

## Exact Joint Probability Distribution for Centrosymmetric Structure Factors. Derivation and Application to the $\sum_1$ Relationship in the Space Group $P\bar{1}$

BY URI SHMUELI

*Department of Chemistry, Tel Aviv University, Ramat Aviv, 69 978 Tel Aviv, Israel*

AND GEORGE H. WEISS

*Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20205, USA*

(Received 5 December 1984; accepted 6 March 1985)

### Abstract

Recent applications of exact random-walk techniques to crystallographic structure-factor statistics have now been extended to multivariate joint probability density functions of several structure factors. The technique of deriving such multivariate exact density functions is introduced, and is applied to the study of the simplest sign relationship:  $\sum_1$ , in the space group  $P\bar{1}$ . An exact expression is obtained for the probability that the sign of the normalized structure factor  $E(2h\ 2k\ 2l)$  is positive, given the magnitudes of  $E(2h\ 2k\ 2l)$  and  $E(h\ k\ l)$ , and an extensive numerical examination of this new expression - as compared with the conventional asymptotic formula for this probability - is presented. It is shown that the asymptotic formula usually underestimates the probability that  $E(2h\ 2k\ 2l)$  is positive, the discrepancies between the exact and asymptotic results being rather serious when the atomic composition of the asymmetric unit is heterogeneous (even moderately so), and when the number of atoms is small; a paucity of atoms leads to significant discrepancies even in the equal-atom case. On the other hand, for large asymmetric units of low heterogeneity and for high  $E$  values, the exact and asymptotic expressions agree very well in their predictions. The qualitative behaviour of the new exact expression is consistent with the known features of the  $\sum_1$  relationship and its statistical interpretations.

### Introduction

Recent applications of exact random-walk techniques to the derivation of probability density functions (p.d.f.'s) of a single structure factor (Shmueli, Weiss, Kiefer & Wilson, 1984; Shmueli, Weiss & Kiefer, 1985) yield exact expressions for the desired p.d.f.'s, which are readily computed to any required accuracy. These new p.d.f.'s are given as Fourier or Fourier-Bessel series, possessing favourable convergence properties. Although these new p.d.f.'s have not yet

been generalized to all the space groups, like the existing Gram-Charlier expansions (Shmueli & Wilson, 1981; Shmueli & Kaldor, 1981, 1983), their performance is much superior to that of the Gram-Charlier series, especially in cases of large discrepancies between observed distributions and those predicted (Wilson, 1949) on the basis of the central limit theorem (e.g. Cramer, 1951). It should be pointed out that random-walk p.d.f.'s are not new in crystallography (e.g. Hauptman & Karle, 1952; Wilson, 1952; Bertaut, 1955*a*); the novel aspect pursued in this study is their representation in terms of Fourier series.

These successful developments, and the simplicity and generality of the underlying statistical principles (e.g. Weiss & Kiefer, 1983), suggested to us an investigation of the joint p.d.f.'s of several structure factors with the obvious purpose of constructing exact expressions applicable to direct methods of sign and phase determination. Such expressions are of interest for two main reasons: (i) the existing direct-methods formalisms are either based on the central limit theorem approximation, or they consist of Gram-Charlier-type generalizations of the latter (for a comprehensive review see, for example, Giacovazzo, 1980), and (ii) the expressions used in practice are well known to be approximate, and the goodness of these approximations cannot be predicted in advance except for equal-atom structures that obey the assumptions of the Wilson (1949) statistics, in which case the approximations are known to be very good.

Some preliminary results of our study of the exact statistical aspects of the  $\sum_1$  and  $\sum_2$  relationships, in the space group  $P\bar{1}$ , have been communicated (Weiss, Shmueli, Kiefer & Wilson, 1984*a*) and are given elsewhere (Weiss, Shmueli, Kiefer & Wilson, 1984*b*).

The purpose of the present paper is (i) to introduce the mathematical technique for the systematic derivation of the exact joint p.d.f. of several structure factors, (ii) to apply this technique to the derivation of an exact probability for a positive sign of the structure factor  $E_{2h}$ , given the magnitudes  $|E_h|$  and  $|E_{2h}|$ , where  $E_h$  is the normalized structure factor of reflection

$h(hkl)$ , and (iii) to compare the exact probability for the above  $\sum_1$  relationship, in the space group  $P\bar{1}$ , with the usual asymptotic one – the hyperbolic tangent formula (Cochran & Woolfson, 1955; Klug, 1958) – with regard to the magnitudes of the  $|E|$  values, the number of atoms in the asymmetric unit, and the effect of the presence of heavy atoms.

As will be seen from what follows, the exact expression for the probability for the positive sign of  $E_{2h}$  reduces to the Cochran & Woolfson (1955) formula in the limit of a large number of equal atoms. Significant deviations of the asymptotic result from that derived below may be caused by a paucity of atoms or a heterogeneous composition of the asymmetric unit.

### Mathematical analysis

The technique proposed by Shmueli *et al.* (1984) for the representation of the p.d.f. of a single structure factor as a Fourier series is readily generalized to the multidimensional case.

Let  $\mathbf{E} = (E_1, E_2, \dots, E_d)$  be the vector of the component normalized structure factors, where

$$E_s = E_{h_s} = \sum_{j=1}^{N/k} n_j \xi_j(\mathbf{h}_s), \quad (1)$$

$$n_j = f_j / \left( \sum_{j=1}^N f_j^2 \right)^{1/2} \quad (2)$$

is the normalized scattering factor, and

$$\xi(\mathbf{h}_s) = \sum_{q=1}^k \exp [2\pi i \mathbf{h}_s^T (\mathbf{P}_q \mathbf{r}_j + \mathbf{t}_q)] \quad (3)$$

is the trigonometric structure factor (*International Tables for X-ray Crystallography*, 1952) of the  $j$ th atom. This representation explicitly includes its dependence on the space-group symmetry of the crystal; in (1)  $N$  is the number of atoms in the unit cell,  $k$  is the order of the point group times the multiplicity of the Bravais lattice,  $f_j$  is the scattering factor of the  $j$ th atom, including its vibration parameters,  $\mathbf{h}_s$  is the diffraction vector corresponding to the normalized structure factor  $E_s$ , and  $\mathbf{P}_q$  and  $\mathbf{t}_q$  are the rotation and translation parts of the  $q$ th space-group operator respectively.

This analysis is confined to the centrosymmetric case, and is based on the following assumptions:

- (i) All the atoms are located in general positions.
- (ii) Dispersion is neglected.
- (iii) The atomic phase factors,  $2\pi \mathbf{h} \cdot \mathbf{r}$ , are uniformly distributed in the  $(0, 2\pi)$  range, and the position vectors of different atoms are rationally independent (Hauptman & Karle, 1953; Hauptman, 1964).
- (iv) The contributions of different atoms, within the asymmetric unit of the space group, are independent.

We wish to derive the joint probability density function of the vector  $\mathbf{E}$ , the components of which can be regarded as random variables.

Clearly, each of the individual  $E$ 's may be non zero only in the range  $-E_{\max} < E_s < E_{\max}$ , where

$$E_{\max} = \sum_{j=1}^N n_j \equiv 1/\alpha. \quad (4)$$

The required p.d.f. can therefore be expanded in the multiple Fourier series

$$p(\mathbf{E}) = (\alpha/2)^d \sum_{\mathbf{u}} C_{\mathbf{u}} \exp(-\pi i \alpha \mathbf{u}^T \mathbf{E}), \quad (5)$$

where  $\mathbf{u} = (u_1, u_2, \dots, u_d)$  is a vector of the summation indices. The coefficients,  $C_{\mathbf{u}}$ , are found from

$$\begin{aligned} C_{\mathbf{u}} &= \int_{-1/\alpha}^{1/\alpha} \dots \int_{-1/\alpha}^{1/\alpha} p(\mathbf{E}) \exp(\pi i \alpha \mathbf{u}^T \mathbf{E}) d^m \mathbf{E} \quad (6) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\mathbf{E}) \exp(\pi i \alpha \mathbf{u}^T \mathbf{E}) d^m \mathbf{E} \\ &= C(\pi \alpha \mathbf{u}), \quad (7) \end{aligned}$$

where  $m$  is the number of independent component structure factors,  $E_s$ , and  $C(\omega)$  is the characteristic function (*e.g.* Kendall & Stuart, 1963) corresponding to the p.d.f.  $p(\mathbf{E})$ . For example, if  $p(\mathbf{E}) = p(E_h, E_{2h})$ , a double Fourier series (5) is needed for the p.d.f., while only one integral needs to be evaluated in (6) since  $E_h$  and  $E_{2h}$  are correlated. We therefore have  $d=2$  and  $m=1$  for the above example, which will be considered in detail below.

The characteristic function,  $C(\omega)$ , can be regarded as an expectation value of the exponential  $\exp(i\omega^T \mathbf{E})$ , and can thus be evaluated from known statistical properties of the component structure factors. We thus have

$$\begin{aligned} C(\omega) &= \langle \exp(i\omega^T \mathbf{E}) \rangle \\ &= \left\langle \exp \left( i \sum_{r=1}^d \omega_r E_r \right) \right\rangle \\ &= \left\langle \exp \left[ i \sum_{r=1}^d \sum_{j=1}^{N/k} \omega_r n_j \xi_j(\mathbf{h}_r) \right] \right\rangle \\ &= \prod_{j=1}^{N/k} \left\langle \exp \left[ i n_j \sum_{r=1}^d \omega_r \xi_j(\mathbf{h}_r) \right] \right\rangle \quad (8) \end{aligned}$$

$$= \prod_{j=1}^{N/k} C_j(\omega), \quad (9)$$

where  $C_j(\omega)$  is the contribution of the  $j$ th atom to the characteristic function, and use has been made of (1) and assumption (iv) concerning the independence of the atomic contributions.

The construction of the joint p.d.f. (5) thus requires:

- (i) evaluation of the atomic characteristic functions  $C_j(\omega)$ , from (3) and (8), and

(ii) replacement of the components of  $\omega$  by the appropriate functions of the summation indices; in the present case,  $\omega_r \rightarrow \pi\alpha u_r$ .

The above formalism is a general one for all the centrosymmetric space groups, and is valid under the assumptions stated; it does not account for non-crystallographic symmetry. The corresponding treatment of non-centrosymmetric space groups is closely analogous to the above, and effects of non-crystallographic symmetry can be allowed for by suitably modified trigonometric structure factors (e.g. Shmueli, Weiss & Kiefer, 1985).

### The $\Sigma_1$ relationship in $P\bar{1}$

The required joint p.d.f. is  $p(E_h, E_{2h})$ , which will enable us to find an exact expression for the probability that  $E_{2h}$  is positive - given the magnitudes  $|E_h|$  and  $|E_{2h}|$ .

We have, for the space group  $P\bar{1}$ ,

$$\xi_j(\mathbf{h}) = 2 \cos \theta_j \quad \text{and} \quad \xi_j(2\mathbf{h}) = 2 \cos 2\theta_j, \quad (10)$$

where  $\theta_j = 2\pi\mathbf{h}\cdot\mathbf{r}_j$  is taken as uniform in the  $[0, 2\pi]$  range (see above). The atomic characteristic function,  $C_j(\omega_1, \omega_2)$  is thus given by (8) as

$$\begin{aligned} C_j(\omega_1, \omega_2) &= \langle \exp [in_j(\omega_1 \cos \theta_j + \omega_2 \cos 2\theta_j)] \rangle \\ &= (1/2\pi) \int_{-\pi}^{\pi} \exp [in_j(\omega_1 \cos \theta \\ &\quad + \omega_2 \cos 2\theta)] d\theta. \end{aligned} \quad (11)$$

It is convenient to reduce this integral to series form, by making use of the expansion

$$\exp(iu \cos \theta) = \sum_{s=-\infty}^{\infty} i^s J_s(u) \exp(is\theta) \quad (12)$$

[Gradshteyn & Ryzhik (1980), entry: 8.511(4)]. We can then readily perform the integration over  $\theta$  in (11), and making use of some basic properties of Bessel functions of integer order (e.g. Abramowitz & Stegun, 1972), we obtain

$$\begin{aligned} C_j(\omega_1, \omega_2) &= \sum_{m=-\infty}^{\infty} i^{3m} J_{2m}(2n_j\omega_1) J_m(2n_j\omega_2) \\ &= R_j + iI_j, \end{aligned} \quad (13)$$

where

$$\begin{aligned} R_j(\omega_1, \omega_2) &= J_0(2n_j\omega_1) J_0(2n_j\omega_2) \\ &\quad + 2 \sum_{m=1}^{\infty} (-1)^m J_{4m}(2n_j\omega_1) J_{2m}(2n_j\omega_2) \end{aligned} \quad (14)$$

and

$$I_j(\omega_1, \omega_2) = 2 \sum_{m=0}^{\infty} (-1)^{m+1} J_{4m+2}(2n_j\omega_1) J_{2m+1}(2n_j\omega_2). \quad (15)$$

The characteristic function corresponding to the required p.d.f. is now obtained by substituting (13)–(15) into (9), and the Fourier coefficients, to be used in (5), follow readily. On making the replacements:  $\omega_1 \rightarrow \pi\alpha s$ ,  $\omega_2 \rightarrow \pi\alpha t$ , and introducing the abbreviations

$$R_{jst} = R_j(\pi\alpha s, \pi\alpha t) \quad \text{and} \quad I_{jst} = I_j(\pi\alpha s, \pi\alpha t), \quad (16)$$

the Fourier coefficient becomes

$$C_{st} = \prod_{j=1}^{N/2} (R_{jst} + iI_{jst}) = R_{st} + iI_{st}. \quad (17)$$

The Fourier series in (5) can be simplified by making use of the symmetry of its coefficients. Thus,  $I_{st}$  is a sum of products, each containing an odd number of  $I_{jst}$ . Similarly,  $R_{st}$  consists of the sum of all terms, in which there is an even number of  $I_{jst}$ . Consequently, one can determine the relevant properties of  $R_{st}$  and  $I_{st}$  by examining those of  $R_{jst}$  and  $I_{jst}$ . The explicit formulae in (14) and (15) allow us to write

$$\begin{aligned} R_{j\bar{s}t} &= R_{jst} = R_{j\bar{s}\bar{t}} = R_{jst}, \\ I_{j\bar{s}t} &= I_{jst} \quad \text{and} \quad I_{j\bar{s}\bar{t}} = -I_{jst}, \end{aligned} \quad (18)$$

which leads to

$$\begin{aligned} R_{s\bar{t}} &= R_{s\bar{t}} = R_{s\bar{t}} = R_{st}, \\ I_{s\bar{t}} &= I_{st} \quad \text{and} \quad I_{s\bar{t}} = -I_{st}. \end{aligned} \quad (19)$$

Making use of these relationships, we can express the required p.d.f. as

$$\begin{aligned} p(E_1, E_2) &= (\alpha^2/4) \left\{ 1 + 2 \sum_{s=1}^{\infty} [R_{s0} \cos(\pi\alpha E_1) \right. \\ &\quad + R_{0s} \cos(\pi\alpha E_2)] \\ &\quad + 4 \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} [R_{st} \cos(\pi\alpha E_2) \\ &\quad \left. + I_{st} \sin(\pi\alpha E_2)] \cos(\pi\alpha E_1) \right\}, \end{aligned} \quad (20)$$

where  $E_1 = E_h$  and  $E_2 = E_{2h}$ . Notice that in this formula

$$R_{s0} = R_{0s} = \prod_{j=1}^{N/2} J_0(2\pi\alpha n_j). \quad (21)$$

The conditional density,  $p(E_2|E_1)$ , is given by

$$p(E_2|E_1) = p(E_1, E_2)/p(E_1), \quad (22)$$

where  $p(E_1)$  is obtained as

$$p(E_1) = (\alpha/2) \left\{ 1 + 2 \sum_{s=1}^{\infty} R_{s0} \cos(\pi\alpha E_1) \right\}, \quad (23)$$

which, together with the explicit expression (21), agrees with the p.d.f. of the structure factor as given by Shmueli *et al.* (1984).

Finally, the probability that  $E_{2h}$  is positive, conditioned on the knowledge of  $|E_h|$  and to be denoted by  $p_+(2\mathbf{h}|\mathbf{h})$ , is readily obtained using Bertaut's iden-

tity:  $p_+ = (1 + p_-/p_+)^{-1}$  (Bertaut, 1955*b*) with (20). We obtain

$$p_+(2\mathbf{h}|\mathbf{h}) = \frac{1}{2}(1 + \Omega/\Gamma) \quad (24)$$

where

$$\Omega = \alpha^2 \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} I_{st} \cos(\pi s \alpha E_{\mathbf{h}}) \sin(\pi t \alpha |E_{2\mathbf{h}}|) \quad (25)$$

and

$$\Gamma = (\alpha^2/4) \left\{ 1 + 2 \sum_{s=1}^{\infty} R_{s0} [\cos(\pi s \alpha E_{\mathbf{h}}) + \cos(\pi s \alpha |E_{2\mathbf{h}}|)] + 4 \sum_{s=1}^{\infty} \sum_{t=1}^{\infty} R_{st} \cos(\pi s \alpha E_{\mathbf{h}}) \cos(\pi t \alpha |E_{2\mathbf{h}}|) \right\}. \quad (26)$$

Numerical results, obtained using these exact formulae, will be presented and discussed in the next section.

Equations (20) and (24) are exact. We can, however, recover the classical approximation, first derived by Vand & Pepinsky (1953) and by Cochran & Woolfson (1955), by keeping the lowest-order Bessel functions in (14) and (15). The rederivation of the classical approximation from the exact result is given in Appendix A. While it is also possible to derive systematic corrections to the approximate result, the algebra becomes extremely complicated, so that there is little advantage in not using the exact series representations (20) and (24) when the asymptotic classical approximation is inadequate. The magnitudes of the differences between approximate and exact formulae for  $p_+(2\mathbf{h}|\mathbf{h})$  are discussed below.

### Examples and discussion

The exact expression for the probability  $p_+(2\mathbf{h}|\mathbf{h})$ , given by (24), has been evaluated for a range of hypothetical atomic compositions and values of  $|E_{2\mathbf{h}}|$  and  $|E_{\mathbf{h}}|$ . These results were accompanied by the evaluation of corresponding values of the asymptotic probability formula

$$p_+(2\mathbf{h}|\mathbf{h}) \approx \frac{1}{2} + \frac{1}{2} \tanh [(\sigma_3/2\sigma_2^{3/2})K] \quad (27)$$

(Cochran & Woolfson, 1955; Klug, 1958), where  $K = |E_{2\mathbf{h}}|(E_{\mathbf{h}}^2 - 1)$ .

The computations were programmed in Fortran and use has been made of a local library subroutine in the evaluation of the Bessel functions  $J_n(x)$  for a large range of their orders. Preliminary tests showed that satisfactory convergence of the various summations can be attained by taking the first ten terms of the Bessel-function series (14) and (15), and the first twenty terms in a single Fourier summation in (24). The assumed composition of the asymmetric unit of

the space group  $P\bar{1}$  was taken as  $C_nX_m$ , and the composition-dependent quantities that appear in (14), (15), (24) and (27) were expressed in terms of  $n$ ,  $m$  and  $\rho$  - the ratio of the atomic numbers  $Z_C$  and  $Z_X$  (e.g. Shmueli, 1979). We thus have

$$u_C = 1/[2(n + m\rho)], \quad u_X = \rho/[2(n + m\rho)], \\ \alpha = [2(nu_C^2 + mu_X^2)]^{1/2}, \quad |E|_{\max} = 1/\alpha, \\ \frac{1}{2}\sigma_3\sigma_2^{-3/2} = (n + m\rho^3)/[2(n + m\rho^2)]^{3/2},$$

where  $\rho = Z_X/Z_C$ ,  $u_C$  and  $u_X$  are unitary scattering factors of C and X respectively and  $\alpha$  is defined by (4). Note that the unitary and normalized scattering factors [cf. (2)] are related by  $u_j = \alpha n_j$ .

The exact and approximate expressions for  $p_+(2\mathbf{h}|\mathbf{h})$  were compared with respect to the effects of (i) the values of  $|E_{\mathbf{h}}|$  and  $|E_{2\mathbf{h}}|$ , (ii) the number of atoms in the asymmetric unit, and (iii) the atomic heterogeneity, as given by the value of  $\rho$  defined above. Only sample comparisons can be presented within the available space, and these are by no means an exhaustive description of the effects of the above three strongly interrelated variables.

Fig. 1 shows the effect of the magnitude of the  $|E|$  values, for the moderately heterogeneous assumed composition  $C_{30}Kr_2$  ( $\rho = 6$ ). The magnitude of  $E_{2\mathbf{h}}$  was taken as 1.75, and  $|E_{\mathbf{h}}|$  was allowed to vary from 0.5 to 3.0 in steps of 0.1. Similar calculations for higher values of  $\rho$  result in steeper sigmoid curves for both approximate and exact probabilities, and in increasing discrepancies between the probability values both above and below the crossover point. As

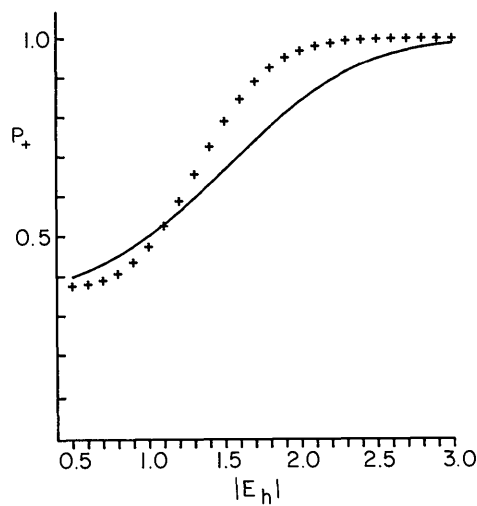


Fig. 1. Effect of the magnitudes of  $E$  values. The exact and approximate probabilities  $p_+(2\mathbf{h}|\mathbf{h})$  for the positive sign of  $E_{2\mathbf{h}}$  as a function of  $|E_{\mathbf{h}}|$ , for the assumed composition  $C_{30}Kr_2$  of the asymmetric unit of  $P\bar{1}$ . The magnitude of  $E_{2\mathbf{h}}$  is taken as 1.75. The solid line denotes the approximate probability, computed from (27), and the centres of the '+' signs denote the values of the exact probability, obtained from (24).

the heterogeneity is decreased both curves approach each other, until in the equal-atom case they very nearly coalesce - for 32 atoms in the asymmetric unit; effects of paucity of atoms are negligible for this number of equal atoms. A common feature of such comparisons is that the approximate expression underestimates the probability for high values of  $|E_h|$ , and overestimates it at the low side of this  $|E|$ . No such crossover is reported by Klug (1958), who bases his conclusions on the first few terms in the Edgeworth expansion for the relevant joint p.d.f. The crossover is, however, present in Giacobozzo's (1978, p. 565) work on one-phase seminvariants.

The effect of the number of atoms in the asymmetric unit of  $P\bar{1}$  is shown in Fig. 2. The atomic composition is here chosen as  $C_nX_2$ , and the probabilities are computed, for fixed  $|E|$  values, as a function of the number of carbons. Equal and reasonably large values of  $|E|$  have been chosen for these illustrations ( $|E_h| = |E_{2h}| = 2.25$ ). Fig. 2(a) shows the result for the equal-atom case, while the composition for Fig. 2(b) is  $C_nKr_2$ . It is seen from Fig. 2(a) that effects of paucity of atoms, in the equal-atom case, can lead to significant discrepancies, while the exact and approximate probabilities approach each other asymptotically with increasing  $n$ . In the presence of two moderately heavy atoms [Fig. 2(b)] the range of  $n$  over which discrepancies persist is more extended than in the equal-atom case. The exact expression is required for accurate evaluation of  $p_+(2h|h)$  when the number of atoms in the unit cell is small.

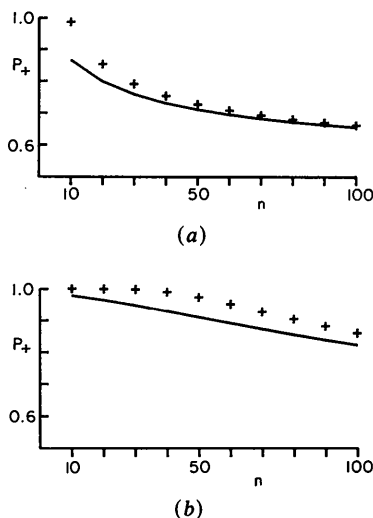


Fig. 2. Effect of the number of (light) atoms in the asymmetric unit. The assumed composition of the asymmetric unit of  $P\bar{1}$  is  $C_nX_2$ , and the exact and approximate probabilities  $p_+(2h|h)$  are displayed as functions of  $n$ . The solid line denotes the approximate probability, computed from (27), and the centres of the '+' signs denote the values of the exact probability, obtained from (24). The  $|E|$  values are taken as  $|E_{2h}| = |E_h| = 2.25$ . (a) The equal-atom case ( $X = C$ ). (b) A moderately heterogeneous asymmetric unit ( $X = Kr$ ,  $\rho = 6$ ).

An illustration of the effect of atomic heterogeneity is given in Fig. 3 and in Table 1, where the data used for the construction of this figure are summarized. The test asymmetric unit has the composition  $C_{30}X_2$ , with  $\rho = Z_X/Z_C$  varying from 1 to 15, the  $|E|$  values corresponding to Figs. 3(a), (b) and (c) being  $|E_h| = |E_{2h}| = 1.50, 2.00$  and  $2.25$  respectively. The high values of the exact probability, at large heterogeneities, are associated with the smooth transition into the region of the Harker-Kasper inequality

$$U_h^2 \leq \frac{1}{2}(1 + U_{2h}) \quad (28)$$

(Harker & Kasper, 1948), which has been observed in our calculations. However, the exact probabilities remain unity, or thereabout, below the inequality threshold as well.

As shown by the above illustrations, the behaviour of the exact probability (24) is qualitatively consistent with the known features of the  $\sum_1$  relationship, and its statistical interpretations. On the other hand, the

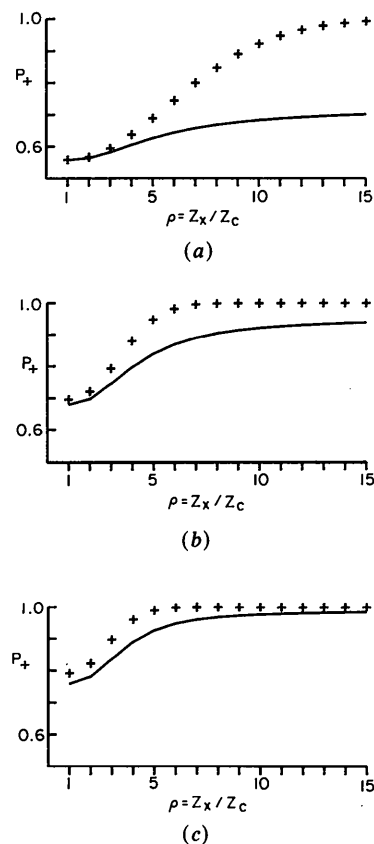


Fig. 3. Effects of atomic heterogeneity for selected  $|E|$  values. The assumed composition of the asymmetric unit of  $P\bar{1}$  is  $C_{30}X_2$ , and the exact and approximate probabilities  $p_+(2h|h)$  are displayed as functions of the ratio of atomic numbers,  $\rho = Z_X/Z_C$ . The solid line denotes the approximate probability, given by (27), and the centres of the '+' signs denote the values of the exact probability, obtained from (24). (a)  $|E_{2h}| = |E_h| = 1.50$ , (b)  $|E_{2h}| = |E_h| = 2.00$ , (c)  $|E_{2h}| = |E_h| = 2.25$ .

Table 1. Comparison of asymptotic and exact probabilities  $p_{\pm}(2\mathbf{h}|\mathbf{h})$ 

The atomic composition of the asymmetric unit was chosen for this comparison as  $C_{30}X_2$ , where the atomic number of  $X$  has been varied from  $Z_X = 6$  to  $Z_X = 90$ , in steps of 6. The first column contains the heterogeneity indicator  $\rho = Z_X/Z_C$ . The headings 'asympt.' refer to probabilities computed from (27), and 'exact' denote those obtained from (24), (25) and (26). The probabilities listed in each column pertain to the  $|E|$  values shown in the corresponding column heading.

$\rho$	$ E_{\mathbf{h}}  =  E_{2\mathbf{h}}  = 1.50$ $p_{+}(2\mathbf{h} \mathbf{h})$		$ E_{\mathbf{h}}  =  E_{2\mathbf{h}}  = 2.00$ $p_{+}(2\mathbf{h} \mathbf{h})$		$ E_{\mathbf{h}}  =  E_{2\mathbf{h}}  = 2.25$ $p_{+}(2\mathbf{h} \mathbf{h})$	
	asympt.	exact	asympt.	exact	asympt.	exact
1	0.5583	0.5589	0.6792	0.6961	0.7582	0.7912
2	0.5647	0.5671	0.6970	0.7208	0.7806	0.8217
3	0.5829	0.5946	0.7449	0.7930	0.8365	0.8960
4	0.6057	0.6370	0.7979	0.8807	0.8901	0.9603
5	0.6269	0.6886	0.8403	0.9474	0.9262	0.9901
6	0.6444	0.7449	0.8702	0.9821	0.9478	0.9984
7	0.6583	0.8002	0.8907	0.9953	0.9607	0.9999
8	0.6690	0.8497	0.9048	0.9991	0.9686	1.0000
9	0.6773	0.8907	0.9147	0.9999	0.9738	1.0000
10	0.6839	0.9231	0.9219	1.0000	0.9773	1.0000
11	0.6890	0.9476	0.9273	1.0000	0.9797	1.0000
12	0.6932	0.9656	0.9314	1.0000	0.9815	1.0000
13	0.6965	0.9783	0.9345	1.0000	0.9829	1.0000
14	0.6993	0.9869	0.9371	1.0000	0.9839	1.0000
15	0.7016	0.9925	0.9391	1.0000	0.9847	1.0000

asymptotic expression (27) may be quite inaccurate, even for moderate atomic heterogeneities and, more importantly, for the equal-atom case when the number of atoms is small.

It can be stated on the basis of the above results, and other calculations for different choices of the parameters, that the discrepancies between the exact and the asymptotic probabilities for the positive sign of  $E_{2\mathbf{h}}$  decrease as  $K$  increases. The region of greatest discrepancy, for a given  $K$ , seems to depend on the value of  $K$  itself. Thus, the discrepancy is greatest at high atomic heterogeneities for low values of  $K$  [cf. Fig. 3(a)], and the range of maximum discrepancy moves to lower heterogeneities as  $K$  increases.

The actual discrepancies may be somewhat different when the commonly used atomic-number approximation is replaced by properly averaged values of the atomic scattering factors (e.g. Shmueli, 1982). However, this is not likely to change the above picture to any appreciable extent.

A result of an immediate possible practical significance is the fact that the universally used asymptotic expression (27) often gives what are regarded as unacceptably low indications, in situations in which the values of the exact probability (24) would most usually be considered as acceptable. It therefore appears that the use of the exact equation (24), in estimating the reliability of  $\sum_1$  relationships, should very significantly promote the determination of the signs of structure seminvariants in the space group  $P1$  - whenever there is some atomic heterogeneity in the structure, or the number of (equal or unequal) atoms in the asymmetric unit of  $P1$  is small.

The above exact formalism, and the possibility of its evaluation to any accuracy that is required cannot, of course, be expected to provide any remedy for wrong sign indications that are due to features that have not been allowed for by the underlying assumptions (see above). The assumptions of central importance appear to be the uniform distribution of the atomic phase factors,  $2\pi\mathbf{h}\cdot\mathbf{r}_j$ , and the independence of the atomic contributions. While the requirement that all the atoms be located in general positions [assumption (i)] is in many cases sufficient to ensure (in practice) the uniform distribution of atomic phase factors, this may fail to an unknown extent when conspicuous rational dependence is present in the structure. Thus, for example, with all the atoms in general positions, but a heavy scatterer located at  $1/2, 1/2, 1/4$ , in the space group  $P\bar{1}$ , this uniformity may certainly be affected and hence also the validity of the probability formulae (see, for example, Hauptman & Karle, 1959). An investigation of this and other kinds of rational dependence (Hauptman, 1964), in the context of the present and other formalisms of structure-factor statistics, certainly seems to be of interest.

The validity of the almost universally made assumption that the atomic contributions to the structure factor are independent [assumption (iv)] has also been questioned (see, for example, Wilson, 1981) but, to the authors' knowledge, the effects that this assumption may have on the univariate or multivariate p.d.f.'s of structure factors have not been investigated to any depth. In any case, the removal of the assumptions concerning rational and statistical independence should involve an incorporation of known structural features into the statistical formalism, rather than their 'subtraction' - as might be done, for example, in the case of fixed special positions.

A comment on the computational aspects of the present work also seems to be in order. The statement that the first ten terms of (14) or (15) are sufficient to achieve satisfactory convergence pertains to outstandingly heavy atoms only, and fewer terms are needed in order to achieve the same convergence for atoms of the first few rows of the Periodic Table. The computing times for evaluating (24) are less than 0.1 c.p.u. s per relationship, and no need was felt in the present calculations of introducing computational acceleration techniques. Also, such techniques (although usually helpful) do not yet appear to be vital in the computation of the exact probability for the positive sign of the triple product  $E(\mathbf{h})E(\mathbf{k})E(\mathbf{h}+\mathbf{k})$  (Shmueli & Weiss, 1985). They may, however, be called for in similar studies of relationships of higher orders - at least with the computer that was used in this work.

Still in the computational context, we wish to point out that the main part of the computation, the evaluation of Bessel functions of general order, is

quite readily performed with the available polynomial approximations and recursion formulae (e.g. Abramowitz & Stegun, 1972). However, the numerical stability of such computations usually calls for a judicious use of forward and backward recursion techniques, which should be implemented in any program for computing arrays of Bessel functions of general order.

We wish to thank Professors C. Giacomazzo, H. Hauptman, A. J. C. Wilson and M. M. Woolfson for their encouraging comments and questions related to the first version of the present manuscript, and the referees of this paper for pointing out the need to elaborate on certain 'standard' assumptions. All the computations were done with the aid of a Cyber 170-855 at the Tel Aviv University Computation Center. The general-order Bessel functions were computed with the library subroutine *BESRI*, written by Mr David Sagin of the Computation Center.

## APPENDIX A

### Rederivation of the approximate $p_+(2\mathbf{h}|\mathbf{h})$

It is well known that the various forms of the central limit theorem can be derived from analytic properties of the characteristic function as  $\omega \rightarrow 0$  (e.g. Kendall & Stuart, 1963). For example, if one has a sum of identically distributed independent random variables,  $s_n = (x_1 + \dots + x_n)/n^{1/2}$ , where  $\langle x_i \rangle = 0$  and  $\langle x_i^2 \rangle = 1$ , the characteristic function of  $s_n$  can be expressed as  $C^n(\omega/n^{1/2})$ , in terms of the characteristic functions of the  $x$ 's. The central limit theorem is a consequence of expanding this function as

$$\begin{aligned} C^n(\omega/n^{1/2}) &= \exp [n \ln C(\omega/n^{1/2})] \\ &\sim \exp [n \ln (1 - \omega^2/2n + \dots)] \\ &\sim \exp (-\omega^2/2), \end{aligned} \quad (A1)$$

on retaining lowest-order terms in  $n$ . But this is equivalent to retaining just the lowest-order terms in an expansion of the characteristic function and exponentiating them. This strategy can also be carried out in the present case. Corrections to the Gaussian can be derived by expanding higher-order terms (in  $1/n$ ) in a Taylor series, leading to Edgeworth expansions (Giacovazzo, 1980).

To carry out this program, given the exact expressions in (13)–(15), let us consider the lowest-order terms coming from (14) and (15):

$$\begin{aligned} R_j(\omega_1, \omega_2) &\sim J_0(2n_j\omega_1)J_0(2n_j\omega_2) \\ &\sim (1 - n_j^2\omega_1^2)(1 - n_j^2\omega_2^2) \\ &\sim 1 - n_j^2(\omega_1^2 + \omega_2^2) \end{aligned} \quad (A2)$$

and

$$\begin{aligned} I_j(\omega_1, \omega_2) &\sim J_2(2n_j\omega_1)J_1(2n_j\omega_2) \\ &\sim -n_j^3\omega_1^2\omega_2. \end{aligned} \quad (A3)$$

The remaining Bessel functions give rise to higher-order corrections and are therefore neglected. Using these results we can write, for the lowest-order terms in the characteristic functions,

$$\begin{aligned} C_j(\omega_1, \omega_2) &\sim 1 - n_j^2(\omega_1^2 + \omega_2^2) - in_j^3\omega_1^2\omega_2 \\ &\sim \exp [-n_j^2(\omega_1^2 + \omega_2^2) - in_j^3\omega_1\omega_2], \end{aligned} \quad (A4)$$

where the approximation  $1 + z \sim \exp(z)$  has been used. Substituting (A4) into (9), we obtain the approximate characteristic function as

$$\begin{aligned} C(\omega_1, \omega_2) &\approx \prod_{j=1}^{N/2} \exp [-n_j^2(\omega_1^2 + \omega_2^2) - in_j^3\omega_1^2\omega_2] \\ &\sim \exp [-\frac{1}{2}(\omega_1^2 + \omega_2^2)] \\ &\quad \times [1 - i(\sigma_3/2\sigma_2^{3/2})\omega_1^2\omega_2], \end{aligned} \quad (A5)$$

where

$$\sigma_m = \sum_{j=1}^N f_j^m \quad (A6)$$

and the second line in (A5) has again been approximated by expanding the second exponential to lowest orders in  $\omega_1$  and  $\omega_2$ .

We can now obtain the asymptotic joint p.d.f. by Fourier inversion of the characteristic function in (A5):

$$\begin{aligned} p(E_{\mathbf{h}}, E_{2\mathbf{h}}) &\approx (1/4\pi^2) \iint_{-\infty}^{\infty} C(\omega_1, \omega_2) \\ &\quad \times \exp [-i(\omega_1 E_{\mathbf{h}} + \omega_2 E_{2\mathbf{h}})] d\omega_1 d\omega_2 \\ &\approx (1/2\pi) \exp [-\frac{1}{2}(E_{\mathbf{h}}^2 + E_{2\mathbf{h}}^2)] \\ &\quad \times [1 + (\sigma_3/2\sigma_2^{3/2})E_{2\mathbf{h}}(E_{\mathbf{h}}^2 - 1)]. \end{aligned} \quad (A7)$$

If the last factor, in square brackets, is again approximated by an exponential, we obtain

$$\begin{aligned} p(E_{\mathbf{h}}, E_{2\mathbf{h}}) &\sim \frac{1}{2\pi} \exp [-\frac{1}{2}(E_{\mathbf{h}}^2 + E_{2\mathbf{h}}^2) \\ &\quad + (\sigma_3/2\sigma_2^{3/2})E_{2\mathbf{h}}(E_{\mathbf{h}}^2 - 1)] \end{aligned} \quad (A8)$$

and the classical hyperbolic tangent formula for  $p_+(2\mathbf{h}|\mathbf{h})$  (Cochran & Woolfson, 1955; Klug, 1958), shown in (27) in the text, can be readily derived from (A8).

The linear approximation to the tangent formula

$$p_+(2\mathbf{h}|\mathbf{h}) \sim \frac{1}{2} + \frac{1}{4}\sigma_3\sigma_2^{-3/2}|E_{2\mathbf{h}}|(E_{\mathbf{h}}^2 - 1) \quad (A9)$$

(Hauptman & Karle, 1953; Giacomazzo, 1980) can be derived directly from (A7).

## References

- ABRAMOWITZ, M. & STEGUN, I. (1972). *Handbook of Mathematical Functions*. New York: Dover.
- BERTAUT, E. F. (1955a). *Acta Cryst.* **8**, 537-550.
- BERTAUT, E. F. (1955b). *Acta Cryst.* **8**, 823-832.
- COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1-12.
- CRAMER, H. (1951). *Mathematical Methods of Statistics*. Princeton Univ. Press.
- GIACOVAZZO, C. (1978). *Acta Cryst.* **A34**, 562-574.
- GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*. New York: Academic Press.
- GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Tables of Integrals, Series and Products*. New York: Academic Press.
- HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70-75.
- HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421-1433.
- HAUPTMAN, H. & KARLE, J. (1952). *Acta Cryst.* **5**, 48-59.
- HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*. ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.
- HAUPTMAN, H. & KARLE, J. (1959). *Acta Cryst.* **12**, 846-850.
- International Tables for X-ray Crystallography* (1952). Vol. I. Birmingham: Kynoch Press.
- KENDALL, M. G. & STUART, A. (1963). *The Advanced Theory of Statistics*, Vol. 1, 3rd ed. London: Charles Griffin.
- KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.
- SHMUELI, U. (1979). *Acta Cryst.* **A35**, 282-286.
- SHMUELI, U. (1982). *Acta Cryst.* **A38**, 362-371.
- SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* **A37**, 76-80.
- SHMUELI, U. & KALDOR, U. (1983). *Acta Cryst.* **A39**, 619-621.
- SHMUELI, U. & WEISS, G. H. (1985). In preparation.
- SHMUELI, U., WEISS, G. H. & KIEFER, J. E. (1985). *Acta Cryst.* **A41**, 55-59.
- SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* **A40**, 651-660.
- SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* **A37**, 342-353.
- VAND, V. & PEPINSKY, R. (1953). *The Statistical Approach to X-Ray Structure Analysis*. Pennsylvania State Univ. Press.
- WEISS, G. H. & KIEFER, J. E. (1983). *J. Phys. A*, **16**, 489-495.
- WEISS, G. H., SHMUELI, U., KIEFER, J. E. & WILSON, A. J. C. (1984a). *Acta Cryst.* **A40**, C419.
- WEISS, G. H., SHMUELI, U., KIEFER, J. E. & WILSON, A. J. C. (1984b). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON. New York: Adenine Press. In the press.
- WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.
- WILSON, A. J. C. (1952). *Nuovo Cimento*, **9**, 50-55.
- WILSON, A. J. C. (1981). *Acta Cryst.* **A37**, 808-810.

*Acta Cryst.* (1985). **A41**, 408-413

## One-Wavelength Technique: Some Probabilistic Formulas Using the Anomalous Dispersion Effect

BY G. CASCARANO AND C. GIACOVAZZO

*Centro di Ricerca Interdipartimentale di Cristallografia, Università di Bari, 70121 Bari, and Dipartimento Geomineralogico, Università Piazza Umberto I, 1-70121, Bari, Italy*

(Received 16 November 1984; accepted 14 March 1985)

### Abstract

The method of joint probability distributions has been applied to structure factors in order to assign the phases relative to the complete structure when the phases corresponding to the anomalous scatterers are known. Formulas have been obtained that generalize Sim's [*Acta Cryst.* (1959), **12**, 813-815] distribution to the case in which the anomalous dispersion effect is present.

### Notation

$p, q$  number of anomalous and non-anomalous scatterers respectively in the unit cell

$N$  number of atoms in the unit cell ( $N = p + q$ )

$f = f' + if''$  general expression for the atomic scattering factor

$F^+, F^-$  structure factors of the reflexions  $\mathbf{h}$  and  $-\mathbf{h}$  respectively

$F_p^+, F_p^-, F_q^+, F_q^-$  structure factors of the  $p$  anomalous scatterers and of the

$F'$

$F_p', F_p''$

$\sum = \sum_{j=1}^N (f_j'^2 + f_j''^2)$

$\varphi_j = f_j / \sum^{1/2}$

$\varphi_j', \varphi_j''$   
 $R = F / \sum^{1/2}$

$q$  non-anomalous scatterers for the reflexions  $\mathbf{h}$  and  $-\mathbf{h}$  respectively

structure factor (imaginary component of anomalous dispersion omitted)

contribution of anomalous scatterers due to the real and to the imaginary component of anomalous dispersion

$$F_p' = \sum_{j=1}^p f_j' \exp 2\pi i \mathbf{h} \cdot \mathbf{r}_j$$

$$F_p'' = \sum_{j=1}^p f_j'' \exp 2\pi i \mathbf{h} \cdot \mathbf{r}_j$$

average value of  $|F_{\mathbf{h}}|^2$  at given  $|\mathbf{h}|$

normalized scattering factor of the  $j$ th atom

real and imaginary parts of  $\varphi_j$   
normalized structure factor