q}.$$

This approximation scheme amounts to using the ‘conjugate distribution’ (Khinchin, 1949)

$$P_{\boldsymbol{\tau}}(\mathbf{X}_j) = P(\mathbf{X}_j) \frac{\exp(\boldsymbol{\tau} \cdot \mathbf{X}_j)}{M(\boldsymbol{\tau})}$$

instead of the original distribution $P(\mathbf{X}_j) = P_0(\mathbf{X}_j)$ for the common distribution of all N random vectors \mathbf{X}_j . The exponential modulation results from the analytic continuation of the characteristic (or moment-generating) function into \mathbb{C}^n , as in Section 1.3.2.4.2.10. The saddlepoint approximation \mathcal{P}^{SP} is only the leading term of an asymptotic expansion (called the *saddlepoint expansion*) for \mathcal{P} , which is actually the Edgeworth expansion associated with $P_{\boldsymbol{\tau}}^*$.

1.3.4.5.2.2. The statistical theory of phase determination

The methods of probability theory just surveyed were applied to various problems formally similar to the crystallographic phase problem [e.g. the ‘problem of the random walk’ of Pearson (1905)] by Rayleigh (1880, 1899, 1905, 1918, 1919) and Kluyver (1906). They became the basis of the statistical theory of communication with the classic papers of Rice (1944, 1945).

The Gram–Charlier and Edgeworth series were introduced into crystallography by Bertaut (1955*a,b,c*, 1956*a*) and by Klug (1958), respectively, who showed them to constitute the mathematical basis of numerous formulae derived by Hauptman & Karle (1953). The saddlepoint approximation was introduced by Bricogne (1984) and was shown to be related to variational methods involving the

maximization of certain entropy criteria. This connection exhibits most of the properties of the Fourier transform at play simultaneously, and will now be described as a final illustration.

(a) Definitions and conventions

Let H be a set of unique non-origin reflections \mathbf{h} for a crystal with lattice Λ and space group G . Let H contain n_a acentric and n_c centric reflections. Structure-factor values attached to all reflections in H will comprise $n = 2n_a + n_c$ real numbers. For \mathbf{h} acentric, $\alpha_{\mathbf{h}}$ and $\beta_{\mathbf{h}}$ will be the real and imaginary parts of the complex structure factor; for \mathbf{h} centric, $\gamma_{\mathbf{h}}$ will be the real coordinate of the (possibly complex) structure factor measured along a real axis rotated by one of the two angles $\theta_{\mathbf{h}}, \pi$ apart, to which the phase is restricted modulo 2π (Section 1.3.4.2.2.5). These n real coordinates will be arranged as a column vector containing the acentric then the centric data, *i.e.* in the order

$$\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \alpha_{n_a}, \beta_{n_a}, \gamma_1, \gamma_2, \dots, \gamma_{n_c}.$$

(b) Vectors of trigonometric structure-factor expressions

Let $\boldsymbol{\xi}(\mathbf{x})$ denote the vector of trigonometric structure-factor expressions associated with $\mathbf{x} \in D$, where D denotes the asymmetric unit. These are defined as follows:

$$\begin{aligned} \alpha_{\mathbf{h}}(\mathbf{x}) + i\beta_{\mathbf{h}}(\mathbf{x}) &= \Xi(\mathbf{h}, \mathbf{x}) \quad \text{for } \mathbf{h} \text{ acentric} \\ \gamma_{\mathbf{h}}(\mathbf{x}) &= \exp(-i\theta_{\mathbf{h}}) \Xi(\mathbf{h}, \mathbf{x}) \quad \text{for } \mathbf{h} \text{ centric,} \end{aligned}$$

where

$$\Xi(\mathbf{h}, \mathbf{x}) = \frac{1}{|G_{\mathbf{x}}|} \sum_{g \in G} \exp\{2\pi i \mathbf{h} \cdot [S_g(\mathbf{x})]\}.$$

According to the convention above, the coordinates of $\boldsymbol{\xi}(\mathbf{x})$ in \mathbb{R}^n will be arranged in a column vector as follows:

$$\begin{aligned} \xi_{2r-1}(\mathbf{x}) &= \alpha_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = 1, \dots, n_a, \\ \xi_{2r}(\mathbf{x}) &= \beta_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = 1, \dots, n_a, \\ \xi_{n_a+r}(\mathbf{x}) &= \gamma_{\mathbf{h}_r}(\mathbf{x}) \quad \text{for } r = n_a + 1, \dots, n_a + n_c. \end{aligned}$$

(c) Distributions of random atoms and moment-generating functions

Let position \mathbf{x} in D now become a random vector with probability density $m(\mathbf{x})$. Then $\boldsymbol{\xi}(\mathbf{x})$ becomes itself a random vector in \mathbb{R}^n , whose distribution $p(\boldsymbol{\xi})$ is the image of distribution $m(\mathbf{x})$ through the mapping $\mathbf{x} \rightarrow \boldsymbol{\xi}(\mathbf{x})$ just defined. The locus of $\boldsymbol{\xi}(\mathbf{x})$ in \mathbb{R}^n is a compact algebraic manifold \mathcal{L} (the multidimensional analogue of a Lissajous curve), so that p is a singular measure (a distribution of order 0, Section 1.3.2.3.4, concentrated on that manifold) with compact support. The average with respect to p of any function Ω over \mathbb{R}^n which is infinitely differentiable in a neighbourhood of \mathcal{L} may be calculated as an average with respect to m over D by the ‘induction formula’:

$$\langle p, \Omega \rangle = \int_D m(\mathbf{x}) \Omega[\boldsymbol{\xi}(\mathbf{x})] d^3 \mathbf{x}.$$

In particular, one can calculate the moment-generating function M for distribution p as

$$M(\mathbf{t}) \equiv \langle p_{\boldsymbol{\xi}}, \exp(\mathbf{t} \cdot \boldsymbol{\xi}) \rangle = \int_D m(\mathbf{x}) \exp[\mathbf{t} \cdot \boldsymbol{\xi}(\mathbf{x})] d^3 \mathbf{x}$$

and hence calculate the moments μ (respectively cumulants κ) of p by differentiation of M (respectively $\log M$) at $\mathbf{t} = \mathbf{0}$:

1.3. FOURIER TRANSFORMS IN CRYSTALLOGRAPHY

$$\begin{aligned}\mu_{r_1 r_2 \dots r_n} &\equiv \int_D m(\mathbf{x}) \xi_1^{r_1}(\mathbf{x}) \xi_2^{r_2}(\mathbf{x}) \dots \xi_n^{r_n}(\mathbf{x}) d^3 \mathbf{x} \\ &= \frac{\partial^{r_1 + \dots + r_n} (M)}{\partial t_1^{r_1} \dots \partial t_n^{r_n}} \\ \kappa_{r_1 r_2 \dots r_n} &= \frac{\partial^{r_1 + \dots + r_n} (\log M)}{\partial t_1^{r_1} \dots \partial t_n^{r_n}}.\end{aligned}$$

The structure-factor algebra for group G (Section 1.3.4.2.2.9) then allows one to express products of ξ 's as linear combinations of other ξ 's, and hence to express all moments and cumulants of distribution $p(\xi)$ as linear combinations of real and imaginary parts of Fourier coefficients of the prior distribution of atoms $m(\mathbf{x})$. This plays a key role in the use of non-uniform distributions of atoms.

(d) The joint probability distribution of structure factors

In the random-atom model of an equal-atom structure, N atoms are placed randomly, independently of each other, in the asymmetric unit D of the crystal with probability density $m(\mathbf{x})$. For point atoms of unit weight, the vector \mathbf{F} of structure-factor values for reflections $\mathbf{h} \in H$ may be written

$$\mathbf{F} = \sum_{l=1}^N \xi^{[l]},$$

where the N copies $\xi^{[l]}$ of random vector ξ are independent and have the same distribution $p(\xi)$.

The joint probability distribution $\mathcal{P}(\mathbf{F})$ is then [Section 1.3.4.5.2.1(e)]

$$\mathcal{P}(\mathbf{X}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \exp[N \log M(it) - it \cdot \mathbf{X}] d^n t.$$

For low dimensionality n it is possible to carry out the Fourier transformation numerically after discretization, provided $M(it)$ is sampled sufficiently finely that no aliasing results from taking its N th power (Barakat, 1974). This exact approach can also accommodate heterogeneity, and has been used first in the field of intensity statistics (Shmueli *et al.*, 1984, 1985; Shmueli & Weiss, 1987, 1988), then in the study of the Σ_1 and Σ_2 relations in triclinic space groups (Shmueli & Weiss, 1985, 1986). Some of these applications are described in Chapter 2.1 of this volume. This method could be extended to the construction of any joint probability distribution (j.p.d.) in any space group by using the generic expression for the moment-generating function (m.g.f.) derived by Bricogne (1984). It is, however, limited to small values of n by the necessity to carry out n -dimensional FFTs on large arrays of sample values.

The asymptotic expansions of Gram–Charlier and Edgeworth have good convergence properties only if $F_{\mathbf{h}}$ lies in the vicinity of $\langle F_{\mathbf{h}} \rangle = N \mathcal{F}[m](\mathbf{h})$ for all $\mathbf{h} \in H$. Previous work on the j.p.d. of structure factors has used for $m(\mathbf{x})$ a uniform distribution, so that $\langle \mathbf{F} \rangle = \mathbf{0}$; as a result, the corresponding expansions are accurate only if all moduli $|F_{\mathbf{h}}|$ are small, in which case the j.p.d. contains little phase information.

The saddlepoint method [Section 1.3.4.5.2.1(f)] constitutes the method of choice for evaluating the joint probability $\mathcal{P}(\mathbf{F}^*)$ of structure factors when some of the moduli in \mathbf{F}^* are large. As shown previously, this approximation amounts to using the ‘conjugate distribution’

$$p_{\tau}(\xi) = p(\xi) \frac{\exp(\tau \cdot \xi)}{M(\tau)}$$

instead of the original distribution $p(\xi) = p_0(\xi)$ for the distribution of random vector ξ . This conjugate distribution p_{τ} is induced from

the modified distribution of atoms

$$q_{\tau}(\mathbf{x}) = m(\mathbf{x}) \frac{\exp[\tau \cdot \xi(\mathbf{x})]}{M(\tau)}, \quad (\text{SP1})$$

where, by the induction formula, $M(\tau)$ may be written as

$$M(\tau) = \int_D m(\mathbf{x}) \exp[\tau \cdot \xi(\mathbf{x})] d^3 \mathbf{x} \quad (\text{SP2})$$

and where τ is the unique solution of the saddlepoint equation:

$$\nabla_{\tau} (\log M^N) = \mathbf{F}^*. \quad (\text{SP3})$$

The desired approximation is then

$$\mathcal{P}^{\text{SP}}(\mathbf{F}^*) = \frac{\exp(\mathbf{S})}{\sqrt{\det(2\pi\mathcal{Q})}},$$

where

$$\mathbf{S} = \log M^N(\tau) - \tau \cdot \mathbf{F}^*$$

and where

$$\mathcal{Q} = \nabla \nabla^T (\log M^N) = \mathbf{N}\mathbf{Q}.$$

Finally, the elements of the Hessian matrix $\mathbf{Q} = \nabla \nabla^T (\log M)$ are just the trigonometric second-order cumulants of distribution p , and hence can be calculated *via* structure-factor algebra from the Fourier coefficients of $q_{\tau}(\mathbf{x})$. All the quantities involved in the expression for $\mathcal{P}^{\text{SP}}(\mathbf{F}^*)$ are therefore effectively computable from the initial data $m(\mathbf{x})$ and \mathbf{F}^* .

(e) Maximum-entropy distributions of atoms

One of the main results in Bricogne (1984) is that the modified distribution $q_{\tau}(\mathbf{x})$ in (SP1) is the unique distribution which has maximum entropy $\mathcal{I}_m(q)$ relative to $m(\mathbf{x})$, where

$$\mathcal{I}_m(q) = - \int_D q(\mathbf{x}) \log \left[\frac{q(\mathbf{x})}{m(\mathbf{x})} \right] d^3 \mathbf{x},$$

under the constraint that \mathbf{F}^* be the centroid vector of the corresponding conjugate distribution $\mathcal{P}_{\tau}(\mathbf{F})$. The traditional notation of maximum-entropy (ME) theory (Jaynes, 1957, 1968, 1983) is in this case (Bricogne, 1984)

$$q^{\text{ME}}(\mathbf{x}) = m(\mathbf{x}) \frac{\exp[\lambda \cdot \xi(\mathbf{x})]}{Z(\lambda)} \quad (\text{ME1})$$

$$Z(\lambda) = \int_D m(\mathbf{x}) \exp[\lambda \cdot \xi(\mathbf{x})] d^3 \mathbf{x} \quad (\text{ME2})$$

$$\nabla_{\lambda} (\log Z^N) = \mathbf{F}^* \quad (\text{ME3})$$

so that Z is identical to the m.g.f. M , and the coordinates τ of the saddlepoint are the Lagrange multipliers λ for the constraints \mathbf{F}^* .

Jaynes’s ME theory also gives an estimate for $\mathcal{P}(\mathbf{F}^*)$:

$$\mathcal{P}^{\text{ME}}(\mathbf{F}^*) \approx \exp(\mathcal{I}),$$

where

$$\mathcal{I} = \log Z^N - \lambda \cdot \mathbf{F}^* = N \mathcal{I}_m(q^{\text{ME}})$$

is the total entropy and is the counterpart to \mathbf{S} under the equivalence just established.

\mathcal{P}^{ME} is identical to \mathcal{P}^{SP} , but lacks the denominator. The latter, which is the normalization factor of a multivariate Gaussian with covariance matrix \mathcal{Q} , may easily be seen to arise through Szegő’s theorem (Sections 1.3.2.6.9.4, 1.3.4.2.1.10) from the extra logarithmic term in Stirling’s formula

$$\log(q!) \approx q \log q - q + \frac{1}{2} \log(2\pi q)$$

1. GENERAL RELATIONSHIPS AND TECHNIQUES

(see, for instance, Reif, 1965) beyond the first two terms which serve to define entropy, since

$$\frac{1}{n} \log \det (2\pi\mathbf{Q}) \approx \int_{\mathbb{R}^3/\mathbb{Z}^3} \log 2\pi q^{\text{ME}}(\mathbf{x}) d^3\mathbf{x}.$$

The relative effect of this extra normalization factor depends on the ratio

$$\frac{n}{N} = \frac{\text{dimension of } \mathbf{F} \text{ over } \mathbb{R}}{\text{number of atoms}}.$$

The above relation between entropy maximization and the saddlepoint approximation is the basis of a Bayesian statistical approach to the phase problem (Bricogne, 1988) where the assumptions under which joint distributions of structure factors are sought incorporate many new ingredients (such as molecular boundaries, isomorphous substitutions, known fragments, noncrystallographic symmetries, multiple crystal forms) besides trial phase choices for basis reflections. The ME criterion intervenes in the construction of $q^{\text{ME}}(\mathbf{x})$ under these assumptions, and the distribution $q^{\text{ME}}(\mathbf{x})$ is a very useful computational intermediate in obtaining the approximate joint probability $\mathcal{P}^{\text{SP}}(\mathbf{F}^*)$ and the associated conditional distributions and likelihood functions.

(f) Role of the Fourier transformation

The formal developments presented above make use of the following properties of the Fourier transformation:

- (i) the convolution theorem, which turns the convolution of probability distributions into the multiplication of their characteristic functions;
- (ii) the differentiation property, which confers moment-generating properties to characteristic functions;
- (iii) the reciprocity theorem, which allows the retrieval of a probability distribution from its characteristic or moment-generating function;
- (iv) the Paley–Wiener theorem, which allows the analytic continuation of characteristic functions associated to probability

distributions with compact support, and thus gives rise to conjugate families of distributions;

(v) Bertaut’s structure-factor algebra (a discrete symmetrized version of the convolution theorem), which allows the calculation of all necessary moments and cumulants when the dimension n is small;

(vi) Szegő’s theorem, which provides an asymptotic approximation of the normalization factor when n is large.

This multi-faceted application seems an appropriate point at which to end this description of the Fourier transformation and of its use in crystallography.

Acknowledgements

Many aspects of the theory of discrete Fourier transform algorithms and of its extension to incorporate crystallographic symmetry have been the focus of a long-standing collaborative effort between Professor Louis Auslander, Professor Richard Tolimieri, their co-workers and the writer. I am most grateful to them for many years of mathematical stimulation and enjoyment, for introducing me to the ‘big picture’ of the discrete Fourier transform which they have elaborated over the past decade, and for letting me describe here some of their unpublished work. In particular, the crystallographic extensions of the Rader/Winograd algorithms presented in Section 1.3.4.3.4.3 were obtained by Richard Tolimieri, in a collaboration partially supported by NIH grant GM 32362 (to the writer).

I am indebted to the Editor for many useful and constructive suggestions of possible improvements to the text, only a few of which I have been able to implement. I hope to incorporate many more of them in the future. I also wish to thank Dr D. Sayre for many useful comments on an early draft of the manuscript.

This contribution was written during the tenure of a Visiting Fellowship at Trinity College, Cambridge, with partial financial support from Trinity College and the MRC Laboratory of Molecular Biology. I am most grateful to both institutions for providing ideal working conditions.