

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

$$\lambda_1 = 2.404825577$$

$$\lambda_2 = 5.5200781103$$

$$\lambda_3 = 8.6537279129$$

$$\lambda_4 = 11.7915344390$$

$$\lambda_5 = 14.9309177085$$

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