

2.1. Statistical properties of the weighted reciprocal lattice

BY U. SHMUELI AND A. J. C. WILSON

2.1.1. Introduction

The structure factor of the hkl reflection is given by

$$F(hkl) = \sum_{j=1}^N f_j \exp[2\pi i(hx_j + ky_j + lz_j)], \quad (2.1.1.1)$$

where f_j is the atomic scattering factor [complex if there is appreciable dispersion; see Chapter 1.2 and *IT C* (1999, Section 4.2.6)], x_j, y_j, z_j are the fractional coordinates of the j th atom and N is the number of atoms in the unit cell. The present chapter is concerned with the statistical properties of the structure factor F and the intensity $I = FF^*$, such as their average values, variances, higher moments and their probability density distributions.

Equation (2.1.1.1) expresses F as a function of two conceptually different sets of variables: hkl taking on integral values in reciprocal space and xyz in general having non-integral values in direct space, although the special positions tabulated for each space group in *IT A* (1983) may include the integers 0 and 1. In special positions, the non-integers often include rational fractions, but in general positions they are in principle irrational. Although hkl and xyz appear to be symmetrical variables in (2.1.1.1), these limitations on their values mean that one can consider two different sets of statistical properties. In the first we seek, for example, the average intensity of the hkl reflection (indices fixed) as the positional parameters of the N atoms are distributed with equal probability over the continuous range 0–1. In the second, we seek, for example, the average intensity of the observable reflections (or of a subgroup of them having about the same value of $\sin \theta$) with the values of xyz held constant at the values they have, or are postulated to have, in a crystal structure. Other examples are obtained by substituting the words ‘probability density’ for ‘average intensity’. For brevity, we may call the statistics resulting from the first process *fixed-index* (continuously variable parameters being understood), and those resulting from the second process *fixed-parameter* (integral indices being understood). Theory based on the first process is (comparatively) easy; theory based on the second hardly exists, although there is a good deal of theory concerning the conditions under which the two processes will lead to the same result (Hauptman & Karle, 1953; Giacovazzo, 1977, 1980). Mathematically, of course, the condition is that the phase angle

$$\vartheta = 2\pi(hx + ky + lz) \quad (2.1.1.2)$$

should be distributed with uniform probability over the range 0– 2π , whichever set of variables is regarded as fixed, but it is not clear when this distribution can be expected in practice for fixed-parameter averaging. The usual conclusion is that the uniform distribution will be realized if there are enough atoms, if the atomic coordinates do not approximate to rational fractions, if there are enough reflections and if stereochemical effects are negligible (Shmueli *et al.*, 1984).

Obviously, the second process (fixed parameters, varying integral indices) corresponds to the observable reality, and various approximations to it have been attempted, in preference to assuming its equivalence with the first. For example, a third (approximate) method of averaging has been used (Wilson, 1949, 1981): xyz were held fixed and hkl are treated as continuous variables.

2.1.2. The average intensity of general reflections

2.1.2.1. Mathematical background

The process may be illustrated by evaluating, or attempting to evaluate, the average intensity of reflection by the three processes.

The intensity of reflection is given by multiplying equation (2.1.1.1) by its complex conjugate:

$$I = FF^* \quad (2.1.2.1)$$

$$= \sum_{j,k} f_j f_k^* \exp\{2\pi i[h(x_j - x_k) + \dots]\} \quad (2.1.2.2)$$

$$= \Sigma + \sum_{j \neq k} f_j f_k^* \exp\{2\pi i[h(x_j - x_k) + \dots]\}, \quad (2.1.2.3)$$

where

$$\Sigma = \sum_j f_j f_j^* \quad (2.1.2.4)$$

is the sum of the squares of the moduli of the atomic scattering factors. Wilson (1942) argued, without detailed calculation, that the average value of the exponential term would be zero and hence that

$$\langle I \rangle = \Sigma. \quad (2.1.2.5)$$

Averaging equation (2.1.2.3) for hkl fixed, xyz ranging uniformly over the unit cell – the first process described above – gives this result identically, without complication or approximation. Ordinarily the second process cannot be carried out. We can, however, postulate a special case in which it is possible. We take a homoatomic structure and before averaging we correct the f 's for temperature effects and the fall-off with $\sin \theta$, so that ff^* is the same for all the atoms and is independent of hkl . If the range of hkl over which the expression for I has to be averaged is taken as a parallelepiped in reciprocal space with h ranging from $-H$ to $+H$, k from $-K$ to $+K$, l from $-L$ to $+L$, equation (2.1.2.2) can be factorized into the product of the sums of three geometrical progressions. Algebraic manipulation then easily leads to

$$\begin{aligned} \langle I \rangle &= ff^* \sum_j \sum_k \frac{\sin \pi N_H (x_j - x_k)}{N_H \sin \pi (x_j - x_k)} \\ &\times \frac{\sin \pi N_K (y_j - y_k)}{N_K \sin \pi (y_j - y_k)} \frac{\sin \pi N_L (z_j - z_k)}{N_L \sin \pi (z_j - z_k)}, \end{aligned} \quad (2.1.2.6)$$

where $N_H = 2H + 1$, $N_K = 2K + 1$ and $N_L = 2L + 1$. The terms with $j = k$ give Σ , but the remaining terms are not zero. Because of the periodic nature of the trigonometric terms, the effective coordinate differences are never greater than 0.5 and in a structure of any complexity there will be many much less than 0.5. For $HKL = 000$, in fact, $\langle I \rangle$ becomes the square of the modulus of the sum of the atomic scattering factors,

$$\langle I \rangle = \Phi \Phi^*, \quad (2.1.2.7)$$

where

$$\Phi = \sum_{j=1}^N f_j, \quad (2.1.2.8)$$

and not the sum of the squares of their moduli; for larger HKL , $\langle I \rangle$ rapidly decreases to Σ and then oscillates about that value. Wilson (1949, especially Section 2.1.1) suggested that the regions of averaging should be chosen so that at least one index of every reflection is ≥ 2 if $\langle I \rangle$ is to be identified with Σ , and this has proven to be a useful rule-of-thumb.

The third process of averaging replaces the sum over integral values of the indices by an integration over continuous values, the appropriate values of the limits in this example being $-(H + 1/2)$ to $+(H + 1/2)$. The effect is to replace the sines in the

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

denominators, but not in the numerators, of equation (2.1.2.6) by their arguments, and this is equivalent to the approximation $\sin x \simeq x$ in the denominators only. This is a good approximation for atoms close together in the structure and thus giving the largest terms in the sums in equation (2.1.2.6), and gives the correct sign and order of magnitude even for x having its maximum value of $\pi/2$.

2.1.2.2. Physical background

The preceding section has used mathematical arguments. From a physical point of view, the radiation diffracted by atoms that are resolved will interfere destructively, so that the resulting intensity will be the sum of the intensities diffracted by individual atoms, whereas that from completely unresolved atoms will interfere constructively, so that *amplitudes* rather than intensities add. In intermediate cases there will be partial constructive interference. Resolution in accordance with the Rayleigh (1879) criterion requires that $s = (2 \sin \theta) / \lambda$ should be greater than half the reciprocal of the minimum interatomic distance in the crystal (Wilson, 1979); full resolution requires a substantial multiple of this. This criterion is essentially equivalent to that proposed from the study of a special case of the second process in the preceding section.

2.1.2.3. An approximation for organic compounds

In organic compounds there are very many interatomic distances of about 1.5 or 1.4 Å. Adoption of the preceding criterion would mean that the inner portion of the region of reciprocal space accessible by the use of copper $K\alpha$ radiation is not within the sphere of intensity statistics based on fixed-index (first process) averaging. No substantial results are available for fixed-parameter (second process) averaging, and very few from the approximation to it (third process).

To the extent to which the third process is acceptable, an approximation to the variation of $\langle I \rangle$ with $\sin \theta$ is obtainable. The exponent in equation (2.1.2.2) can be written as

$$2\pi i s r_{jk} \cos \psi, \quad (2.1.2.9)$$

where s is the radial distance in reciprocal space, r_{jk} is the distance from the j th to the k th atom and ψ is the angle between the vectors s and r . Averaging over a sphere of radius s , with ψ treated as the co-latitude, gives

$$\langle I \rangle = \sum_j \sum_k f_j f_k \frac{\sin 2\pi s r_{jk}}{2\pi s r_{jk}}. \quad (2.1.2.10)$$

This is the familiar Debye expression. It has the correct limits for s zero and s large, and is in accord with the argument from resolution.

2.1.2.4. Effect of centring

In the preceding discussion there has been a tacit assumption that the lattice is primitive. A centred crystal can always be referred to a primitive lattice and if this is done no change is required. If the centred lattice is retained, many reflections are identically zero and the intensity of the non-zero reflections is enhanced by a factor of two (I and C lattices) or four (F lattice), so that the average intensity of all the reflections, zero and non-zero taken together, is unchanged.

Other symmetry elements affect only zones and rows of reflections, and so do not affect the general average when the total number of reflections is large. Their effect on zones and rows is discussed in Section 2.1.3.

Table 2.1.3.1. *Intensity-distribution effects of symmetry elements causing systematic absences*

Abbreviations and orientation of axes: A = acentric distribution, C = centric distribution, Z = systematically zero, S = distribution parameter, $\langle I \rangle$ = average intensity. Axes are parallel to \mathbf{c} , planes are perpendicular to \mathbf{c} .

Element	Reflections	Distribution	S/Σ	$\langle I \rangle/\Sigma$
2_1	hkl	A	1	1
	$hk0$	C	1	1
	$00l$	$(Z + A)/2$	1	2
$3_{1, 3_2}$	hkl	A	1	1
	$hk0$	A	1	1
	$00l$	$(2Z + A)/3$	1	3
$4_{1, 4_3}$	hkl	A	1	1
	$hk0$	C	1	1
	$00l$	$(3Z + A)/4$	1	4
4_2	hkl	A	1	1
	$hk0$	C	1	1
	$00l$	$(Z + A)/2$	2	4
$6_{1, 6_5}$	hkl	A	1	1
	$hk0$	C	1	1
	$00l$	$(5Z + A)/6$	1	6
$6_{2, 6_4}$	hkl	A	1	1
	$hk0$	C	1	1
	$00l$	$(2Z + A)/3$	2	6
6_3	hkl	A	1	1
	$hk0$	C	1	1
	$00l$	$(Z + A)/2$	3	6
a	hkl	A	1	1
	$hk0$	$(Z + A)/2$	1	2
	$00l$	C	1	1
	$0k0$	A	2	2
C, I	All	$(Z + A)/2$	1	2
F	All	$(3Z + A)/2$	1	4

2.1.3. The average intensity of zones and rows

2.1.3.1. Symmetry elements producing systematic absences

Symmetry elements can be divided into two types: those that cause systematic absences and those that do not. Those producing systematic absences (glide planes and screw axes) produce at the same time groups of reflections (confined to zones and rows in reciprocal space, respectively) with an average intensity an integral* multiple of the general average. The effects for single symmetry elements of this type are given in Table 2.1.3.1 for the general reflections hkl and separately for any zones and rows that are affected. The 'average multipliers' are given in the column headed $\langle I \rangle/\Sigma$; 'distribution' and 'distribution parameters' are treated in Section 2.1.5. As for the centring, the fraction of reflections missing and the integer multiplying the average are related in such a way that the overall intensity is unchanged. The

* The multiple is given as an exact integer for fixed-index averaging, an approximate integer for fixed-parameter averaging. Statements should be understood to refer to fixed-index averaging unless the contrary is explicitly stated.

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

Table 2.1.3.2. *Intensity-distribution effects of symmetry elements not causing systematic absences*

Abbreviations and orientation of axes: *A* = acentric distribution, *C* = centric distribution, *S* = distribution parameter, $\langle I \rangle$ = average intensity. Axes are parallel to *c*, planes are perpendicular to *c*.

Element	Reflections	Distribution	$S/\Sigma = \langle I \rangle/\Sigma$
1	All	<i>A</i>	1
$\bar{1}$	All	<i>C</i>	1
2	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>C</i>	1
	<i>00l</i>	<i>A</i>	2
$\bar{2} = m$	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>A</i>	2
	<i>00l</i>	<i>C</i>	1
3	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>A</i>	1
	<i>00l</i>	<i>A</i>	3
$\bar{3}$	<i>hkl</i>	<i>C</i>	1
	<i>hk0</i>	<i>C</i>	1
	<i>00l</i>	<i>C</i>	3
4	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>C</i>	1
	<i>00l</i>	<i>A</i>	4
$\bar{4}$	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>C</i>	1
	<i>00l</i>	<i>C</i>	2
6	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>C</i>	1
	<i>00l</i>	<i>A</i>	6
$\bar{6} = 3/m$	<i>hkl</i>	<i>A</i>	1
	<i>hk0</i>	<i>A</i>	2
	<i>00l</i>	<i>C</i>	3

mechanism for compensation for the reflections with enhanced intensity is obvious.

2.1.3.2. *Symmetry elements not producing systematic absences*

Certain symmetry elements not producing absences (mirror planes and rotation axes) cause equivalent atoms to coincide in a plane or a line projection and hence produce a zone or row in reciprocal space for which the average intensity is an integral multiple of the general average (Wilson, 1950); the effects of single such symmetry elements are given in Table 2.1.3.2. There is, however, no obvious mechanism for compensation for this enhancement. When reflections are few this may be an important matter in assigning an approximate absolute scale by comparing observed and calculated intensities. Wilson (1964), Nigam (1972) and Nigam & Wilson (1980), noting that in such cases the finite size of atoms results in forbidden ranges of positional parameters, have shown that there is a diminution of the intensity of layers (rows) in the immediate neighbourhood of the enhanced zones (rows), just sufficient to compensate for the enhancement. In forming general averages, therefore, reflections from enhanced zones or rows should be included at their full intensity, not divided by the multiplier; the

matter is discussed in more detail by Wilson (1987a). It should be noted, however, that organic structures containing molecules related by rotation axes are rare, and such structures related by mirror planes are even rarer (Wilson, 1993).

2.1.3.3. *More than one symmetry element*

Further alterations of the intensities occur if two or more such symmetry elements are present in the space group. The effects were treated in detail by Rogers (1950), who used them to construct a table for the determination of space groups by supplementing the usual knowledge of Laue group with statistical information. Only two pairs of space groups, the orthorhombic *I222* and *I2₁2₁2₁*, and their cubic supergroups *I23* and *I2₁3₁*, remained unresolved. Examination of this table shows that what statistical information does is to resolve the Laue group into point groups; the further resolution into space groups is equivalent to the use of Table 3.2 in *IT A* (1983). The statistical consequences of each point group, as given by Rogers, are reproduced in Table 2.1.3.3.

2.1.4. Probability density distributions – mathematical preliminaries

For the purpose of this chapter, ‘ideal’ probability distributions or probability density functions are the asymptotic forms obtained by the use of the central-limit theorem when the number of atoms in the unit cell, *N*, is sufficiently large. In order to derive them it is necessary to outline the properties of characteristic functions and to state alternative conditions for the validity of the central-limit theorem; the distributions themselves are derived in Section 2.1.5.

2.1.4.1. *Characteristic functions*

The average value of $\exp(itx)$ is very important in probability theory; it is called the characteristic function of the distribution $f(x)$ and is denoted by $C_x(t)$ or, when no confusion can arise, by $C(t)$. It exists for all legitimate distributions, whether discrete or continuous. In the continuous case it is given by

$$C(t) = \int_{-\infty}^{\infty} \exp(itx)f(x) dx, \quad (2.1.4.1)$$

and is thus the Fourier transform of $f(x)$. In many cases it can be obtained from known integrals. For example, for the Cauchy distribution,

$$C(t) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{\exp(itx)}{a^2 + x^2} dx \quad (2.1.4.2)$$

$$= \exp(-a|t|), \quad (2.1.4.3)$$

and for the normal distribution,

$$C(t) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right) \exp(itx) dx \quad (2.1.4.4)$$

$$= \exp\left(imt - \frac{\sigma^2 t^2}{2}\right). \quad (2.1.4.5)$$

Since the characteristic function is the Fourier transform of the distribution function, the converse is true, and if the characteristic function is known the probability distribution function can be obtained by the use of Fourier inversion theorem,

$$f(x) = (1/2\pi) \int_{-\infty}^{\infty} \exp(-itx)C(t) dt. \quad (2.1.4.6)$$

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

Table 2.1.3.3. Average multiples for the 32 point groups (modified from Rogers, 1950).

The multiple gives S/Σ for the row and zone corresponding to the principal axis of the point-group symbol; those for the secondary and tertiary axes are given when the symbol contains such axes.

Point group	Principal		Secondary		Tertiary	
	Row	Zone	Row	Zone	Row	Zone
1 $\bar{1}$	1 1	1 1				
2 m $2/m$	2 1 2	1 2 2				
222 $mm2$ mmm	2 2 4	1 2 2	2 2 4	1 2 2	2 4 4	1 1 2
4 $\bar{4}$ $4/m$	4 2 4	1 1 2				
422 $4mm$ $\bar{4}2m$ $4/mmm$	4 8 4 8	1* 1 1 2	2 2 2 4	1 2 1 2	2 2 2 4	1 2 2 2
3 $\bar{3}$	3 3	1 1				
321 $3m1$ $31m$	3 6 6	1 1 1	2 1 2	1 2 2	1 2 2	1 1 1
6 $\bar{6}$ $6/m$	6 3 6	1 2 2				
622 $6mm$ $\bar{6}m2$ $6/mmm$	6 12 6 12	1 1 2 2	2 2 2 4	1 2 2 2	2 2 4 4	1 2 1 2
231 $m\bar{3}1$	2 4	1 2	3 3	1 1	1 1	1 1
432 $\bar{4}3m$ $m\bar{3}m$	4 4 8	1 1 2	3 6 6	1 1 2	2 2 4	1 2 2

Note. The pairs of point groups, 1 and $\bar{1}$ and 3 and $\bar{3}$, not distinguished by average multiples, may be distinguished by their centric and acentric probability density functions.

* The entry for the principal zone for the point group 422 was given incorrectly as 2 in the first edition of this volume.

An alternative approach to the derivation of the distribution from a known characteristic function will be discussed below.

The most important property of characteristic functions in crystallography is the following: if x and y are independent random variables with characteristic functions $C_x(t)$ and $C_y(t)$, the characteristic function of their sum

$$z = x + y \quad (2.1.4.7)$$

is the product

$$C_z(t) = C_x(t)C_y(t). \quad (2.1.4.8)$$

Obviously this can be extended to any number of independent random variables.

When the moments exist, the characteristic function can be expanded in a power series in which the k th term is $m_k(it)^k/k!$. If the power series

$$\exp(itx) = 1 + itx + \frac{(it)^2x^2}{2!} + \frac{(it)^3x^3}{3!} + \dots \quad (2.1.4.9)$$

is substituted in equation (2.1.4.1), one obtains

$$C(t) = 1 + itm'_1 + \frac{(it)^2m'_2}{2!} + \frac{(it)^3m'_3}{3!} + \dots \quad (2.1.4.10)$$

The moments are written with primes in order to indicate that equation (2.1.4.10) is valid for moments about an arbitrary origin as well as for moments about the mean. If the random variable is transformed by a change of origin and scale, say

$$y = \frac{x - a}{b}, \quad (2.1.4.11)$$

the characteristic function for y becomes

$$C_y(t) = b \exp(-iat/b)C_x(t). \quad (2.1.4.12)$$

2.1.4.2. The cumulant-generating function

A function that is often more useful than the characteristic function is its logarithm, the cumulant-generating function:

$$K(t) = \log C(t) = k_1 + \frac{k_2(it)^2}{2!} + \frac{k_3(it)^3}{3!} + \dots, \quad (2.1.4.13)$$

where the k 's are called the *cumulants* and may be regarded as being defined by the equation. They can be evaluated in terms of the moments by combining the series (2.1.4.10) for $C(t)$ with the ordinary series for the logarithm and equating the coefficients of t^r . In most cases the process as described is tedious, but it can be shortened by use of a general method [Stuart & Ord (1994), Section 3.14, pp. 87–88; Exercise 3.19, p. 119]. Obviously, the cumulants exist only if the moments exist. The first few relations are

$$\begin{aligned} k_0 &= 0 \\ k_1 &= m'_1 \\ k_2 &= m_2 = m'_2 - (m'_1)^2 \\ k_3 &= m_3 = m'_3 - 3m'_2m'_1 + 2(m'_1)^2 \\ k_4 &= m_4 - 3(m_2)^2 \\ &= m'_4 - m'_3m'_1 - 3(m'_2)^2 + 12m'_2(m'_1)^2 - 6(m'_1)^4. \end{aligned} \quad (2.1.4.14)$$

Such expressions and their converses up to k_{10} are given by Stuart & Ord (1994, pp. 88–91). Since all the cumulants except k_1 can be expressed in terms of the central moments only (*i.e.*, those unprimed), only k_1 is changed by a change of the origin. Because of this property, they are sometimes called the semi-invariants (or seminvariants) of the distribution. Since addition of random variables is equivalent to the multiplication of their characteristic functions [equation (2.1.4.8)] and multiplication of functions is equivalent to the addition of their logarithms, each cumulant of the distribution of the sum of a number of random variables is equal to the sum of the cumulants of the distribution functions of the individual variables – hence the name cumulants. Although the cumulants (except k_1) are independent of a change of origin, they are not independent of a change of scale. As for the moments, a

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

change of scale simply multiplies them by a power of the scale factor; if $y = x/b$

$$(k_y)_r = (k_x)_r/b^r. \quad (2.1.4.15)$$

The cumulants of the normal distribution are particularly simple. From equation (2.1.4.5), the cumulant-generating function of a normal distribution is

$$K(t) = imt - \sigma^2 t^2/2 \quad (2.1.4.16)$$

$$k_1 = m \quad (2.1.4.17)$$

$$k_2 = \sigma^2, \quad (2.1.4.18)$$

all cumulants with $r > 2$ are identically zero.

2.1.4.3. The central-limit theorem

A simple form of this important theorem can be stated as follows:

If x_1, x_2, \dots, x_n are independent and identically distributed random variables, each of them having the same mean m and variance σ^2 , then the sum

$$S_n = \sum_{j=1}^n x_j \quad (2.1.4.19)$$

tends to be normally distributed – independently of the distribution(s) of the individual random variables – with mean nm and variance $n\sigma^2$, provided n is sufficiently large.

In order to prove this theorem, let us define a standardized random variable corresponding to the sum S_n , *i.e.*, such that its mean is zero and its variance is unity:

$$\hat{S}_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\sum_{j=1}^n (x_j - m)}{\sigma\sqrt{n}} \equiv \sum_{j=1}^n \frac{W_j}{\sqrt{n}}, \quad (2.1.4.20)$$

where $W_j = (x_j - m)/\sigma$ is a standardized single random variable. The characteristic function of \hat{S}_n is therefore given by

$$C_n(\hat{S}_n, t) = \langle \exp(it\hat{S}_n) \rangle = \left\langle \exp \left[it \sum_{j=1}^n \frac{W_j}{\sqrt{n}} \right] \right\rangle \quad (2.1.4.21)$$

$$= \prod_{j=1}^n \left\langle \exp \left[it \frac{W_j}{\sqrt{n}} \right] \right\rangle \quad (2.1.4.22)$$

$$= \left\{ \left\langle \exp \left[it \frac{W_1}{\sqrt{n}} \right] \right\rangle \right\}^n, \quad (2.1.4.23)$$

where the brackets $\langle \rangle$ denote the operation of averaging with respect to the appropriate probability density function (p.d.f.) [cf. equation (2.1.4.1)]. Equation (2.1.4.22) follows from equation (2.1.4.21) by the assumption of independence, while the assumption of identically distributed variables leads to the identity of the characteristic functions of the individual variables – as seen in equation (2.1.4.23).

On the assumption that moments of all the orders exist – a most plausible assumption in situations usually encountered in structure-factor statistics – we can now expand the characteristic function of a single variable in a power series [cf. equation (2.1.4.10)]:

$$\begin{aligned} \left\langle \exp \left[it \frac{W_1}{\sqrt{n}} \right] \right\rangle &= \left\langle \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \frac{W_1^r}{n^{r/2}} \right\rangle \\ &= \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \frac{\langle W_1^r \rangle}{n^{r/2}} \\ &\equiv 1 - \frac{t^2}{2n} + \frac{\zeta(t, n)}{n}, \end{aligned} \quad (2.1.4.24)$$

since $\langle W_1 \rangle = 0$, $\langle W_1^2 \rangle = 1$, and the quantity denoted by $\zeta(t, n)$ in (2.1.4.24) is given by

$$\zeta(t, n) = \sum_{r=3}^{\infty} \frac{(it)^r}{r!} \frac{\langle W_1^r \rangle}{n^{(r/2)-1}}. \quad (2.1.4.25)$$

The characteristic function of \hat{S}_n is therefore

$$\langle \exp(it\hat{S}_n) \rangle = \left[1 - \frac{t^2}{2n} + \frac{\zeta(t, n)}{n} \right]^n. \quad (2.1.4.26)$$

Now, as is seen from (2.1.4.25), for every fixed t the quantity $\zeta(t, n)$ tends to zero as n tends to infinity. The cumulant-generating function of the standardized sum then becomes

$$\log C_n(\hat{S}_n, t) = n \log \left[1 - \frac{1}{n} \left(\frac{t^2}{2} - \zeta(t, n) \right) \right] \quad (2.1.4.27)$$

and the logarithm on the right-hand side of equation (2.1.4.27) has the form $\log(1 - z)$ with $|z| \rightarrow 0$ as $n \rightarrow \infty$. We may therefore use the expansion

$$\log(1 - z) = - \left(z + \frac{z^2}{2} + \frac{z^3}{3} + \dots \right),$$

which is valid for $|z| < 1$. We then obtain

$$\begin{aligned} \log C_n(\hat{S}_n, t) &= -n \left[\frac{1}{n} \left(\frac{t^2}{2} - \zeta(t, n) \right) + \frac{1}{2n^2} \left(\frac{t^2}{2} - \zeta(t, n) \right)^2 \right. \\ &\quad \left. + \frac{1}{3n^3} \left(\frac{t^2}{2} - \zeta(t, n) \right)^3 + \dots \right] \\ &= -\frac{t^2}{2} + \zeta(t, n) - \frac{1}{2n} \left(\frac{t^2}{2} - \zeta(t, n) \right)^2 \\ &\quad - \frac{1}{3n^2} \left(\frac{t^2}{2} - \zeta(t, n) \right)^3 - \dots \end{aligned}$$

and finally, for every fixed t ,

$$\lim_{n \rightarrow \infty} \log C_n(\hat{S}_n, t) = -\frac{t^2}{2}. \quad (2.1.4.28)$$

Since the logarithm is a continuous function of t , it follows directly that

$$\lim_{n \rightarrow \infty} C_n(\hat{S}_n, t) = \exp \left(-\frac{t^2}{2} \right). \quad (2.1.4.29)$$

The right-hand side of (2.1.4.29) is just the characteristic function of a standardized normal p.d.f., *i.e.*, a normal p.d.f. with zero mean and unit variance [cf. equation (2.1.4.5)]. The asymptotic expression for the p.d.f. of the standardized sum is therefore obtained as

$$p(\hat{S}) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{\hat{S}^2}{2} \right),$$

which proves the above version of the central-limit theorem.

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

Surprisingly, this theorem has a very wide applicability and values of n as low as 30 are often large enough for the theorem to be useful. Situations in which the normal p.d.f. must be modified or replaced by an altogether different one are dealt with in Sections 2.1.7 and 2.1.8 of this chapter.

2.1.4.4. Conditions of validity

The above outline of a proof of the central-limit theorem depended on the existence of moments of all orders. The components of structure factors always possess finite moments of all orders, but the existence of moments beyond the second is not necessary for the validity of the theorem and it can be proved under much less stringent conditions. In fact, if all the random variables in equation (2.1.4.19) have the *same* distribution – as in a homoatomic structure – the only requirement is that the second moments of the distributions should exist [the Lindeberg–Lévy theorem (*e.g.* Cramér, 1951)]. If the distributions are not the same – as in a heteroatomic structure – some further condition is necessary to ensure that no individual random variable dominates the sum. The Liapounoff proof requires the existence of third absolute moments, but this is regarded as aesthetically displeasing; a theorem that ultimately involves only means and variances should require only means and variances in the proof. The Lindeberg–Cramér conditions meet this aesthetic criterion. Roughly, the conditions are that S^2 , the variance of the sum, should tend to infinity and σ_j^2/S^2 , where σ_j^2 is the variance of the j th random variable, should tend to zero for all j as n tends to infinity. The precise formulation is quoted by Kendall & Stuart (1977, p. 207).

2.1.4.5. Non-independent variables

The central-limit theorem, under certain conditions, remains valid even when the variables summed in equation (2.1.4.19) are not independent. The conditions have been investigated by Bernstein (1922, 1927); roughly they amount to requiring that the variables should not be too closely correlated. The theorem applies, in particular, when each x_r is related to a finite number, $f(n)$, of its neighbours, when the x 's are said to be $f(n)$ dependent. The $f(n)$ dependence seems plausible for crystallographic applications, since the positions of atoms close together in a structure are closely correlated by interatomic forces, whereas those far apart will show little correlation if there is any flexibility in the asymmetric unit when unconstrained. Harker's (1953) idea of 'globs' seems equivalent to $f(n)$ dependence. Long-range stereochemical effects, as in pseudo-graphitic aromatic hydrocarbons, would presumably produce long-range correlations and make $f(n)$ dependence less plausible. If Bernstein's conditions are satisfied, the central-limit theorem would apply, but the actual value of $\langle x^2 \rangle - \langle x \rangle^2$ would have to be used for the variance, instead of the sum of the variances of the random variables in (2.1.4.19). Because of the correlations the two values are no longer equal.

French & Wilson (1978) seem to have been the first to appeal explicitly to the central-limit theorem extended to non-independent variables, but many previous workers [for typical references, see Wilson (1981)] tacitly made the replacement – in the X-ray case substituting the local mean intensity for the sum of the squares of the atomic scattering factors.

2.1.5. Ideal probability density distributions

In applications of the central-limit theorem, and its extensions, to intensity statistics the x_j 's of equation (2.1.4.19) have the form (atomic scattering factor of the j th atom) times (a trigonometric expression characteristic of the space group and Wyckoff position; also known as the trigonometric structure factor). These trigono-

metric expressions for all the space groups, and general Wyckoff positions, are given in Tables A1.4.3.1 through A1.4.3.7, and their first few even moments (fixed-index averaging) are given in Table 2.1.7.1. One cannot, of course, conclude that the magnitudes of the structure factor always have a normal distribution – even if the structure is homoatomic; one must look at each problem and see what components of the structure factor can be put in the form (2.1.4.19), deduce the m and σ^2 to be used for each, and combine the components to obtain the asymptotic (large N , not large x) expression for the problem in question. Ordinarily the components are the real and the imaginary parts of the structure factor; the structure factor is purely real only if the structure is centrosymmetric, the space-group origin is chosen at a crystallographic centre and the atoms are non-dispersive.

2.1.5.1. Ideal acentric distributions

The ideal acentric distributions are obtained by applying the central-limit theorem to the real and the imaginary parts of the structure factor, as given by equation (2.1.1.1). Consider first a crystal with no rotational symmetry (space group $P1$). The real part, A , of the structure factor is then given by

$$A = \sum_{j=1}^N f_j \cos \vartheta_j, \quad (2.1.5.1)$$

where N is the number of atoms in the unit cell and ϑ_j is the phase angle of the j th atom. The central-limit theorem then states that A tends to be normally distributed about its mean value with variance equal to its mean-square deviation from its mean. Under the assumption that the phase angles ϑ_j are uniformly distributed on the 0 – 2π range, the mean value of each cosine is zero, so that its variance is

$$\sigma^2 = \sum_{j=1}^N f_j^2 \langle \cos^2 \vartheta_j \rangle. \quad (2.1.5.2)$$

Under the same assumption, the mean value of each $\cos^2 \vartheta$ is one-half, so that the variance becomes

$$\sigma^2 = (1/2) \sum_{j=1}^N f_j^2 = (1/2)\Sigma, \quad (2.1.5.3)$$

where Σ is the sum of the squares of the atomic scattering factors [*cf.* equation (2.1.2.4)]. The asymptotic form of the distribution of A is therefore given by

$$p(A) dA = (\pi\Sigma)^{-1/2} \exp(-A^2/\Sigma) dA. \quad (2.1.5.4)$$

A similar calculation, with sines instead of cosines, gives an analogous distribution for the imaginary part B , so that the joint probability of the real and imaginary parts of F is

$$p(A, B) dA dB = (\pi\Sigma)^{-1} \exp[-(A^2 + B^2)/\Sigma] dA dB. \quad (2.1.5.5)$$

Ordinarily, however, we are more interested in the distribution of the magnitude, $|F|$, of the structure factor than in the distribution of A and B . Using polar coordinates in equation (2.1.5.5) [$A = |F| \cos \phi$, $B = |F| \sin \phi$] and integrating over the angle ϕ gives

$$p(|F|) d|F| = (2|F|/\Sigma) \exp(-|F|^2/\Sigma) d|F|. \quad (2.1.5.6)$$

It is usually convenient, in structure-factor and intensity statistics, to express the results in terms of the normalized structure factor E and its magnitude $|E|$. If $|F|$ has been put on an absolute scale (see Section 2.2.4.3), we have

$$E = \frac{F}{\sqrt{\Sigma}} \quad \text{and} \quad |E| = \frac{|F|}{\sqrt{\Sigma}}, \quad (2.1.5.7)$$

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

so that

$$p(|E|) d|E| = 2|E| \exp(-|E|^2) d|E| \quad (2.1.5.8)$$

is the normalized-structure-factor version of (2.1.5.6).

Distributions resulting from noncentrosymmetric crystals are known as *acentric* distributions; those arising from centrosymmetric crystals are known as *centric*. These adjectives are used to describe *distributions*, not crystal symmetry.

2.1.5.2. Ideal centric distributions

When a non-dispersive crystal is centrosymmetric, and the space-group origin is chosen at a crystallographic centre of symmetry, the imaginary part B of its structure amplitude is zero. In the simplest case, space group $P\bar{1}$, the contribution of the j th atom plus its centrosymmetric counterpart is $2f_j \cos \vartheta_j$. The calculation of $p(A)$ goes through as before, with allowance for the fact that there are $N/2$ pairs instead of N independent atoms, giving

$$p(A) dA = (2\pi\Sigma)^{-1/2} \exp[-A^2/(2\Sigma)] dA \quad (2.1.5.9)$$

or equivalently

$$p(|F|) d|F| = [2/(\pi\Sigma)]^{1/2} \exp[-|F|^2/(2\Sigma)] d|F| \quad (2.1.5.10)$$

or

$$p(|E|) d|E| = (2/\pi)^{1/2} \exp(-|E|^2/2) d|E|. \quad (2.1.5.11)$$

2.1.5.3. Effect of other symmetry elements on the ideal acentric and centric distributions

Additional crystallographic symmetry elements do not produce any essential alterations in the ideal centric or acentric distribution; their main effect is to replace the parameter Σ by a 'distribution parameter', called S by Wilson (1950) and Rogers (1950), in certain groups of reflections. In addition, in noncentrosymmetric space groups, the distribution of certain groups of reflections becomes centric, though the general reflections remain acentric. The changes are summarized in Tables 2.1.3.1 and 2.1.3.2. The values of S are integers for lattice centring, glide planes and those screw axes that produce absences, and approximate integers for rotation axes and mirror planes; the modulations of the average intensity in reciprocal space outlined in Section 2.1.3.2 apply.

It should be noted that if intensities are normalized to the average of the group to which they belong, rather than to the general average, the distributions given in equations (2.1.5.8) and (2.1.5.11) are not affected.

2.1.5.4. Other ideal distributions

The distributions just derived are asymptotic, as they are limiting values for large N . They are the only ideal distributions, in this sense, when there is only strict crystallographic symmetry and no dispersion. However, other ideal (asymptotic) distributions arise when there is noncrystallographic symmetry, or if there is dispersion. The *subcentric* distribution,

$$p(|E|) d|E| = \frac{2|E|}{(1-k^2)^{1/2}} \exp[-|E|^2/(1-k^2)] \times I_0\left(\frac{k|E|^2}{1-k^2}\right) d|E|, \quad (2.1.5.12)$$

where $I_0(x)$ is a modified Bessel function of the first kind and k is the ratio of the scattering from the centrosymmetric part to the total scattering, arises when a noncentrosymmetric crystal contains centrosymmetric parts or when dispersion introduces effective

noncentrosymmetry into the scattering from a centrosymmetric crystal (Srinivasan & Parthasarathy, 1976, ch. III; Wilson, 1980a,b; Shmueli & Wilson, 1983). The *bicentric* distribution

$$p(|E|) d|E| = \pi^{-3/2} \exp(-|E|^2/8) K_0(|E|^2/8) d|E| \quad (2.1.5.13)$$

arises, for example, when the 'asymmetric unit in a centrosymmetric crystal is a centrosymmetric molecule' (Lipson & Woolfson, 1952); $K_0(x)$ is a modified Bessel function of the second kind. There are higher hypercentric, hyperparallel and sesquicentric analogues (Wilson, 1952; Rogers & Wilson, 1953; Wilson, 1956). The ideal subcentric and bicentric distributions are expressed in terms of known functions, but the higher hypercentric and the sesquicentric distributions have so far been studied only through their moments and integral representations. Certain hypersymmetric distributions can be expressed in terms of Meijer's G functions (Wilson, 1987b).

2.1.5.5. Relation to distributions of I

When only the intrinsic probability distributions are being considered, it does not greatly matter whether the variable chosen is the intensity of reflection (I), or its positive square root, the modulus of the structure factor ($|F|$), since both are necessarily real and non-negative. In an obvious notation, the relation between the intensity distribution and the structure-factor distribution is

$$p_I(I) = (1/2)I^{-1/2} p_{|F|}(I^{1/2}) \quad (2.1.5.14)$$

or

$$p_{|F|}(|F|) = 2|F| p_I(|F|^2). \quad (2.1.5.15)$$

Statistical fluctuations in counting rates, however, introduce a small but finite probability of negative observed intensities (Wilson, 1978a, 1980a) and thus of imaginary structure factors. This practical complication is treated in *IT C* (1999, Parts 7 and 8).

Both the ideal centric and acentric distributions are simple members of the family of gamma distributions, defined by

$$\gamma_n(x) dx = [\Gamma(n)]^{-1} x^{n-1} \exp(-x) dx, \quad (2.1.5.16)$$

where n is a parameter, not necessarily integral, and $\Gamma(n)$ is the gamma function. Thus the ideal acentric intensity distribution is

$$p(I) dI = \exp(-I/\Sigma) d(I/\Sigma) \quad (2.1.5.17)$$

$$= \gamma_1(I/\Sigma) d(I/\Sigma) \quad (2.1.5.18)$$

and the ideal centric intensity distribution is

$$p(I) dI = (2\Sigma/\pi)^{1/2} \exp[-I/(2\Sigma)] d[I/(2\Sigma)] \quad (2.1.5.19)$$

$$= \gamma_{1/2}[I/(2\Sigma)] d[I/(2\Sigma)]. \quad (2.1.5.20)$$

The properties of gamma distributions and of the related beta distributions, summarized in Table 2.1.5.1, are used in Section 2.1.6 to derive the probability density functions of sums and of ratios of intensities drawn from one of the ideal distributions.

2.1.5.6. Cumulative distribution functions

The integral of the probability density function $f(x)$ from the lower end of its range up to an arbitrary value x is called the cumulative probability distribution, or simply the distribution function, $F(x)$, of x . It can always be written

$$F(x) = \int_{-\infty}^x f(u) du; \quad (2.1.5.21)$$

if the lower end of its range is not actually $-\infty$ one takes $f(x)$ as identically zero between $-\infty$ and the lower end of its range. For the distribution of A [equation (2.1.5.4) or (2.1.5.9)] the lower limit is in

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

Table 2.1.5.1. *Some properties of gamma and beta distributions*

If x_1, x_2, \dots, x_n are independent gamma-distributed variables with parameters p_1, p_2, \dots, p_n , their sum is a gamma-distributed variable with $p = p_1 + p_2 + \dots + p_n$.

If x and y are independent gamma-distributed variables with parameters p and q , then the ratio $u = x/y$ has the distribution $\beta_2(u; p, q)$.

With the same notation, the ratio $v = x/(x+y)$ has the distribution $\beta_1(v; p, q)$.

Differences and products of gamma-distributed variables do not lead to simple results. For proofs, details and references see Kendall & Stuart (1977).

Name of the distribution, its functional form, mean and variance
<p>Gamma distribution with parameter p:</p> $\gamma_p(x) = [\Gamma(x)]^{-1} x^{p-1} \exp(-x); \quad p \leq x < \infty, \quad p > 0$ <p>mean: $\langle x \rangle = p$; variance: $\langle (x - \langle x \rangle)^2 \rangle = p$.</p>
<p>Beta distribution of first kind with parameters p and q:</p> $\beta_1(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1}; \quad 0 \leq x \leq \infty, \quad p, q > 0$ <p>mean: $\langle x \rangle = p/(p+q)$;</p> <p>variance: $\langle (x - \langle x \rangle)^2 \rangle = pq/[(p+q)^2(p+q+1)]$.</p>
<p>Beta distribution of second kind with parameters p and q:</p> $\beta_2(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1+x)^{-p-q}; \quad 0 \leq x < \infty, \quad p, q > 0$ <p>mean: $\langle x \rangle = p/(q-1)$;</p> <p>variance: $\langle (x - \langle x \rangle)^2 \rangle = p(p+q-1)/[(q-1)(q-2)]$.</p>

fact $-\infty$; for the distribution of $|F|$, $|E|$, I and I/Σ the lower end of the range is zero. In such cases, equation (2.1.5.21) becomes

$$F(x) = \int_0^x f(x) dx. \quad (2.1.5.22)$$

In crystallographic applications the cumulative distribution is usually denoted by $N(x)$, rather than by the capital letter corresponding to the probability density function designation. The cumulative forms of the ideal acentric and centric distributions (Howells *et al.*, 1950) have found many applications. For the acentric distribution of $|E|$ [equation (2.1.5.8)] the integration is readily carried out:

$$N(|E|) = 2 \int_0^{|E|} y \exp(-y^2) dy = 1 - \exp(-|E|^2). \quad (2.1.5.23)$$

The integral for the centric distribution of $|E|$ [equation (2.1.5.11)] cannot be expressed in terms of elementary functions, but the integral required has so many important applications in statistics that it has been given a special name and symbol, the error function $\text{erf}(x)$, defined by

$$\text{erf}(x) = (2/\pi^{1/2}) \int_0^x \exp(-t^2) dt. \quad (2.1.5.24)$$

For the centric distribution, then

$$N(|E|) = (2/\pi)^{1/2} \int_0^{|E|} y \exp(-y^2/2) dy \quad (2.1.5.25)$$

$$= \text{erf}(|E|/2^{1/2}). \quad (2.1.5.26)$$

The error function is extensively tabulated [see *e.g.* Abramowitz & Stegun (1972), pp. 310–311, and a closely related function on pp. 966–973].

2.1.6. Distributions of sums, averages and ratios

2.1.6.1. Distributions of sums and averages

In Section 2.1.2.1, it was shown that the average intensity of a sufficient number of reflections is Σ [equation (2.1.2.4)]. When the number of reflections is not 'sufficient', their mean value will show statistical fluctuations about Σ ; such statistical fluctuations are in addition to any systematic variation resulting from non-independence of atomic positions, as discussed in Sections 2.1.2.1–2.1.2.3. We thus need to consider the probability density functions of sums like

$$J_n = \sum_{i=1}^n G_i, \quad (2.1.6.1)$$

and averages like

$$Y = J_n/n, \quad (2.1.6.2)$$

where G_i is the intensity of the i th reflection. The probability density distributions are easily obtained from a property of gamma distributions: If x_1, x_2, \dots, x_n are independent gamma-distributed variables with parameters p_1, p_2, \dots, p_n , their sum is a gamma-distributed variable with parameter p equal to the sum of the parameters. The sum of n intensities drawn from an acentric distribution thus has the distribution

$$p(J_n) dJ_n = \gamma_n(J_n/\Sigma) d(J_n/\Sigma); \quad (2.1.6.3)$$

the parameters of the variables added are all equal to unity, so that their sum is p . Similarly, the sum of n intensities drawn from a centric distribution has the distribution

$$p(J_n) dJ_n = \gamma_{n/2}[J_n/(2\Sigma)] d[J_n/(2\Sigma)]; \quad (2.1.6.4)$$

each parameter has the value of one-half. The corresponding distributions of the averages of n intensities are then

$$p(Y) dY = \gamma_n(nY/\Sigma) d(nY/\Sigma) \quad (2.1.6.5)$$

for the acentric case, and

$$p(Y) dY = \gamma_{n/2}[nY/(2\Sigma)] d[nY/(2\Sigma)] \quad (2.1.6.6)$$

for the centric. In both cases the expected value of Y is Σ and the variances are Σ^2/n and $2\Sigma^2/n$, respectively, just as would be expected.

2.1.6.2. Distribution of ratios

Ratios like

$$S_{n,m} = J_n/K_m, \quad (2.1.6.7)$$

where J_n is given by equation (2.1.6.1),

$$K_m = \sum_{j=1}^m H_j, \quad (2.1.6.8)$$

and the H_j 's are the intensities of a set of reflections (which may or may not overlap with those included in J_n), are used in correlating intensities measured under different conditions. They arise in

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

correlating reflections on different layer lines from the same or different specimens, in correlating the same reflections from different crystals, in normalizing intensities to the local average or to Σ , and in certain systematic trial-and-error methods of structure determination (see Rabinovich & Shakked, 1984, and references therein). There are three main cases:

(i) G_i and H_i refer to the *same* reflection; for example, they might be the observed and calculated quantities for the hkl reflection measured under different conditions or for different crystals of the same substance; or

(ii) G_i and H_i are *unrelated*; for example, the observed and calculated values for the hkl reflection for a completely wrong trial structure, or values for entirely different reflections, as in reducing photographic measurements on different layer lines to the same scale; or

(iii) the G_i 's are a subset of the H_i 's, so that $G_i = H_i$ for $i < n$ and $m > n$.

Aside from the scale factor, in case (i) G_i and H_i will differ chiefly through relatively small statistical fluctuations and uncorrected systematic errors, whereas in case (ii) the differences will be relatively large because of the inherent differences in the intensities. Here we are concerned only with cases (ii) and (iii); the practical problems of case (i) are postponed to *IT C* (1999).

There is little in the crystallographic literature concerning the probability distribution of sums like (2.1.6.1) or ratios like (2.1.6.7); certain results are reviewed by Srinivasan & Parthasarathy (1976, ch. 5), but with a bias toward partially related structures that makes it difficult to apply them to the immediate problem.

In case (ii) (G_i and H_i independent), acentric distribution, Table 2.1.5.1 gives the distribution of the ratio

$$u = nY/(mZ) \quad (2.1.6.9)$$

$$p(u) du = \beta_2[nY/(mZ); n, m] d[nY/(mZ)], \quad (2.1.6.10)$$

where β_2 is a beta distribution of the second kind, Y is given by equation (2.1.6.2) and Z by

$$Z = K_m/m, \quad (2.1.6.11)$$

where n is the number of intensities included in the numerator and m is the number in the denominator. The expected value of Y/Z is then

$$\langle Y/Z \rangle = \frac{m}{m-1} = 1 + \frac{1}{m} + \dots \quad (2.1.6.12)$$

with variance

$$\sigma^2 = \frac{(n+m-1)m^2}{(m-1)^2(m-2)n}. \quad (2.1.6.13)$$

One sees that Y/Z is a biased estimate of the scaling factor between two sets of intensities and the bias, of the order of m^{-1} , depends only on the number of intensities averaged in the denominator. This may seem odd at first sight, but it becomes plausible when one remembers that the mean of a quantity is an unbiased estimator of itself, but the reciprocal of a mean is not an unbiased estimator of the mean of a reciprocal. The mean exists only if $m > 1$ and the variance only for $m > 2$.

In the centric case, the expression for the distribution of the ratio of the two means Y and Z becomes

$$p(u) du = \beta_2[nY/(mZ); n/2, m/2] d[nY/(mZ)] \quad (2.1.6.14)$$

with the expected value of Y/Z equal to

$$\langle Y/Z \rangle = \frac{m}{m-2} = 1 + \frac{2}{m} + \dots \quad (2.1.6.15)$$

and with its variance equal to

$$\sigma^2 = \frac{2(n+m-2)m^2}{(m-2)^2(m-4)n}. \quad (2.1.6.16)$$

For the same number of reflections, the bias in $\langle Y/Z \rangle$ and the variance for the centric distribution are considerably larger than for the acentric. For both distributions the variance of the scaling factor approaches zero when n and m become large. The variances are large for m small, in fact 'infinite' if the number of terms averaged in the denominator is sufficiently small. These biases are readily removed by multiplying Y/Z by $(m-1)/m$ or $(m-2)/m$. Many methods of estimating scaling factors – perhaps most – also introduce bias (Wilson, 1975; Lomer & Wilson, 1975; Wilson, 1976, 1978c) that is not so easily removed. Wilson (1986a) has given reasons for supposing that the bias of the ratio (2.1.6.7) approximates to

$$1 + \frac{\sigma^2(I)}{m\langle I \rangle^2}, \quad (2.1.6.17)$$

whatever the intensity distribution. Equations (2.1.6.12) and (2.1.6.15) are consistent with this.

2.1.6.3. Intensities scaled to the local average

When the G_i 's are a subset of the H_i 's, the beta distributions of the second kind are replaced by beta distributions of the first kind, with means and variances readily found from Table 2.1.5.1. The distribution of such a ratio is chiefly of interest when Y relates to a single reflection and Z relates to a group of m intensities including Y . This corresponds to normalizing intensities to the local average. Its distribution is

$$p(I/\langle I \rangle) d(I/\langle I \rangle) = \beta_1(I/n\langle I \rangle; 1, n-1) d(I/n\langle I \rangle) \quad (2.1.6.18)$$

in the acentric case, with an expected value of $I/\langle I \rangle$ of unity; there is no bias, as is obvious *a priori*. The variance of $I/\langle I \rangle$ is

$$\sigma^2 = \frac{n-1}{n+1}, \quad (2.1.6.19)$$

which is less than the variance of the intensities normalized to an 'infinite' population by a fraction of the order of $2/n$. Unlike the variance of the scaling factor, the variance of the normalized intensity approaches unity as n becomes large. For intensities having a centric distribution, the distribution normalized to the local average is given by

$$p(I/\langle I \rangle) d(I/\langle I \rangle) = \beta_1[I/n\langle I \rangle; 1/2, (n-1)/2] d(I/n\langle I \rangle), \quad (2.1.6.20)$$

with an expected value of $I/\langle I \rangle$ of unity and with variance

$$\sigma^2 = \frac{2(n-1)}{n+2}, \quad (2.1.6.21)$$

less than that for an 'infinite' population by a fraction of about $3/n$.

Similar considerations apply to intensities normalized to Σ in the usual way, since they are equal to those normalized to $\langle I \rangle$ multiplied by $\langle I \rangle/\Sigma$.

2.1.6.4. The use of normal approximations

Since J_n and K_m [equations (2.1.6.1) and (2.1.6.8)] are sums of identically distributed variables conforming to the conditions of the central-limit theorem, it is tempting to approximate their distributions by normal distributions with the correct mean and variance. This would be reasonably satisfactory for the distributions of J_n and K_m themselves for quite small values of n and m , but unsatisfactory for the distribution of their ratio for any values of n and m , even large. The ratio of two variables with normal distributions is

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

notorious for its rather indeterminate mean and 'infinite' variance, resulting from the 'tail' of the denominator distributions extending through zero to negative values. The leading terms of the ratio distribution are given by Kendall & Stuart (1977, p. 288).

2.1.7. Non-ideal distributions: the correction-factor approach

2.1.7.1. Introduction

The probability density functions (p.d.f.'s) of the magnitude of the structure factor, presented in Section 2.1.5, are based on the central-limit theorem discussed above. In particular, the centric and acentric p.d.f.'s given by equations (2.1.5.11) and (2.1.5.8), respectively, are expected to account for the statistical properties of diffraction patterns obtained from crystals consisting of nearly equal atoms, which obey the fundamental assumptions of uniformity and independence of the atomic contributions and are not affected by noncrystallographic symmetry and dispersion. It is also assumed there that the number of atoms in the asymmetric unit is large. Distributions of structure-factor magnitudes which are based on the central-limit theorem, and thus obey the above assumptions, have been termed 'ideal', and the subjects of the following sections are those distributions for which some of the above assumptions/restrictions are not fulfilled; the latter distributions will be called 'non-ideal'.

We recall that the assumption of uniformity consists of the requirement that the fractional part of the scalar product $hx + ky + lz$ be uniformly distributed over the $[0, 1]$ interval, which holds well if x, y, z are rationally independent (Hauptman & Karle, 1953), and permits one to regard the atomic contribution to the structure factor as a random variable. This is of course a necessary requirement for any statistical treatment. If, however, the atomic composition of the asymmetric unit is widely heterogeneous, the structure factor is then a sum of unequally distributed random variables and the Lindeberg-Lévy version of the central-limit theorem (cf. Section 2.1.4.4) cannot be expected to apply. Other versions of this theorem might still predict a normal p.d.f. of the sum, but at the expense of a correspondingly large number of terms/atoms. It is well known that atomic heterogeneity gives rise to severe deviations from ideal behaviour (e.g. Howells *et al.*, 1950) and one of the aims of crystallographic statistics has been the introduction of a correct dependence on the atomic composition into the non-ideal p.d.f.'s [for a review of the early work on non-ideal distributions see Srinivasan & Parthasarathy (1976)]. A somewhat less well known fact is that the dependence of the p.d.f.'s of $|E|$ on space-group symmetry becomes more conspicuous as the composition becomes more heterogeneous (e.g. Shmueli, 1979; Shmueli & Wilson, 1981). Hence both the composition and the symmetry dependence of the intensity statistics are of interest. Other problems, which likewise give rise to non-ideal p.d.f.'s, are the presence of heavy atoms in (variable) special positions, heterogeneous structures with complete or partial noncrystallographic symmetry, and the presence of outstandingly heavy dispersive scatterers.

The need for theoretical representations of non-ideal p.d.f.'s is exemplified in Fig. 2.1.7.1(a), which shows the ideal centric and acentric p.d.f.'s together with a frequency histogram of $|E|$ values, recalculated for a centrosymmetric structure containing a platinum atom in the asymmetric unit of $P\bar{1}$ (Faggiani *et al.*, 1980). Clearly, the deviation from the Gaussian p.d.f., predicted by the central-limit theorem, is here very large and a comparison with the possible ideal distributions can (in this case) lead to wrong conclusions.

Two general approaches have so far been employed in derivations of non-ideal p.d.f.'s which account for the above-mentioned problems: the correction-factor approach, to be dealt

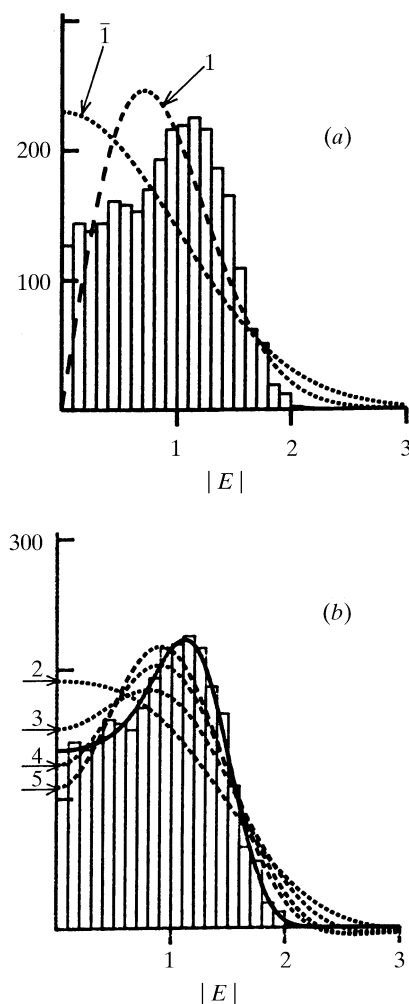


Fig. 2.1.7.1. Atomic heterogeneity and intensity statistics. The histogram appearing in (a) and (b) was constructed from $|E|$ values which were recalculated from atomic parameters published for the centrosymmetric structure of $C_6H_{18}Cl_2N_4O_4Pt$ (Faggiani *et al.*, 1980). The space group of the crystal is $P\bar{1}$, $Z = 2$, i.e. all the atoms are located in general positions. (a) A comparison of the recalculated distribution of $|E|$ with the ideal centric [equation (2.1.5.11)] and acentric [equation (2.1.5.8)] p.d.f.'s, denoted by $\bar{1}$ and 1 , respectively. (b) The same recalculated histogram along with the centric correction-factor p.d.f. [equation (2.1.7.5)], truncated after two, three, four and five terms (dashed lines), and with that accurately computed for the correct space-group Fourier p.d.f. [equations (2.1.8.5) and (2.1.8.22)] (solid line).

with in the following sections, and the more recently introduced Fourier method, to which Section 2.1.8 is dedicated. In what follows, we introduce briefly the mathematical background of the correction-factor approach, apply this formalism to centric and acentric non-ideal p.d.f.'s, and present the numerical values of the moments of the trigonometric structure factor which permit an approximate evaluation of such p.d.f.'s for all the three-dimensional space groups.

2.1.7.2. Mathematical background

Suppose that $p(x)$ is a p.d.f. which accurately describes the experimental distribution of the random variable x , where x is related to a sum of random variables and can be assumed to obey (to some approximation) an ideal p.d.f., say $p^{(0)}(x)$, based on the central-limit theorem. In the correction-factor approach we seek to represent $p(x)$ as

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

$$p(x) = p^{(0)}(x) \sum_k d_k f_k(x), \quad (2.1.7.1)$$

where d_k are coefficients which depend on the cause of the deviation of $p(x)$ from the central-limit theorem approximation and $f_k(x)$ are suitably chosen functions of x . A choice of the set $\{f_k\}$ is deemed suitable, if only from a practical point of view, if it allows the convenient introduction of the cause of the above deviation of $p(x)$ into the expansion coefficients d_k . This requirement is satisfied – also from a theoretical point of view – by taking $f_k(x)$ as a set of polynomials which are orthogonal with respect to the ideal p.d.f., taken as their weight function (e.g. Cramér, 1951). That is, the functions $f_k(x)$ so chosen have to obey the relationship

$$\int_a^b f_k(x) f_m(x) p^{(0)}(x) dx = \delta_{km} = \begin{cases} 1, & \text{if } 2k = m \\ 0, & \text{if } 2k \neq m \end{cases}, \quad (2.1.7.2)$$

where $[a, b]$ is the range of existence of all the functions involved. It can be readily shown that the coefficients d_k are given by

$$d_k = \int_a^b f_k(x) p(x) dx = \langle f_k(x) \rangle = \sum_{n=0}^k c_n^{(k)} \langle x^n \rangle, \quad (2.1.7.3)$$

where the brackets $\langle \rangle$ in equation (2.1.7.3) denote averaging with respect to the *unknown* p.d.f. $p(x)$ and $c_n^{(k)}$ is the coefficient of the n th power of x in the polynomial $f_k(x)$. The coefficients d_k are thus directly related to the moments of the non-ideal distribution and the coefficients of the powers of x in the orthogonal polynomials. The latter coefficients can be obtained by the Gram–Schmidt procedure (e.g. Spiegel, 1974), or by direct use of the Szegő determinants (e.g. Cramér, 1951), for any weight function that has finite moments. However, the feasibility of the present approach depends on our ability to obtain the moments $\langle x^n \rangle$ *without* the knowledge of the non-ideal p.d.f., $p(x)$.

2.1.7.3. Application to centric and acentric distributions

We shall summarize here the non-ideal centric and acentric distributions of the magnitude of the normalized structure factor E (e.g. Shmueli & Wilson, 1981; Shmueli, 1982). We assume that (i) all the atoms are located in general positions and have rationally independent coordinates, (ii) all the scatterers are dispersionless, and (iii) there is no noncrystallographic symmetry. Arbitrary atomic composition and space-group symmetry are admitted. The appropriate weight functions and the corresponding orthogonal polynomials are

$p^{(0)}(E)$	$f_k(x)$	Non-ideal distribution
$(2/\pi)^{1/2} \exp(- E ^2/2)$	$He_{2k}(E)/[(2k)!]^{1/2}$	Centric
$2 E \exp(- E ^2)$	$L_k(E ^2)$	Acentric

(2.1.7.4)

where He_k and L_k are Hermite and Laguerre polynomials, respectively, as defined, for example, by Abramowitz & Stegun (1972). Equations (2.1.7.2), (2.1.7.3) and (2.1.7.4) suffice for the general formulation of the above non-ideal p.d.f.'s of $|E|$. Their full derivation entails (i) the expression of a sufficient number of moments of $|E|$ in terms of absolute moments of the trigonometric structure factor (e.g. Shmueli & Wilson, 1981; Shmueli, 1982) and (ii) calculation of the latter moments for the various symmetries (Wilson, 1978b; Shmueli & Kaldor, 1981, 1983). The notation below is similar to that employed by Shmueli (1982).

These non-ideal p.d.f.'s of $|E|$, for which the first five expansion terms are available, are given by

$$p_c(|E|) = p_c^{(0)}(|E|) \left[1 + \sum_{k=2}^{\infty} \frac{A_{2k}}{(2k)!} He_{2k}(|E|) \right] \quad (2.1.7.5)$$

and

$$p_a(|E|) = p_a^{(0)}(|E|) \left[1 + \sum_{k=2}^{\infty} \frac{(-1)^k B_{2k}}{k!} L_k(|E|^2) \right] \quad (2.1.7.6)$$

for centrosymmetric and noncentrosymmetric space groups, respectively, where $p_c^{(0)}(|E|)$ and $p_a^{(0)}(|E|)$ are the ideal centric and acentric p.d.f.'s [see (2.1.7.4)] and the unified form of the coefficients A_{2k} and B_{2k} , for $k = 2, 3, 4$ and 5 , is

$$\begin{aligned} A_4 \quad \text{or} \quad B_4 &= a_4 Q_4 \\ A_6 \quad \text{or} \quad B_6 &= a_6 Q_6 \\ A_8 \quad \text{or} \quad B_8 &= a_8 Q_8 + U(a_4^2 Q_4^2 - \gamma_4^2) \\ A_{10} \quad \text{or} \quad B_{10} &= a_{10} Q_{10} + V(a_4 a_6 Q_4 Q_6 - \gamma_4 \gamma_6 Q_{10}) \\ &\quad + W \gamma_4^2 Q_{10} \end{aligned} \quad (2.1.7.7)$$

(Shmueli, 1982), where $U = 35$ or 18 , $V = 210$ or 100 and $W = 3150$ or 900 according as A_{2k} or B_{2k} is required, respectively, and the other quantities in equation (2.1.7.7) are given below. The *composition-dependent* terms in equations (2.1.7.7) are

$$Q_{2k} = \frac{\sum_{j=1}^m f_j^{2k}}{(\sum_{n=1}^m f_n^2)^k}, \quad (2.1.7.8)$$

where m is the number of atoms in the asymmetric unit, f_j , $j = 1, \dots, m$ are their scattering factors, and the *symmetry* dependence is expressed by the coefficients a_{2k} in equation (2.1.7.7), as follows:

$$a_{2k} = (-1)^{k-1} (k-1)! \alpha_{k0} + \sum_{p=2}^k (-1)^{k-p} (k-p)! \alpha_{kp} \gamma_{2p}, \quad (2.1.7.9)$$

where

$$\alpha_{kp} = \binom{k}{p} \frac{(2k-1)!!}{(2p-1)!!} \quad \text{or} \quad \binom{k}{p} \frac{k!}{p!} \quad (2.1.7.10)$$

according as the space group is centrosymmetric or noncentrosymmetric, respectively, and γ_{2p} in equation (2.1.7.9) is given by

$$\gamma_{2p} = \frac{\langle |T|^{2p} \rangle}{\langle |T|^2 \rangle^p}, \quad (2.1.7.11)$$

where $\langle |T|^k \rangle$ is the k th absolute moment of the trigonometric structure factor

$$T(\mathbf{h}) = \sum_{s=1}^g \exp[2\pi i \mathbf{h}^T (\mathbf{P}_s \mathbf{r} + \mathbf{t}_s)] \equiv \xi(\mathbf{h}) + i\eta(\mathbf{h}). \quad (2.1.7.12)$$

In equation (2.1.7.12), g is the number of general equivalent positions listed in *ITA* (1983) for the space group in question, times the multiplicity of the Bravais lattice, $(\mathbf{P}_s, \mathbf{t}_s)$ is the s th space-group operator and \mathbf{r} is an atomic position vector.

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

Table 2.1.7.1. *Some even absolute moments of the trigonometric structure factor*

The symbols p , q , r and s denote the second, fourth, sixth and eighth absolute moments of the trigonometric structure factor T [equation (2.1.7.12)], respectively, and the columns of the table contain (for some conciseness) p , q , r/p and s/p^2 . The numbers in parentheses, appearing beside some space-group entries, refer to hkl subsets which are defined in the note at the end of the table. These subset references are identical with those given by Shmueli & Kaldor (1981, 1983). The symbols q , r and s are also equivalent to $\gamma_4 P^2$, $\gamma_6 P^3$ and $\gamma_8 P^4$, respectively, where γ_{2n} are the normalized absolute moments given by equation (2.1.7.11).

Space groups(s)	p	q	r/p	s/p^2
Point group: 1 $P1$	1	1	1	1
Point group: $\bar{1}$ $P\bar{1}$	2	6	10	17
Point groups: 2, m All P All C	2 4	6 48	10 160	17 560
Point group: 2/m All P All C	4 8	36 288	100 1600	$306\frac{1}{4}$ 9800
Point group: 222 All P All C and I $F222$	4 8 16	28 224 1792	64 1024 16384	$169\frac{3}{4}$ 5432 173824
Point group: $mm2$ All P All A , C and I $Fmm2$ $Fdd2$ (1) $Fdd2$ (2)	4 8 16 16 16	36 288 2304 2304 1280	100 1600 25600 25600 7168	$306\frac{1}{4}$ 9800 313600 313600 43264
Point group: mmm All P All C and I $Fmmm$ $Fddd$ (1) $Fddd$ (2)	8 16 32 32 32	216 1728 13824 13824 7680	1000 16000 256000 256000 71680	$5359\frac{3}{8}$ 171500 5488000 5488000 757120
Point group: 4 $P4$, $P4_2$ $P4_1^*$ (3) $P4_1^*$ (4) $I4$ $I4_1$ (5) $I4_1$ (6)	4 4 4 8 8 8	36 36 20 288 288 160	100 100 28 1600 1600 448	$306\frac{1}{4}$ $306\frac{1}{4}$ $42\frac{1}{4}$ 9800 9800 1352
Point group: $\bar{4}$ $P\bar{4}$ $I\bar{4}$	4 8	28 224	64 1024	$169\frac{3}{4}$ 5432
Point group: 4/m All P $I4/m$ $I4_1/a$ (7) $I4_1/a$ (8)	8 16 16 16	216 1728 1728 960	1000 16000 16000 4480	$5359\frac{3}{8}$ 171500 171500 23660
Point group: 422 $P422$, $P4_212$, $P4_222$, $P4_2212$ $P122$,* $P4_12_12^*$ (3) $P4_122$,* $P4_12_12^*$ (4) $I422$ $I4_122$ (7) $I4_122$ (8)	8 8 8 16 16 16	136 136 104 1088 1088 832	424 424 208 6784 6784 3328	$1682\frac{1}{8}$ $1682\frac{1}{8}$ $470\frac{1}{8}$ 53828 53828 15044

Space groups(s)	p	q	r/p	s/p^2
Point group: 4mm All P $I4mm$, $I4cm$ $I4_1md$, $I4_1cd$ (7) $I4_1md$, $I4_1cd$ (8)	8 16 16 16	168 1344 1344 832	640 10240 10240 3328	$2970\frac{5}{5}$ 95060 95060 15188
Point groups: $\bar{4}2m$, $4m2$ All P $I\bar{4}m2$, $I\bar{4}2m$, $I\bar{4}c2$ $I\bar{4}2d$ (5) $I\bar{4}2d$ (6)	8 16 16 16	136 1088 1088 832	424 6784 6784 3328	$1682\frac{1}{8}$ 53828 53828 15044
Point group: 4/mmm All P $I4/mmm$, $I4/mcm$ $I4_1/amd$, $I4_1/acd$ (5) $I4_1/amd$, $I4_1/acd$ (6)	16 32 32 32	1008 8064 8064 4992	6400 102400 102400 33280	$51985\frac{15}{16}$ 1663550 1663550 265790
Point group: 3 All P and R	3	15	31	71
Point group: $\bar{3}$ All P and R	6	90	310	1242
Point group: 32 All P and R	6	66	166	508
Point group: 3m $P3m1$, $P31m$, $R3m$ $P3c1$, $P31c$, (3); $R3c$ (1) $P3c1$, $P31c$, (4); $R3c$ (2)	6 6 6	66 66 66	178 178 154	604 604 412
Point group: $\bar{3}m$ $P\bar{3}1m$, $P\bar{3}m1$, $R\bar{3}m$ $P\bar{3}1c$, $P\bar{3}c1$ (3); $R\bar{3}c$ (1) $P\bar{3}1c$, $P\bar{3}c1$ (4); $R\bar{3}c$ (2)	12 12 12	396 396 396	1780 1780 1540	$10578\frac{3}{4}$ $10578\frac{3}{4}$ $7218\frac{3}{4}$
Point group: 6 $P6$ $P6_1^*$ (9) $P6_1^*$ (10) $P6_1^*$ (11) $P6_1^*$ (12) $P6_2^*$ (13) $P6_2^*$ (14) $P6_3$ (3) $P6_3$ (4)	6 6 6 6 6 6 6 6 6	90 90 54 54 90 90 54 90 90	340 340 91 97 280 340 97 340 280	1522 1522 161 193 962 1522 193 1522 962
Point group: $\bar{6}$ $P\bar{6}$	6	90	310	1242
Point group: 6/m $P6/m$ $P6_3/m$ (3)	12 12	540 540	3400 3400	$26643\frac{3}{4}$ $26643\frac{3}{4}$

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

Table 2.1.7.1. Some even absolute moments of the trigonometric structure factor (cont.)

Space groups(s)	p	q	r/p	s/p^2	Space groups(s)	p	q	r/p	s/p^2
$P6_3/m$ (4)	12	540	2800	$16843\frac{3}{4}$	Point group: 432				
Point group: 622					$P4_32, P4_232$	24	1272	4648	$25216\frac{7}{8}$
$P622$	12	324	1150	$5506\frac{1}{4}$	$P4_132^* (15)$	24	1272	4648	$25216\frac{7}{8}$
$P6_122^* (9)$	12	324	1150	$5506\frac{1}{4}$	$P4_132^* (16)$	24	1176	3568	$13916\frac{7}{8}$
$P6_122^* (10)$	12	252	577	$1537\frac{3}{4}$	$P4_132^* (17)$	24	1080	2776	$8664\frac{7}{8}$
$P6_122^* (11)$	12	252	583	$1601\frac{3}{4}$	$P4_132^* (18)$	24	984	2272	$6580\frac{7}{8}$
$P6_122^* (12)$	12	324	1090	$4746\frac{1}{4}$	$I432$	48	10176	74368	806940
$P6_222^* (13)$	12	324	1150	$5506\frac{1}{4}$	$I4_132 (15)$	48	10176	74368	806940
$P6_222^* (14)$	12	252	583	$1601\frac{3}{4}$	$I4_132 (17)$	48	8640	44416	277276
$P6_322 (3)$	12	324	1150	$5506\frac{1}{4}$	$F432$	96	81408	1189888	25822080
$P6_322 (4)$	12	324	1090	$4746\frac{1}{4}$	$F4_132 (15)$	96	81408	1189888	25822080
Point group: 6mm					$F4_132 (18)$	96	62976	581632	6738816
$P6mm$	12	396	1930	$12818\frac{3}{4}$	Point group: $\bar{4}3m$				
$P6cc (3)$	12	396	1930	$12818\frac{3}{4}$	$P\bar{4}3m$	24	1272	5128	$32896\frac{7}{8}$
$P6cc (4)$	12	396	1450	$6098\frac{3}{4}$	$P\bar{4}3n (1)$	24	1272	5128	$32896\frac{7}{8}$
$P6_3cm, P6_3mc (3)$	12	396	1930	$12818\frac{3}{4}$	$P\bar{4}3n (2)$	24	1272	4168	$17536\frac{7}{8}$
$P6_3cm, P6_3mc (4)$	12	396	1630	$8338\frac{3}{4}$	$I\bar{4}3m$	48	10176	82048	1052700
Point groups: $6m2, 62m$					$I\bar{4}3d (15); (20)$	48	10176	82048	1052700
$P6m2, P62m$	12	396	1780	$10578\frac{3}{4}$	$I\bar{4}3d (15); (21)$	48	10176	66688	561180
$P6c2, P62c (3)$	12	396	1780	$10578\frac{3}{4}$	$I\bar{4}3d (17)$	48	8640	44416	277276
$P6c2, P62c (4)$	12	396	1540	$7218\frac{3}{4}$	$F\bar{4}3m$	96	81408	1312768	33686400
Point group: $6/mmm$					$F\bar{4}3c (15)$	96	81408	1312768	33686400
$P6/mmm$	24	2376	19300	$224328\frac{1}{8}$	$F\bar{4}3c (18)$	96	81408	1067008	17957760
$P6/mcc (3)$	24	2376	19300	$224328\frac{1}{8}$	Point group: $m\bar{3}m$				
$P6/mcc (4)$	24	2376	14500	$106728\frac{1}{8}$	$Pm\bar{3}m, Pn\bar{3}m$	48	8784	72160	$972717\frac{13}{16}$
$P6/mcm, P6/mmc (3)$	24	2376	19300	$224328\frac{1}{8}$	$Pn\bar{3}n, Pm\bar{3}n (1)$	48	8784	72160	$972717\frac{13}{16}$
$P6/mcm, P6/mmc (4)$	24	2376	16300	$145928\frac{1}{8}$	$Pn\bar{3}n, Pm\bar{3}n (2)$	48	8784	56800	$488877\frac{13}{16}$
Point group: 23					$Im\bar{3}m$	96	70272	1154560	31126970
$P23, P2_13$	12	276	760	$2695\frac{1}{4}$	$Ia\bar{3}d (15); (20)$	96	70272	1154560	31126970
$I23, I2_13$	24	2208	12160	86248	$Ia\bar{3}d (15); (21)$	96	51840	432640	4497850
$F23$	48	17664	194560	2759936	$Ia\bar{3}d (17)$	96	70272	908800	15644090
Point group: $m\bar{3}$					$Fm\bar{3}m$	192	562176	18472960	996063040
$Pm\bar{3}, Pn\bar{3}, Pa\bar{3}$	24	1800	9400	$67703\frac{1}{8}$	$Fm\bar{3}c (1)$	192	562176	18472960	996063040
$Im\bar{3}, Ia\bar{3}$	48	14400	150400	2166500	$Fm\bar{3}c (2)$	192	562176	14540800	500610880
$Fm\bar{3}$	96	115200	2406400	69328000	$Fd\bar{3}m (1)$	192	562176	18472960	996063040
$Fd\bar{3} (1)$	96	115200	2406400	69328000	$Fd\bar{3}m (2)$	192	414720	7782400	205432640
$Fd\bar{3} (2)$	96	96768	1484800	28183680	$Fd\bar{3}c (1)$	192	562176	18472960	996063040
					$Fd\bar{3}c (2)$	192	414720	6799360	136619840

Note. hkl subsets: (1) $h+k+l=2n$; (2) $h+k+l=2n+1$; (3) $l=2n$; (4) $l=2n+1$; (5) $2h+l=2n$; (6) $2h+l=2n+1$; (7) $2k+l=2n$; (8) $2k+l=2n+1$; (9) $l=6n$; (10) $l=6n+1, 6n+5$; (11) $l=6n+2, 6n+4$; (12) $l=6n+3$; (13) $l=3n$; (14) $l=3n+1, 3n+2$; (15) hkl all even; (16) only one index odd; (17) only one index even; (18) hkl all odd; (19) two indices odd; (20) $h+k+l=4n$; (21) $h+k+l=4n+2$.

* And the enantiomorphous space group.

The cumulative distribution functions, obtained by integrating equations (2.1.7.5) and (2.1.7.6), are given by

$$N_c(|E|) = \operatorname{erf}\left(\frac{|E|}{\sqrt{2}}\right) - \frac{2}{\sqrt{\pi}} \exp\left(-\frac{|E|^2}{2}\right) \times \left[\sum_{k=2}^{\infty} \frac{A_{2k}}{(2k)!} He_{2k-1}(|E|) \right] \quad (2.1.7.13)$$

and

$$N_a(|E|) = 1 - \exp(-|E|^2) + \exp(-|E|^2) \times \left\{ \sum_{k=2}^{\infty} \frac{(-1)^k B_{2k}}{k!} [L_{k-1}(|E|^2) - L_k(|E|^2)] \right\} \quad (2.1.7.14)$$

for centrosymmetric and noncentrosymmetric space groups, respectively, where the coefficients are defined in equations (2.1.7.7)–(2.1.7.12). Note that the first term on the right-hand side

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

Table 2.1.7.2. Closed expressions for γ_{2k} [equation (2.1.7.11)] for space groups of low symmetry

The normalized moments γ_{2k} are expressed in terms of M_k , where

$$M_k = \frac{(2k)!}{2^k(k!)^2} = \frac{(2k-1)!!}{k!},$$

and l' , which takes on the values 1, 2 or 4 according as the Bravais lattice is of type P , one of the types A, B, C or I , or type F , respectively. The expressions for γ_{2k} are identical for all the space groups based on a given point group, except $Fdd2$ and $Fddd$. The expressions are valid for general reflections and under the restrictions given in the text.

Point group(s)	Expression for γ_{2k}
1	1
$\bar{1}, 2, m$	$l'^{k-1} M_k$
$2/m, mm2$	$l'^{k-1} M_k^2$
mmm	$l'^{k-1} M_k^3$
222	$\frac{l'^{k-1}}{2^k(k!)^2} \sum_{p=0}^k (M_p M_{k-p})^3 [p!(k-p)!]^2$

of equation (2.1.7.13) and the first two terms on the right-hand side of equation (2.1.7.14) are just the cumulative distributions derived from the ideal centric and acentric p.d.f.'s in Section 2.1.5.6.

The moments $\langle |T|^{2k} \rangle$ were compiled for all the space groups by Wilson (1978*b*) for $k = 1$ and 2, and by Shmueli & Kaldor (1981, 1983) for $k = 1, 2, 3$ and 4. These results are presented in Table 2.1.7.1. Closed expressions for the normalized moments γ_{2p} were obtained by Shmueli (1982) for the triclinic, monoclinic and orthorhombic space groups except $Fdd2$ and $Fddd$ (see Table 2.1.7.2). The composition-dependent terms, Q_{2k} , are most conveniently computed as weighted averages over the ranges of $(\sin \theta)/\lambda$ which were used in the construction of the Wilson plot for the computation of the $|E|$ values.

2.1.7.4. Fourier versus Hermite approximations

As noted in Section 2.1.8.7 below, the Fourier representation of the probability distribution of $|F|$ is usually much better than the particular orthogonal-function representation discussed in Section 2.1.7.3. Many, perhaps most, non-ideal centric distributions look like slight distortions of the ideal (Gaussian) distribution and have no resemblance to a cosine function. The empirical observation thus seems paradoxical. The probable explanation has been pointed out by Wilson (1986*b*). A truncated Fourier series is a best approximation, in the least-squares sense, to the function represented. The particular orthogonal-function approach used in equation (2.1.7.5), on the other hand, is not a least-squares approximation to $p_c(|E|)$, but is a least-squares approximation to

$$p_c(|E|) \exp(|E|^2/4). \quad (2.1.7.15)$$

The usual expansions (often known as Gram-Charlier or Edgeworth) thus give great weight to fitting the distribution of the (comparatively few) strong reflections, at the expense of a poor fit for the (much more numerous) weak-to-medium ones. Presumably, a similar situation exists for the representation of acentric distributions, but this has not been investigated in detail. Since the centric distributions $p_c(|E|)$ often look nearly Gaussian, one is led to ask if there is an expansion in orthogonal functions that (i) has the leading term $p_c(|E|)$ and (ii) is a least-squares (as well as an

orthogonal-function)* fit to $p_c(|E|)$. One does exist, based on the orthogonal functions

$$f_k = n(x) He_k(2^{1/2}x), \quad (2.1.7.16)$$

where $n(x)$ is the Gaussian distribution (Myller-Lebedeff, 1907). Unfortunately, no reasonably simple relationship between the coefficients d_k and readily evaluated properties of $p_c(|E|)$ has been found, and the Myller-Lebedeff expansion has not, as yet, been applied in crystallography. Although Stuart & Ord (1994, p. 112) dismiss it in a three-line footnote, it does have important applications in astronomy (van der Marel & Franx, 1993; Gerhard, 1993).

2.1.8. Non-ideal distributions: the Fourier method

The starting point of the method described in the previous section is the central-limit theorem approximation, and the method consists of finding correction factors which result in better approximations to the actual p.d.f. Conceptually, this is equivalent to improving the approximation of the characteristic function [cf. equation (2.1.4.10)] over that which led to the central-limit theorem result.

The method to be described in this section does not depend on any initial approximation and will be shown to utilize the dependence of the exact value of the characteristic function on the space-group symmetry, atomic composition and other factors. This approach has its origin in a simple but ingenious observation by Barakat (1974), who noted that if a random variable has lower and upper bounds then the corresponding p.d.f. can be non-zero only within these bounds and can therefore be expanded in an ordinary Fourier series and set to zero (identically) outside the bounded interval. Barakat's (1974) work dealt with intensity statistics of laser speckle, where sinusoidal waves are involved, as in the present problem. This method was applied by Weiss & Kiefer (1983) to testing the accuracy of a steepest-descent approximation to the exact solution of the problem of random walk, and its first application to crystallographic intensity statistics soon followed (Shmueli *et al.*, 1984). Crystallographic (*e.g.* Shmueli & Weiss, 1987; Rabinovich *et al.*, 1991*a,b*) and noncrystallographic (Shmueli *et al.*, 1985; Shmueli & Weiss, 1985*a*; Shmueli, Weiss & Wilson, 1989; Shmueli *et al.*, 1990) symmetry was found to be tractable by this approach, as well as joint conditional p.d.f.'s of several structure factors (Shmueli & Weiss, 1985*b*, 1986; Shmueli, Rabinovich & Weiss, 1989). The Fourier method is illustrated below by deriving the exact counterparts of equations (2.1.7.5) and (2.1.7.6) and specifying them for some simple symmetries. We shall then indicate a method of treating higher symmetries and present results which will suffice for evaluation of Fourier p.d.f.'s of $|E|$ for a wide range of space groups.

2.1.8.1. General representations of p.d.f.'s of $|E|$ by Fourier series

We assume, as before, that (i) the atomic phase factors $\vartheta_j = 2\pi \mathbf{h}^T \mathbf{r}_j$ [cf. equation (2.1.1.2)] are uniformly distributed on $(0-2\pi)$ and (ii) the atomic contributions to the structure factor are independent. For a centrosymmetric space group, with the origin chosen at a centre of symmetry, the random variable is the (real) normalized structure factor E and its bounds are $-E_M$ and E_M , where

* The condition for this desirable property seems to be that the weight function in equation (2.1.7.2) should be unity.

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

$$E_M = \sum_{j=1}^N n_j, \text{ with } n_j = \frac{f_j}{\left(\sum_{k=1}^N f_k^2\right)^{1/2}}. \quad (2.1.8.1)$$

Here, E_M is the maximum possible value of E and f_j is the conventional scattering factor of the j th atom, including its temperature factor. The p.d.f., $p(E)$, can be non-zero in the range $(-E_M, E_M)$ only and can thus be expanded in the Fourier series

$$p(E) = (\alpha/2) \sum_{k=-\infty}^{\infty} C_k \exp(-\pi i k \alpha E), \quad (2.1.8.2)$$

where $\alpha = 1/E_M$. Only the real part of $p(E)$ is relevant. The Fourier coefficients can be obtained in the conventional manner by integrating over the range $(-E_M, E_M)$,

$$C_k = \int_{-E_M}^{E_M} p(E) \exp(\pi i k \alpha E) dE. \quad (2.1.8.3)$$

Since, however, $p(E) = 0$ for $E < -E_M$ and $E > E_M$, it is possible and convenient to replace the limits of integration in equation (2.1.8.3) by infinity. Thus

$$C_k = \int_{-\infty}^{\infty} p(E) \exp(\pi i k \alpha E) dE = \langle \exp(\pi i k \alpha E) \rangle. \quad (2.1.8.4)$$

Equation (2.1.8.4) shows that C_k is a Fourier transform of the p.d.f. $p(E)$ and, as such, it is the value of the corresponding characteristic function at the point $t_k = \pi \alpha k$ [i.e., $C_k = C(\pi \alpha k)$, where the characteristic function $C(t)$ is defined by equation (2.1.4.1)]. It is also seen that C_k is the expected value of the exponential $\exp(\pi i k \alpha E)$. It follows that the feasibility of the present approach depends on one's ability to evaluate the characteristic function in closed form *without* the knowledge of the p.d.f.; this is analogous to the problem of evaluating absolute moments of the structure factor for the correction-factor approach, discussed in Section 2.1.7. Fortunately, in crystallographic applications these calculations are feasible, provided individual isotropic motion is assumed. The formal expression for the p.d.f. of $|E|$, for any centrosymmetric space group, is therefore

$$p(|E|) = \alpha \left[1 + 2 \sum_{k=1}^{\infty} C_k \cos(\pi k \alpha |E|) \right], \quad (2.1.8.5)$$

where use is made of the assumption that $p(E) = p(-E)$, and the Fourier coefficients are evaluated from equation (2.1.8.4).

The p.d.f. of $|E|$ for a noncentrosymmetric space group is obtained by first deriving the joint p.d.f. of the real and imaginary parts of E and then integrating out its phase. The general expression for E is

$$E = A + iB = |E| \cos \varphi + i|E| \sin \varphi, \quad (2.1.8.6)$$

where φ is the phase of E . The required joint p.d.f. is

$$p(A, B) = (\alpha^2/4) \sum_m \sum_n C_{mn} \exp[-\pi i \alpha (mA + nB)], \quad (2.1.8.7)$$

and introducing polar coordinates $m = r \sin \Delta$ and $n = r \cos \Delta$, where $r = \sqrt{m^2 + n^2}$ and $\Delta = \tan^{-1}(m/n)$, we have

$$p(|E|, \varphi) = (\alpha^2/4) |E| \sum_m \sum_n C_{mn} \exp[-\pi i \alpha |E| \times \sqrt{m^2 + n^2} \sin(\varphi + \Delta)]. \quad (2.1.8.8)$$

Integrating out the phase φ , we obtain

$$p(|E|) = (\pi \alpha^2 |E|/2) \sum_m \sum_n C_{mn} J_0(\pi \alpha |E| \sqrt{m^2 + n^2}), \quad (2.1.8.9)$$

where $J_0(x)$ is the Bessel function of the first kind (e.g. Abramowitz & Stegun, 1972). This is a general form of the p.d.f. of $|E|$ for a noncentrosymmetric space group. The Fourier coefficients are obtained, similarly to the above, as

$$C_{mn} = \langle \exp[\pi i \alpha (mA + nB)] \rangle \quad (2.1.8.10)$$

and the average in equation (2.1.8.10), just as that in equation (2.1.8.4), is evaluated in terms of integrals over the appropriate trigonometric structure factors. In terms of the characteristic function for a joint p.d.f. of A and B , the Fourier coefficient in equation (2.1.8.10) is given by $C_{mn} = C(\pi \alpha m, \pi \alpha n)$.

We shall denote the characteristic function by $C(t_1)$ if it corresponds to a Fourier coefficient of a Fourier series for a centrosymmetric space group and by $C(t_1, t_2)$ or by $C(t, \Delta)$, where $t = (t_1^2 + t_2^2)^{1/2}$ and $\Delta = \tan^{-1}(t_1/t_2)$, if it corresponds to a Fourier series for a noncentrosymmetric space group.

2.1.8.2. Fourier-Bessel series

Equations (2.1.8.5) and (2.1.8.9) are the exact counterparts of equations (2.1.7.5) and (2.1.7.6), respectively. The computational effort required to evaluate equation (2.1.8.9) is somewhat greater than that for (2.1.8.5), because a double Fourier series has to be summed. The p.d.f. for *any* noncentrosymmetric space group can be expressed by a double Fourier series, but this can be simplified if the characteristic function depends on $t = (t_1^2 + t_2^2)^{1/2}$ alone, rather than on t_1 and t_2 separately. In such cases the p.d.f. of $|E|$ for a noncentrosymmetric space group can be expanded in a *single* Fourier-Bessel series (Barakat, 1974; Weiss & Kiefer, 1983; Shmueli *et al.*, 1984). The general form of this expansion is

$$p(|E|) = 2\alpha^2 |E| \sum_{u=1}^{\infty} D_u J_0(\alpha \lambda_u |E|), \quad (2.1.8.11)$$

where

$$D_u = \frac{C(\alpha \lambda_u)}{J_1^2(\lambda_u)} \quad (2.1.8.12)$$

and

$$C(\alpha \lambda_u) = \prod_{j=1}^{N/g} C_{ju}, \quad (2.1.8.13)$$

where $J_1(x)$ is the Bessel function of the first kind, and λ_u is the u th root of the equation $J_0(x) = 0$; the atomic contribution C_{ju} to equation (2.1.8.13) is computed as

$$C_{ju} = C(\alpha n_j \lambda_u). \quad (2.1.8.14)$$

The roots λ_u are tabulated in the literature (e.g. Abramowitz & Stegun, 1972), but can be most conveniently computed as follows. The first five roots are given by

$$\lambda_1 = 2.4048255577$$

$$\lambda_2 = 5.5200781103$$

$$\lambda_3 = 8.6537279129$$

$$\lambda_4 = 11.7915344390$$

$$\lambda_5 = 14.9309177085$$

and the higher ones can be obtained from McMahon's approximation (cf. Abramowitz & Stegun, 1972)

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

$$\lambda_u = \beta + \frac{1}{8\beta} - \frac{124}{3(8\beta)^3} + \frac{120928}{15(8\lambda)^5} - \frac{401743168}{105(8\lambda)^7} + \dots, \quad (2.1.8.15)$$

$$= \prod_{j=1}^N J_0(\pi\alpha n_j \sqrt{m^2 + n^2}). \quad (2.1.8.26)$$

where $\beta = (u - \frac{1}{4})\pi$. For $u > 5$ the values given by equation (2.1.8.15) have a relative error less than 10^{-11} so that no refinement of roots of higher orders is needed (Shmueli *et al.*, 1984). Numerical computations of single Fourier–Bessel series are of course faster than those of the double Fourier series, but both representations converge fairly rapidly.

2.1.8.3. Simple examples

Consider the Fourier coefficient of the p.d.f. of $|E|$ for the centrosymmetric space group $P\bar{1}$. The normalized structure factor is given by

$$E = 2 \sum_{j=1}^{N/2} n_j \cos \vartheta_j, \quad \text{with } \vartheta_j = 2\pi \mathbf{h}^T \cdot \mathbf{r}_j, \quad (2.1.8.16)$$

and the Fourier coefficient is

$$C_k = \langle \exp(\pi i k \alpha E) \rangle \quad (2.1.8.17)$$

$$= \left\langle \exp \left[2\pi i k \alpha \sum_{j=1}^{N/2} n_j \cos \vartheta_j \right] \right\rangle \quad (2.1.8.18)$$

$$= \left\langle \prod_{j=1}^{N/2} \exp(2\pi i k \alpha n_j \cos \vartheta_j) \right\rangle \quad (2.1.8.19)$$

$$= \prod_{j=1}^{N/2} \langle \exp(2\pi i k \alpha n_j \cos \vartheta_j) \rangle \quad (2.1.8.20)$$

$$= \prod_{j=1}^{N/2} \left\{ (1/2\pi) \int_{-\pi}^{\pi} \exp(2\pi i k \alpha n_j \cos \vartheta) d\vartheta \right\} \quad (2.1.8.21)$$

$$= \prod_{j=1}^{N/2} J_0(2\pi k \alpha n_j). \quad (2.1.8.22)$$

Equation (2.1.8.20) is obtained from equation (2.1.8.19) if we make use of the assumption of independence, the assumption of uniformity allows us to rewrite equation (2.1.8.20) as (2.1.8.21), and the expression in the braces in the latter equation is just a definition of the Bessel function $J_0(2\pi k \alpha n_j)$ (*e.g.* Abramowitz & Stegun, 1972).

Let us now consider the Fourier coefficient of the p.d.f. of $|E|$ for the noncentrosymmetric space group $P1$. We have

$$A = \sum_{j=1}^N n_j \cos \vartheta_j \quad \text{and} \quad B = \sum_{j=1}^N n_j \sin \vartheta_j. \quad (2.1.8.23)$$

These expressions for A and B are substituted in equation (2.1.8.10), resulting in

$$C_{mn} = \left\langle \prod_{j=1}^N \exp[\pi i \alpha n_j (m \cos \vartheta_j + n \sin \vartheta_j)] \right\rangle \quad (2.1.8.24)$$

$$= \left\langle \prod_{j=1}^N \exp[\pi i \alpha n_j \sqrt{m^2 + n^2} \sin(\vartheta_j + \Delta)] \right\rangle \quad (2.1.8.25)$$

Equation (2.1.8.24) leads to (2.1.8.25) by introducing polar coordinates analogous to those leading to equation (2.1.8.8), and equation (2.1.8.26) is then obtained by making use of the assumptions of independence and uniformity in an analogous manner to that detailed in equations (2.1.8.12)–(2.1.8.22) above.

The right-hand side of equation (2.1.8.26) is to be used as a Fourier coefficient of the double Fourier series given by (2.1.8.9). Since, however, this coefficient depends on $(m^2 + n^2)^{1/2}$ alone rather than on m and n separately, the p.d.f. of $|E|$ for $P1$ can also be represented by a Fourier–Bessel series [*cf.* equation (2.1.8.11)] with coefficient

$$D_u = \frac{1}{J_1^2(\lambda_u)} \prod_{j=1}^N J_0(\alpha n_j \lambda_u), \quad (2.1.8.27)$$

where λ_u is the u th root of the equation $J_0(x) = 0$.

2.1.8.4. A more complicated example

We now illustrate the methodology of deriving characteristic functions for space groups of higher symmetries, following the method of Rabinovich *et al.* (1991*a,b*). The derivation is performed for the space group $P\bar{6}$ [No. 174]. According to Table A1.4.3.6, the real and imaginary parts of the normalized structure factor are given by

$$A = 2 \sum_{j=1}^{N/6} n_j [C(hki)c(lz)]_j$$

$$= 2 \sum_{j=1}^{N/6} n_j \cos \tau_j \sum_{k=1}^3 \cos \alpha_{jk} \quad (2.1.8.28)$$

and

$$B = 2 \sum_{j=1}^{N/6} n_j [S(hki)c(lz)]_j$$

$$= 2 \sum_{j=1}^{N/6} n_j \cos \tau_j \sum_{k=1}^3 \sin \alpha_{jk}, \quad (2.1.8.29)$$

where

$$\alpha_{j1} = 2\pi(hx_j + ky_j),$$

$$\alpha_{j2} = 2\pi(kx_j + iy_j),$$

$$\alpha_{j3} = 2\pi(ix_j + hy_j),$$

$$\tau_j = 2\pi lz_j.$$

Note that $\alpha_{j1} + \alpha_{j2} + \alpha_{j3} = 0$, *i.e.*, one of these contributions depends on the other two; this is a recurring problem in calculations pertaining to trigonal and hexagonal systems. For brevity, we write directly the general form of the characteristic function from which the functional form of the Fourier coefficient can be readily obtained. The characteristic function is given by

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

$$C(t_1, t_2) = \langle \exp[i(t_1A + t_2B)] \rangle \quad (2.1.8.30)$$

$$= \prod_{j=1}^{N/6} \left\langle \exp \left[2in_j \cos \tau_j \sum_{k=1}^3 (t_1 \cos \alpha_{jk} + t_2 \sin \alpha_{jk}) \right] \right\rangle \quad (2.1.8.31)$$

$$= \prod_{j=1}^{N/6} \left\langle \exp \left[2in_j t \cos \tau_j \sum_{k=1}^3 (\sin \Delta \cos \alpha_{jk} + \cos \Delta \sin \alpha_{jk}) \right] \right\rangle \quad (2.1.8.32)$$

$$= \prod_{j=1}^{N/6} \left\langle \exp \left[2in_j t \cos \tau_j \sum_{k=1}^3 \sin(\alpha_{jk} + \Delta) \right] \right\rangle, \quad (2.1.8.33)$$

where $\Delta = \tan^{-1}(t_1/t_2)$, $t = (t_1^2 + t_2^2)^{1/2}$ and the assumption of independence was used. If we further employ the assumption of uniformity, while remembering that the angular variables α_{jk} are not independent, the characteristic function can be written as

$$C(t_1, t_2) = \prod_{j=1}^{N/6} \left\{ (1/2\pi) \int_{-\pi}^{\pi} d\tau \left([1/(2\pi)^2] \right. \right. \\ \times \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} d\alpha_1 d\alpha_2 d\alpha_3 \delta_{2\pi}(\alpha_1 + \alpha_2 + \alpha_3) \\ \left. \left. \times \exp \left[2in_j t \cos \tau \sum_{k=1}^3 \sin(\alpha_k + \Delta) \right] \right) \right\}, \quad (2.1.8.34)$$

where

$$\delta_{2\pi}(\alpha) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \exp(-ik\alpha) \quad (2.1.8.35)$$

is the Fourier representation of the periodic delta function. Equation (2.1.8.34) then becomes

$$C(t_1, t_2) = \prod_{j=1}^{N/6} \left\{ (1/2\pi) \int_{-\pi}^{\pi} d\tau \sum_{k=-\infty}^{\infty} \left[(1/2\pi) \right. \right. \\ \left. \left. \times \int_{-\pi}^{\pi} \exp(-ik\alpha + 2in_j t \cos \tau \sin(\alpha + \Delta)) d\alpha \right]^3 \right\}. \quad (2.1.8.36)$$

If we change the variable α to $\alpha' - \Delta$, $\sin(\alpha + \Delta)$ becomes $\sin \alpha'$ and $-ik\alpha = -ik\alpha' + ik\Delta$. Hence

$$C(t_1, t_2) = \prod_{j=1}^{N/6} \left\{ (1/2\pi) \int_{-\pi}^{\pi} d\tau \sum_{k=-\infty}^{\infty} \exp(3ik\Delta) J_k^3(2n_j t \cos \tau) \right\}. \quad (2.1.8.37)$$

The imaginary part of the summation, involving Bessel functions of odd orders, vanishes upon integration and the latter is restricted to the positive quadrant in τ . Thus, upon replacing cosines by sines (this is permissible at this stage) the atomic contribution to the characteristic function becomes

$$C_j(t, \Delta) = (2/\pi) \int_0^{\pi/2} \left[J_0^3(2n_j t \sin \tau) \right. \\ \left. + 2 \sum_{k=1}^{\infty} \cos(6k\Delta) J_k^3(2n_j t \sin \tau) \right] d\tau \quad (2.1.8.38)$$

and a double Fourier series must be used for the p.d.f.

2.1.8.5. Atomic characteristic functions

Expressions for the atomic contributions to the characteristic functions were obtained by Rabinovich *et al.* (1991a) for a wide range of space groups, by methods similar to those described above. These expressions are collected in Table 2.1.8.1 in terms of symbols which are defined below. The following abbreviations are used in the subsequent definitions of the symbols:

$$s_{\pm} = 2an_j \sin(\tau \pm \rho), \\ c_{\pm} = 2an_j \cos(\tau \pm \rho) \quad \text{and} \\ \sigma_{\pm} = 2an_j \sin(\tau \pm 2\pi/3 + \rho),$$

and the symbols appearing in Table 2.1.8.1 are given below:

$${}^{(a)}L_j(a, \rho) = \langle J_0(s_+) J_0(s_-) \rangle_{\tau} \\ = \sum_{k=-\infty}^{\infty} \cos(4k\rho) J_k^4 \\ = J_0^4(an_j) + 2 \sum_{k=1}^{\infty} \cos(4k\rho) J_k^4(an_j),$$

$${}^{(b)}Q_j^{(1)}(a, \rho) = \langle J_0^2(s_+) J_0^2(s_-) \rangle_{\tau},$$

$${}^{(c)}Q_j^{(2)}(a, \rho) = \langle J_0(s_+) J_0(s_-) J_0(c_+) J_0(c_-) \rangle_{\tau},$$

$${}^{(d)}T_j(a, \rho) = \sum_{k=-\infty}^{\infty} \exp(6ik\rho) J_k^6(an_j) \\ = J_0^6(an_j) + 2 \sum_{k=1}^{\infty} \cos(6k\rho) J_k^6(an_j),$$

$${}^{(e)}H_j^{(1)}(a, \mu) = \left\langle \mathcal{R} \left[S_j^{(1)}(\tau; a, \mu, 0) \right] \right\rangle_{\tau},$$

$${}^{(f)}H_j^{(2)}(a, \mu) = \left\langle \mathcal{R} \left[S_j^{(2)}(\tau; a, \mu, 0) \right] \right\rangle_{\tau},$$

$${}^{(g)}\tilde{H}_j^{(1)}(a, \mu_1, \mu_2, \rho) = \left\langle \mathcal{R} \left[S_j^{(1)}(\tau; a, \mu_1, \rho) \right. \right. \\ \left. \left. \times S_j^{(1)}(\tau; a, \mu_2, -\rho) \right] \right\rangle_{\tau},$$

$${}^{(h)}\tilde{H}_j^{(2)}(a, \mu_1, \mu_2, \rho) = \left\langle \mathcal{R} \left[S_j^{(2)}(\tau; a, \mu_1, \rho) \right. \right. \\ \left. \left. \times S_j^{(2)}(\tau; a, \mu_2, -\rho) \right] \right\rangle_{\tau},$$

where

$$S_j^{(1)}(\tau; a, \mu, \rho) = \sum_{k=-\infty}^{\infty} e^{3ik\mu} J_k^3(s_+)$$

and

$$S_j^{(2)}(\tau; a, \mu, \rho) = \sum_{k=-\infty}^{\infty} e^{3ik\mu} J_k(s_+) J_k(\sigma_+) J_k(\sigma_-).$$

The averages appearing in the above summary are, in general, computed as

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

 Table 2.1.8.1. Atomic contributions to characteristic functions for $p(|E|)$

The table lists symbolic expressions for the atomic contributions to exact characteristic functions (abbreviated as c.f.) for $p(|E|)$, to be computed as single Fourier series (centric), double Fourier series (acentric) and single Fourier–Bessel series (acentric), as defined in Sections 2.1.8.1 and 2.1.8.2. The symbolic expressions are defined in Section 2.1.8.5. The table is arranged by point groups, space groups and parities of the reflection indices analogously to the table of moments, Table 2.1.7.1, and covers all the space groups and statistically different parities of hkl up to and including space group $Fd\bar{3}$. The expressions are valid for atoms in general positions, for general reflections and presume the absence of noncrystallographic symmetry and of dispersive scatterers.

Space group(s)	g	Atomic c.f.	Remarks
Point group: 1 $P1$	1	$J_0(m_j)$	
Point group: $\bar{1}$ $P\bar{1}$	2	$J_0(2t_1n_j)$	
Point groups: 2, m All P	2	$J_0^2(m_j)$	
All C	4	$J_0^2(2m_j)$	
Point group: $2/m$ All P	4	$J_0^2(2t_1n_j)$	
All C	8	$J_0^2(4t_1n_j)$	
Point group: 222 All P	4	$L_j(t, \Delta)^{(a)}$	
All C and I	8	$L_j(2t, \Delta)$	
$F222$	16	$L_j(4t, \Delta)$	
Point group: $mm2$ All P	4	$L_j(t, 0)$	
All C and I	8	$L_j(2t, 0)$	
$Fmm2$	16	$L_j(4t, 0)$	
$Fdd2$	16	$L_j(4t, 0)$	$h + k + l = 2n$
	16	$L_j(4t, \pi/4)$	$h + k + l = 2n + 1$
Point group: mmm All P	8	$L_j(2t_1, 0)$	
All C and I	16	$L_j(4t_1, 0)$	
$Fmmm$	32	$L_j(8t_1, 0)$	
$Fddd$	32	$L_j(8t_1, 0)$	$h + k + l = 2n$
	32	$L_j(8t_1, \pi/4)$	$h + k + l = 2n + 1$
Point group: 4 $P4, P4_2$	4	$L_j(t, 0)$	$l = 2n$
$P4_1^*$	4	$L_j(t, 0)$	$l = 2n + 1$
	4	$L_j(t, \pi/4)$	
$I4$	8	$L_j(2t, 0)$	$2h + l = 2n$
$I4_1$	8	$L_j(2t, 0)$	$2h + l = 2n + 1$
	8	$L_j(2t, \pi/4)$	
Point group: $\bar{4}$ $P\bar{4}$	4	$L_j(t, \Delta)$	
$I\bar{4}$	8	$L_j(2t, \Delta)$	
Point group: $4/m$ All P	8	$L_j(2t_1, 0)$	
$I4/m$	16	$L_j(4t_1, 0)$	
$I4_1/a$	16	$L_j(4t_1, 0)$	$l = 2n$
	16	$L_j(4t_1, \pi/4)$	$l = 2n + 1$
Point group: 422 $P422, P4_21_2, P4_22_2,$ $P4_22_1_2$	8	$Q_j^{(1)}(t, \Delta)^{(b)}$	
$P4_12_2^*, P4_12_12^*$	8	$Q_j^{(1)}(t, \Delta)$	$l = 2n$
	8	$Q_j^{(2)}(t, \Delta)^{(c)}$	$l = 2n + 1$
$I422$	16	$Q_j^{(1)}(2t, \Delta)$	
$I4_122$	16	$Q_j^{(1)}(2t, \Delta)$	$2k + l = 2n$
	16	$Q_j^{(2)}(2t, \Delta)$	$2k + l = 2n + 1$
Point group: $4mm$ All P	8	$Q_j^{(1)}(t, 0)$	
$I4mm, I4cm$	16	$Q_j^{(1)}(2t, 0)$	$2k + l = 2n$
$I4_1md, I4_1cd$	16	$Q_j^{(1)}(2t, 0)$	$2k + l = 2n + 1$
	16	$Q_j^{(1)}(2t, \pi/4)$	
Point groups: $\bar{4}2m,$ $\bar{4}m2$ All P	8	$Q_j^{(1)}(t, \Delta)$	
$I\bar{4}2m, I\bar{4}m2, I\bar{4}c2$	16	$Q_j^{(1)}(2t, \Delta)$	
$I\bar{4}2d$	16	$Q_j^{(1)}(2t, \Delta)$	$2h + l = 2n$
	16	$Q_j^{(2)}(2t, \Delta)$	$2h + l = 2n + 1$
Point group: $4/mmm$ All P	16	$Q_j^{(1)}(2t_1, 0)$	
$I4/mmm, I4/mcm$	32	$Q_j^{(1)}(4t_1, 0)$	$l = 2n$
$I4_1/amd, I4_1/acd$	32	$Q_j^{(1)}(4t_1, 0)$	$l = 2n + 1$
	32	$Q_j^{(1)}(4t_1, \pi/4)$	
Point group: 3 All P and R	3	$J_0^3(m_j)$	
Point group: $\bar{3}$ All P and R	6	$J_0^3(2t_1n_j)$	
Point group: 32 All P and R	6	$T_j(t, \Delta)^{(d)}$	
Point group: $3m$ $P3m1, P3_1m, R3m$	6	$T_j(t, \pi/2)$	$l = 2n (P),$ $h + k + l = 2n (R)$
$P3c1, P3_1c, R3c$	6	$T_j(t, \pi/2)$	$l = 2n + 1 (P),$ $h + k + l = 2n + 1 (R)$
	6	$T_j(t, 0)$	
Point group: $\bar{3}m$ $P\bar{3}m1, P\bar{3}1m, R\bar{3}m$	12	$T_j(2t_1, \pi/2)$	$l = 2n (P),$ $h + k + l = 2n (R)$
$P\bar{3}c1, P\bar{3}1c, R\bar{3}c$	12	$T_j(2t_1, \pi/2)$	$l = 2n + 1 (P),$ $h + k + l = 2n + 1 (R)$
	12	$T_j(2t_1, 0)$	
Point group: 6 $P6$	6	$H_j^{(1)}(t, \pi/2)^{(e)}$	
$P6_1^*$	6	$H_j^{(1)}(t, \pi/2)$	$l = 6n$
	6	$H_j^{(2)}(t, 0)^{(f)}$	$l = 6n + 1, 6n + 5$
	6	$H_j^{(2)}(t, \pi/2)$	$l = 6n + 2, 6n + 4$
	6	$H_j^{(1)}(t, 0)$	$l = 6n + 3$
$P6_2^*$	6	$H_j^{(1)}(t, \pi/2)$	$l = 3n$
	6	$H_j^{(2)}(t, \pi/2)$	$l = 3n \pm 1$
$P6_3$	6	$H_j^{(1)}(t, \pi/2)$	$l = 2n$
	6	$H_j^{(1)}(t, 0)$	$l = 2n + 1$
Point group: $\bar{6}$ $P\bar{6}$	6	$H_j^{(1)}(t, \Delta)$	
Point group: $6/m$ $P6/m$	12	$H_j^{(1)}(2t_1, \pi/2)$	$l = 2n$
$P6_3/m$	12	$H_j^{(1)}(2t_1, \pi/2)$	$l = 2n + 1$
	12	$H_j^{(1)}(2t_1, 0)$	
Point group: 622 $P622$	12	$\tilde{H}_j^{(1)}(t, \pi/2,$ $-\pi/2, \Delta)^{(g)}$	

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

Table 2.1.8.1. Atomic contributions to characteristic functions for $p(|E|)$ (cont.)

Space group(s)	g	Atomic c.f.	Remarks
$P6_122^*$	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)$	$l = 6n$
	12	$\tilde{H}_j^{(2)}(t, 0, 0, \Delta)^{(h)}$	$l = 6n + 1, 6n + 5$
	12	$\tilde{H}_j^{(2)}(t, \pi/2, \pi/2, \Delta)$	$l = 6n + 2, 6n + 4$
	12	$\tilde{H}_j^{(1)}(t, 0, 0, \Delta)$	$l = 6n + 3$
$P6_222^*$	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)$	$l = 3n$
	12	$\tilde{H}_j^{(2)}(t, \pi/2, \pi/2, \Delta)$	$l = 3n \pm 1$
$P6_322$	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, \Delta)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, 0, 0, \Delta)$	$l = 2n + 1$
Point group: $6mm$			
$P6mm$	12	$\tilde{H}_j^{(1)}(t, \pi/2, \pi/2, 0)$	
$P6cc$	12	$\tilde{H}_j^{(1)}(t, \pi/2, \pi/2, 0)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, \pi/2, -\pi/2, 0)$	$l = 2n + 1$
$P6_3cm, P6_3mc$	12	$\tilde{H}_j^{(1)}(t, \pi/2, \pi/2, 0)$	$l = 2n$
	12	$\tilde{H}_j^{(1)}(t, 0, 0, 0)$	$l = 2n + 1$
Point groups: $\bar{6}2m, \bar{6}m2$			
$P\bar{6}2m, P\bar{6}m2$	12	$\tilde{H}_j^{(1)}(t, \Delta, \Delta, 0)$	
$P\bar{6}2c, P\bar{6}c2$	12	$\tilde{H}_j^{(1)}(t, \Delta, \Delta, 0)$	$l = 2n$

* And the enantiomorphous space group.

Space group(s)	g	Atomic c.f.	Remarks	
Point group: $6/mmm$	12	$\tilde{H}_j^{(1)}(t, \Delta + \pi/2, -\Delta - \pi/2, 0)$	$l = 2n + 1$	
	$P6/mmm$	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, \pi/2, 0)$	
	$P6/mcc$	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, \pi/2, 0)$	$l = 2n$
		24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, -\pi/2, 0)$	$l = 2n + 1$
	$P6_3/mcm, P6_3/mmc$	24	$\tilde{H}_j^{(1)}(2t_1, \pi/2, \pi/2, 0)$	$l = 2n$
24		$\tilde{H}_j^{(1)}(2t_1, 0, 0, 0)$	$l = 2n + 1$	
Point group: 23				
$P23, P2_13$	12	$L_j^3(t, \Delta)$		
$I23, I2_13$	24	$L_j^3(2t, \Delta)$		
$F23$	48	$L_j^3(4t, \Delta)$		
Point group: $m\bar{3}$				
$Pm\bar{3}, Pn\bar{3}, Pa\bar{3}$	24	$L_j^3(2t_1, 0)$		
$Im\bar{3}, Ia\bar{3}$	48	$L_j^3(4t_1, 0)$		
$Fm\bar{3}$	96	$L_j^3(8t_1, 0)$		
$Fd\bar{3}$	96	$L_j^3(8t_1, 0)$	$h + k + l = 2n$	
	96	$L_j^3(8t_1, \pi/4)$	$h + k + l = 2n + 1$	

$$\langle f(\tau) \rangle = (2/\pi) \int_0^{\pi/2} f(\tau) d\tau, \quad (2.1.8.39)$$

except $H_j^{(2)}$ and $\tilde{H}_j^{(2)}$ which are computed as

$$\langle f(\tau) \rangle = (3/\pi) \int_0^{\pi/3} f(\tau) d\tau, \quad (2.1.8.40)$$

where $f(\tau)$ is any of the atomic characteristic functions indicated above. The superscripts preceding the symbols in the above summary are appended to the corresponding symbols in Table 2.1.8.1 on their first occurrence.

2.1.8.6. Other non-ideal Fourier p.d.f.'s

As pointed out above, the representation of the p.d.f.'s by Fourier series is also applicable to effects of noncrystallographic symmetry. Thus, Shmueli *et al.* (1985) obtained the following Fourier coefficient for the bicentric distribution in the space group $P\bar{1}$

$$C_k = (2/\pi) \int_0^{\pi/2} \left[\prod_{j=1}^{N/4} J_0(4\pi k \alpha n_j \cos \vartheta) \right] d\vartheta \quad (2.1.8.41)$$

to be used with equation (2.1.8.5). Furthermore, if we use the important property of the characteristic function as outlined in Section 2.1.4.1, it is easy to write down the Fourier coefficient for a $P\bar{1}$ asymmetric unit containing a centrosymmetric fragment centred at a noncrystallographic centre and a number of atoms not related by symmetry. This Fourier for the above partially bicentric arrangement is a product of expressions (2.1.8.17) and (2.1.8.41), with the appropriate number of atoms in each factor (Shmueli & Weiss, 1985a). While the purely bicentric p.d.f. obtained by using (2.1.8.41) with (2.1.8.5) is significantly different from the ideal bicentric p.d.f. given by equation (2.1.5.13) only when the atomic

composition is sufficiently heterogeneous, the above partially bicentric p.d.f. appears to be a useful development even for an equal-atom structure.

The problem of the coexistence of several noncrystallographic centres of symmetry within the asymmetric unit of $P\bar{1}$, and its effect on the p.d.f. of $|E|$, was examined by Shmueli, Weiss & Wilson (1989) by the Fourier method. The latter study indicates that the strongest effect is produced by the presence of a single noncrystallographic centre.

Another kind of noncrystallographic symmetry is that arising from the presence of centrosymmetric fragments in a noncentrosymmetric structure – the subcentric arrangement already discussed in Section 2.1.5.4. A Fourier-series representation of a non-ideal p.d.f. corresponding to this case was developed by Shmueli, Rabinovich & Weiss (1989), and was also applied to the mathematically equivalent effects of dispersion and presence of heavy scatterers in centrosymmetric special positions in a noncentrosymmetric space group.

A variety of other non-ideal p.d.f.'s occur when heavy atoms are present in special positions (Shmueli & Weiss, 1988). Without going into the details of this development, it can be noted that if the atoms are distributed among k types of Wyckoff positions, the characteristic function corresponding to the p.d.f. of $|E|$ is a product of the k characteristic functions, each of which is related to one of these special positions; the same property of the characteristic function as that in Section 2.1.4.1 is here utilized.

2.1.8.7. Comparison of the correction-factor and Fourier approaches

The need for theoretical non-ideal distributions was exemplified by Fig. 2.1.7.1(a), referred to above, and the performance of the two approaches described above, for this particular example, is shown in

2.1. STATISTICAL PROPERTIES OF THE WEIGHTED RECIPROCAL LATTICE

Fig. 2.1.7.1(b). Briefly, the Fourier p.d.f. shows an excellent agreement with the histogram of recalculated $|E|$ values, while the agreement attained by the Hermite correction factor is much less satisfactory, even for the (longest available to us) five-term expansion. It must be pointed out that (i) the inadequacy of 'short' correction factors, in the example shown, is due to the large deviation from the ideal behaviour and (ii) the number of terms used there in the Fourier summation is twenty, whereafter the summation is terminated. Obviously, the computation of twenty (or more) Fourier coefficients is easier than that of five terms in the correction factor. The convergence of the Fourier series is very satisfactory. It appears that the (analytically) exact Fourier approach is the preferred one in cases of large or intermediate deviations, while the correction-factor approach may cope well with small ones. As far as the availability of symmetry-dependent centric and acentric p.d.f.'s is concerned, correction factors are available for all the space groups (see Table 2.1.7.1), while Fourier coefficients of

p.d.f.'s are available for the first 206 space groups (see Table 2.1.8.1). It should be pointed out that p.d.f.'s based on the correction-factor method cope very well with cubic symmetries higher than $Fd\bar{3}$, even if the asymmetric unit of the space group is strongly heterogeneous (Rabinovich *et al.*, 1991b).

Both approaches described in this section are related to the characteristic function of the required p.d.f. The correction-factor p.d.f.'s (2.1.7.5) and (2.1.7.6) can be obtained by expanding the logarithm of the appropriate characteristic function in a series of cumulants [*e.g.* equation (2.1.4.13); see also Shmueli & Wilson (1982)], truncating the series and performing its term-by-term Fourier inversion. The Fourier p.d.f., on the other hand, is computed by forming a Fourier series whose coefficients are *exact* analytical forms of the characteristic function at points related to the summation indices [*e.g.* equations (2.1.8.5), (2.1.8.9) and (2.1.8.11), and Table 2.1.8.1] and truncating the series when the terms become small enough.

References

2.1

- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of mathematical functions*. New York: Dover.
- Barakat, R. (1974). *First-order statistics of combined random sinusoidal waves with application to laser speckle patterns*. *Opt. Acta*, **21**, 903–921.
- Bernstein, S. (1922). *Sur la théorème limite du calcul des probabilités*. *Math. Ann.* **85**, 237–241.
- Bernstein, S. (1927). *Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes*. *Math. Ann.* **97**, 1–59.
- Cramér, H. (1951). *Mathematical methods of statistics*. Princeton University Press.
- Faggiani, R., Lippert, B. & Lock, C. J. L. (1980). *Heavy transition metal complexes of biologically important molecules*. 4. Crystal and molecular structure of pentahydroxonium chloro(uracilato-N(1))(ethylenediamine)platinum(II)chloride (H_5O_2)[PtCl(NH₂CH₂CH₂NH₂)(C₄H₅N₂O₂)]Cl, and chloro(thyminato-N(1))(ethylenediamine)platinum(II), PtCl(NH₂CH₂CH₂NH₂)(C₅H₅N₂O₂). *Inorg. Chem.* **19**, 295–300.
- French, S. & Wilson, K. (1978). *On the treatment of negative intensity observations*. *Acta Cryst.* **A34**, 517–525.
- Gerhard, O. E. (1993). *Line-of-sight velocity profiles in spherical galaxies: breaking the degeneracy between anisotropy and mass*. *Mon. Not. R. Astron. Soc.* **265**, 213–230.
- Giacovazzo, C. (1977). *On different probabilistic approaches to quartet theory*. *Acta Cryst.* **A33**, 50–54.
- Giacovazzo, C. (1980). *Direct methods in crystallography*. London: Academic Press.
- Harker, D. (1953). *The meaning of the average of $|F|^2$ for large values of interplanar spacing*. *Acta Cryst.* **6**, 731–736.
- Hauptman, H. & Karle, J. (1953). *Solution of the phase problem. I. The centrosymmetric crystal*. Am. Crystallogr. Assoc. Monograph No. 3. Dayton, Ohio: Polycrystal Book Service.
- Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *The probability distribution of X-ray intensities. II. Experimental investigation and the X-ray detection of centers of symmetry*. *Acta Cryst.* **3**, 210–214.
- International Tables for Crystallography* (1983). Vol. A. *Space-group symmetry*, edited by T. Hahn. Dordrecht: Kluwer Academic Publishers.
- International Tables for Crystallography* (1999). Vol. C. *Mathematical, physical and chemical tables*, edited by A. J. C. Wilson & E. Prince. Dordrecht: Kluwer Academic Publishers.
- Kendall, M. & Stuart, A. (1977). *The advanced theory of statistics*, Vol. 1, 4th ed. London: Griffin.
- Lipson, H. & Woolfson, M. M. (1952). *An extension of the use of intensity statistics*. *Acta Cryst.* **5**, 680–682.
- Lomer, T. R. & Wilson, A. J. C. (1975). *Scaling of intensities*. *Acta Cryst.* **B31**, 646–647.
- Marel, R. P. van der & Franx, M. (1993). *A new method for the identification of non-Gaussian line profiles in elliptical galaxies*. *Astrophys. J.* **407**, 525–539.
- Myller-Lebedeff, W. (1907). *Die Theorie der Integralgleichungen in Anwendung auf einige Reihenentwicklungen*. *Math. Ann.* **64**, 388–416.
- Nigam, G. D. (1972). *On the compensation of X-ray intensity*. *Indian J. Pure Appl. Phys.* **10**, 655–656.
- Nigam, G. D. & Wilson, A. J. C. (1980). *Compensation of excess intensity in space group P2*. *Acta Cryst.* **A36**, 832–833.
- Rabinovich, D. & Shakked, Z. (1984). *A new approach to structure determination of large molecules by multi-dimensional search methods*. *Acta Cryst.* **A40**, 195–200.
- Rabinovich, S., Shmueli, U., Stein, Z., Shashua, R. & Weiss, G. H. (1991a). *Exact random-walk models in crystallographic statistics. VI. P.d.f.'s of $|E|$ for all plane groups and most space groups*. *Acta Cryst.* **A47**, 328–335.
- Rabinovich, S., Shmueli, U., Stein, Z., Shashua, R. & Weiss, G. H. (1991b). *Exact random-walk models in crystallographic statistics. VII. An all-space-group study of the effects of atomic heterogeneity on the p.d.f.'s of $|E|$* . *Acta Cryst.* **A47**, 336–340.
- Rayleigh, Lord (1879). *Investigations in optics with special reference to the spectroscope*. *Philos. Mag.* **8**, 261–274.
- Rogers, D. (1950). *The probability distribution of X-ray intensities. IV. New methods of determining crystal classes and space groups*. *Acta Cryst.* **3**, 455–464.
- Rogers, D. & Wilson, A. J. C. (1953). *The probability distribution of X-ray intensities. V. A note on some hypersymmetric distributions*. *Acta Cryst.* **6**, 439–449.
- Shmueli, U. (1979). *Symmetry- and composition-dependent cumulative distributions of the normalized structure amplitude for use in intensity statistics*. *Acta Cryst.* **A35**, 282–286.
- Shmueli, U. (1982). *A study of generalized intensity statistics: extension of the theory and practical examples*. *Acta Cryst.* **A38**, 362–371.
- Shmueli, U. & Kaldor, U. (1981). *Calculation of even moments of the trigonometric structure factor*. *Methods and results*. *Acta Cryst.* **A37**, 76–80.
- Shmueli, U. & Kaldor, U. (1983). *Moments of the trigonometric structure factor*. *Acta Cryst.* **A39**, 615–621.
- Shmueli, U., Rabinovich, S. & Weiss, G. H. (1989). *Exact conditional distribution of a three-phase invariant in the space group P1. I. Derivation and simplification of the Fourier series*. *Acta Cryst.* **A45**, 361–367.
- Shmueli, U., Rabinovich, S. & Weiss, G. H. (1990). *Exact random-walk models in crystallographic statistics. V. Non-symmetrically bounded distributions of structure-factor magnitudes*. *Acta Cryst.* **A46**, 241–246.
- Shmueli, U. & Weiss, G. H. (1985a). *Centric, bicentric and partially bicentric intensity statistics*. *Structure and statistics in crystallography*, edited by A. J. C. Wilson, pp. 53–66. Guilderland: Adenine Press.
- Shmueli, U. & Weiss, G. H. (1985b). *Exact joint probability distributions for centrosymmetric structure factors. Derivation and application to the Σ_1 relationship in the space group P1*. *Acta Cryst.* **A41**, 401–408.
- Shmueli, U. & Weiss, G. H. (1986). *Exact joint distribution of E_h , E_k and E_{h+k} and the probability for the positive sign of the triple product in the space group P1*. *Acta Cryst.* **A42**, 240–246.
- Shmueli, U. & Weiss, G. H. (1987). *Exact random-walk models in crystallographic statistics. III. Distributions of $|E|$ for space groups of low symmetry*. *Acta Cryst.* **A43**, 93–98.
- Shmueli, U. & Weiss, G. H. (1988). *Exact random-walk models in crystallographic statistics. IV. P.d.f.'s of $|E|$ allowing for atoms in special positions*. *Acta Cryst.* **A44**, 413–417.
- Shmueli, U., Weiss, G. H. & Kiefer, J. E. (1985). *Exact random-walk models in crystallographic statistics. II. The bicentric distribution in the space group P1*. *Acta Cryst.* **A41**, 55–59.
- Shmueli, U., Weiss, G. H., Kiefer, J. E. & Wilson, A. J. C. (1984). *Exact random-walk models in crystallographic statistics. I. Space groups $P\bar{1}$ and P1*. *Acta Cryst.* **A40**, 651–660.
- Shmueli, U., Weiss, G. H. & Wilson, A. J. C. (1989). *Explicit Fourier representations of non-ideal hypercentric p.d.f.'s of $|E|$* . *Acta Cryst.* **A45**, 213–217.
- Shmueli, U. & Wilson, A. J. C. (1981). *Effects of space-group symmetry and atomic heterogeneity on intensity statistics*. *Acta Cryst.* **A37**, 342–353.
- Shmueli, U. & Wilson, A. J. C. (1982). *Intensity statistics: non-ideal distributions in theory and practice*. In *Crystallographic statistics: progress and problems*, edited by S. Ramaseshan, M. F. Richardson & A. J. C. Wilson, pp. 83–97. Bangalore: Indian Academy of Sciences.
- Shmueli, U. & Wilson, A. J. C. (1983). *Generalized intensity statistics: the subcentric distribution and effects of dispersion*. *Acta Cryst.* **A39**, 225–233.
- Spiegel, M. R. (1974). *Theory and problems of Fourier analysis*. Schaum's Outline Series. New York: McGraw-Hill.

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

2.1 (cont.)

- Srinivasan, R. & Parthasarathy, S. (1976). *Some statistical applications in X-ray crystallography*. Oxford: Pergamon Press.
- Stuart, A. & Ord, K. (1994). *Kendall's advanced theory of statistics*. Vol. 1. *Distribution theory*, 6th ed. London: Edward Arnold.
- Weiss, G. H. & Kiefer, J. E. (1983). *The Pearson random walk with unequal step sizes*. *J. Phys. A*, **16**, 489–495.
- Wilson, A. J. C. (1942). *Determination of absolute from relative intensity data*. *Nature (London)*, **150**, 151–152.
- Wilson, A. J. C. (1949). *The probability distribution of X-ray intensities*. *Acta Cryst.* **2**, 318–320.
- Wilson, A. J. C. (1950). *The probability distribution of X-ray intensities. III. Effects of symmetry elements on zones and rows*. *Acta Cryst.* **3**, 258–261.
- Wilson, A. J. C. (1952). *Hypercentric and hyperparallel distributions of X-ray intensities*. *Research (London)*, **5**, 588–589.
- Wilson, A. J. C. (1956). *The probability distribution of X-ray intensities. VII. Some sesquicentric distributions*. *Acta Cryst.* **9**, 143–144.
- Wilson, A. J. C. (1964). *The probability distribution of X-ray intensities. VIII. A note on compensation for excess average intensity*. *Acta Cryst.* **17**, 1591–1592.
- Wilson, A. J. C. (1975). *Effect of neglect of dispersion on apparent scale and temperature factors*. In *Anomalous scattering*, edited by S. Ramaseshan & S. C. Abrahams, pp. 325–332. Copenhagen: Munksgaard.
- Wilson, A. J. C. (1976). *Statistical bias in least-squares refinement*. *Acta Cryst.* **A32**, 994–996.
- Wilson, A. J. C. (1978a). *On the probability of measuring the intensity of a reflection as negative*. *Acta Cryst.* **A34**, 474–475.
- Wilson, A. J. C. (1978b). *Variance of X-ray intensities: effect of dispersion and higher symmetries*. *Acta Cryst.* **A34**, 986–994.
- Wilson, A. J. C. (1978c). *Statistical bias in scaling factors: Erratum*. *Acta Cryst.* **B34**, 1749.
- Wilson, A. J. C. (1979). *Problems of resolution and bias in the experimental determination of the electron density and other densities in crystals*. *Acta Cryst.* **A35**, 122–130.
- Wilson, A. J. C. (1980a). *Relationship between 'observed' and 'true' intensity: effects of various counting modes*. *Acta Cryst.* **A36**, 929–936.
- Wilson, A. J. C. (1980b). *Effect of dispersion on the probability distribution of X-ray reflections*. *Acta Cryst.* **A36**, 945–946.
- Wilson, A. J. C. (1981). *Can intensity statistics accommodate stereochemistry?* *Acta Cryst.* **A37**, 808–810.
- Wilson, A. J. C. (1986a). *Distributions of sums and ratios of sums of intensities*. *Acta Cryst.* **A42**, 334–339.
- Wilson, A. J. C. (1986b). *Fourier versus Hermite representations of probability distributions*. *Acta Cryst.* **A42**, 81–83.
- Wilson, A. J. C. (1987a). *Treatment of enhanced zones and rows in normalizing intensities*. *Acta Cryst.* **A43**, 250–252.
- Wilson, A. J. C. (1987b). *Functional form of the ideal hypersymmetric distributions of structure factors*. *Acta Cryst.* **A43**, 554–556.
- Wilson, A. J. C. (1993). *Space groups rare for organic structures. III. Symmorphisms and inherent symmetry*. *Acta Cryst.* **A49**, 795–806.
- Argos, P. & Rossmann, M. G. (1980). *Molecular replacement method*. In *Theory and practice of direct methods in crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 381–389. New York: Plenum.
- Avrami, M. (1938). *Direct determination of crystal structure from X-ray data*. *Phys. Rev.* **54**, 300–303.
- Baggio, R., Woolfson, M. M., Declercq, J.-P. & Germain, G. (1978). *On the application of phase relationships to complex structures. XVI. A random approach to structure determination*. *Acta Cryst.* **A34**, 883–892.
- Banerjee, K. (1933). *Determination of the signs of the Fourier terms in complete crystal structure analysis*. *Proc. R. Soc. London Ser. A*, **141**, 188–193.
- Bertaut, E. F. (1955a). *La méthode statistique en cristallographie. I*. *Acta Cryst.* **8**, 537–543.
- Bertaut, E. F. (1955b). *La méthode statistique en cristallographie. II. Quelques applications*. *Acta Cryst.* **8**, 544–548.
- Bertaut, E. F. (1960). *Ordre logarithmique des densités de répartition. I*. *Acta Cryst.* **13**, 546–552.
- Beurskens, P. T., Beurskens, G., de Gelder, R., Garcia-Granda, S., Gould, R. O., Israel, R. & Smits, J. M. M. (1999). *The DIRDIF-99 program system*. Crystallography Laboratory, University of Nijmegen, The Netherlands.
- Beurskens, P. T., Gould, R. O., Bruins Slot, H. J. & Bosman, W. P. (1987). *Translation functions for the positioning of a well oriented molecular fragment*. *Z. Kristallogr.* **179**, 127–159.
- Beurskens, P. T., Prick, A. J., Doesburg, H. M. & Gould, R. O. (1979). *Statistical properties of normalized difference-structure factors for non-centrosymmetric structures*. *Acta Cryst.* **A35**, 765–772.
- Böhme, R. (1982). *Direkte Methoden für Strukturen mit Überstruktureffekten*. *Acta Cryst.* **A38**, 318–326.
- Bouman, J. (1956). *A general theory of inequalities*. *Acta Cryst.* **9**, 777–780.
- Bricogne, G. (1984). *Maximum entropy and the foundation of direct methods*. *Acta Cryst.* **A40**, 410–415.
- Britten, P. L. & Collins, D. M. (1982). *Information theory as a basis for the maximum determinant*. *Acta Cryst.* **A38**, 129–132.
- Buerger, M. J. (1959). *Vector space and its applications in crystal structure investigation*. New York: John Wiley.
- Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (1999). *SIR99, a program for the automatic solution of small and large crystal structures*. *Acta Cryst.* **A55**, 991–999.
- Burla, M. C., Cascarano, G., Giacovazzo, C., Nunzi, A. & Polidori, G. (1987). *A weighting scheme for tangent formula development. III. The weighting scheme of the SIR program*. *Acta Cryst.* **A43**, 370–374.
- Busetta, B. (1976). *An example of the use of quartet and triplet structure invariants when enantiomorph discrimination is difficult*. *Acta Cryst.* **A32**, 139–143.
- Busetta, B., Giacovazzo, C., Burla, M. C., Nunzi, A., Polidori, G. & Viterbo, D. (1980). *The SIR program. I. Use of negative quartets*. *Acta Cryst.* **A36**, 68–74.
- Camalli, M., Giacovazzo, C. & Spagna, R. (1985). *From a partial to the complete crystal structure. II. The procedure and its applications*. *Acta Cryst.* **A41**, 605–613.
- Cascarano, G. & Giacovazzo, C. (1983). *One-phase seminvariants of first rank. I. Algebraic considerations*. *Z. Kristallogr.* **165**, 169–174.
- Cascarano, G. & Giacovazzo, C. (1985). *One-wavelength technique: some probabilistic formulas using the anomalous dispersion effect*. *Acta Cryst.* **A41**, 408–413.
- Cascarano, G., Giacovazzo, C., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *The distribution of α_h* . *Acta Cryst.* **A40**, 389–394.
- Cascarano, G., Giacovazzo, C., Calabrese, G., Burla, M. C., Nunzi, A., Polidori, G. & Viterbo, D. (1984). *One-phase seminvariants of first rank. II. Probabilistic considerations*. *Z. Kristallogr.* **167**, 37–47.

2.2

- Allegra, G. (1979). *Derivation of three-phase invariants from the Patterson function*. *Acta Cryst.* **A35**, 213–220.
- Altomare, A., Burla, M. C., Camalli, M., Cascarano, G. L., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Spagna, R. (1999). *SIR97: a new tool for crystal structure determination and refinement*. *J. Appl. Cryst.* **32**, 115–119.
- Anzenhofer, K. & Hoppe, W. (1962). *Phys. Verh. Mosbach*. **13**, 119.
- Ardito, G., Cascarano, G., Giacovazzo, C. & Luić, M. (1985). *I-Phase seminvariants and Harker sections*. *Z. Kristallogr.* **172**, 25–34.