and x_{il} is a coordinate of the lattice, so that its variation corresponds to equal shifts of all the atoms in the A lattice. Because $\partial E_{Al-(Bl+Cl+\ldots)}/\partial x_l$ can be imagined to be a sum of equal terms, one for each point in the lattice, all equal to $\partial E_{A-(B+C+\ldots)}/\partial x_i$, it follows that each one of these terms must be zero, and therefore also in this case $\partial E/\partial x_l = 0$ (Fig. 2).

References

BERTAUT, F. (1952). *J. Phys. (Paris)*, **13**, 499–505.
BERTAUT, F. (1978). *J. Phys. (Paris)*, **39**, 1331–1348.
CRUICKSHANK, D. W. J. (1957). *Acta Cryst.* **10**, 504–508.
EWALD, P. P. (1921). *Ann. Phys. (Leipzig)*, **64**, 253–287.
FILIPPINI, G., GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1973). *J. Chem. Phys.* **59**, 5088–5101.
FILIPPINI, G., GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1974*a*). *Acta Cryst.* A**30**, 189–196.
FILIPPINI, G., GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1974*b*). *Chem. Phys. Lett.* **26**, 301–304.
FILIPPINI, G., GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1975*a*). *Chem. Phys.* **8**, 136–146.
FILIPPINI, G., GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1975*b*). *Chem. Phys. Lett.* **35**, 17–20.
FILIPPINI, G., GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1976). *Acta Cryst.* A**32**, 259–264.
GAVEZZOTTI, A., MUGNOLI, A., RAIMONDI, M. & SIMONETTA, M. (1972). *J. Chem. Soc. Perkin Trans. 2*, pp. 425–431.
GRAMACCIOLI, C. M., SIMONETTA, M. & SUFFRITTI, G. B. (1973). *Chem. Phys. Lett.* **20**, 23–28.
JONES, R. E. & TEMPLETON, D. H. (1956). *J. Chem. Phys.* **25**, 1062–1063.
KERTESZ, M., KOLLER, J. & AZMAN, A. (1980). In *Recent Advances in the Quantum Theory of Polymers*, p. 62. Berlin: Springer.
KITAIGORODSKII, A. I. (1966). *J. Chim. Phys.* **63**, 9–16.
KROON, P. A. & VOS, A. (1978). *Acta Cryst.* A**34**, 823–824.
LEIBFRIED, G. & LUDWIG, W. (1961). *Solid State Phys.* **12**, 275–444.
LONSDALE, K. & MILLEDGE, J. (1961). *Acta Cryst.* **14**, 54–61.
LUTZ, U. A. & HÄLG, W. (1970). *Solid State Commun.* **8**, 165.
MASON, R. (1964). *Acta Cryst.* **17**, 547–555.
NATKANIEC, I., BOKHENKOV, E. L., DORNER, B., KALUS, J., MACKENZIE, G. A., PAWLEY, G. S., SCHMELZER, U. & SHEKA, E. F. (1980). *J. Phys. C*, **13**, 4265–4283.
NORDMAN, C. E. & SCHMITKONS, D. L. (1965). *Acta Cryst.* **18**, 764–767.
PAWLEY, G. S. (1967). *Phys. Status Solidi*, **20**, 347–360.
PAWLEY, G. S., MACKENZIE, G. A., BOKHENKOV, E. L., SHEKA, E. F., DORNER, B., KALUS, J., SCHMELZER, U. & NATKANIEC, I. (1980). *Mol. Phys.* **39**, 251–260.
TADDEI, G., BONADEO, H., MARZOCCHI, N. P. & CALIFANO, S. (1973). *J. Chem. Phys.* **58**, 966–978.
VENKATARAMAN, G., FELDKAMP, L. A. & SAHNI, V. C. (1975). *Dynamics of Perfect Crystals*, p. 77. Cambridge, Mass: The MIT Press.
WILLIAMS, D. E. (1967). *J. Chem. Phys.* **47**, 4680–4684.
WILLIAMS, D. E. (1971). *Acta Cryst.* A**27**, 452–455.
WILLIAMS, D. E. (1972). *Acta Cryst.* A**28**, 629–635.
WILLIAMS, D. E. (1974). *Acta Cryst.* A**30**, 71–77.

Acta Cryst. (1981). A**37**, 342–353

Effects of Space-Group Symmetry and Atomic Heterogeneity on Intensity Statistics

BY U. SHMUELI

Department of Chemistry, Tel-Aviv University, 69978 Ramat Aviv, Israel

AND A. J. C. WILSON

Department of Physics, University of Birmingham, Birmingham B15 2TT, England

(*Received* 3 *October* 1980; *accepted* 18 *November* 1980)

Abstract

Probability functions, moments and cumulative distributions of the normalized intensity $z = |E|^2$, depending on the space-group symmetry of the crystal and its chemical composition, have been investigated. The probability functions were reduced to a simple, unified representation, by reconsidering some mathematical properties of the corresponding asymptotic expansions. A subsequent unified derivation of the first four even moments of $|E|$, in terms of symmetry and composition, leads to (i) simple and readily computable expressions for $\langle|E|^4\rangle$, $\langle|E|^6\rangle$ and $\langle|E|^8\rangle$ and (ii) a significant simplification of the expansion coefficients which appear in the above asymptotic expansions. The convergence of these expansions is discussed and illustrated by a numerical example. It is shown that the Edgeworth arrangement of these asymptotic expansions is superior to the frequently given Gram–Charlier one. The dependence of the fourth moment of $|E|$ on

0567-7394/81/030342-12$01.00

atomic heterogeneity and the generalized cumulative distribution functions $N(|E|)$ are illustrated for all the symmorphic space groups. The results of this study are directly applicable to practical intensity statistics for structures containing all the atoms in general positions.

Introduction

Intensity statistics, based on the assumptions that (i) the unit cell comprises a large number of atoms, all in general positions, (ii) there are no predominantly heavy atoms in the cell and (iii) there is no symmetry in the structure other than lattice translations and possibly a crystallographic center of symmetry (Wilson, 1949), are nowadays applied in a routine manner to the verification of space-group assignment and to the resolution of space-group ambiguities. For example, a wide range of statistical tests, based on the above assumptions, is computed by the widely used *MULTAN* system (Main, Hull, Lessinger, Germain, Declercq & Woolfson, 1978) and these are successful in many applications. However, the results of such tests may lead to incorrect conclusions if any of the above assumptions is severely violated. Such disturbing factors as hypersymmetry (Lipson & Woolfson, 1952; Rogers & Wilson, 1953), effects of symmetry and composition, and heavily contributing scatterers in special positions (Foster & Hargreaves, 1963*a*) have been the subject of numerous publications and an extensive review was given by Srinivasan & Parthasarathy (1976).

Two recent contributions (Wilson, 1978; Shmueli, 1979) have enabled one to extend the practical statistical tests by taking into account the effects of crystallographic symmetry and composition of the asymmetric unit on the intensity distribution. The fourth moment of the structure amplitude (and hence the second moment of the normalized intensity) was evaluated by Wilson (1978) for all space groups except *Fd3m* and *Fd3c*. This quantity also depends on the composition of the asymmetric unit and thus a generalized variance test is now available. Another important test, the cumulative distribution function $N(|E|)$ [or $N(z)$] (Howells, Phillips & Rogers, 1950) was generalized by Shmueli (1979) and made to depend on crystallographic symmetry and composition of the asymmetric unit. However, the numerical results were applicable to triclinic, monoclinic and orthorhombic systems only. The two contributions are related, in that the symmetry and composition dependence of the generalized $N(|E|)$ is expressed in terms of moments of $|E|$ alone (Shmueli, 1979).

The present paper is an extension and a unified presentation of the two approaches. It is shown that the first three terms of the expansion for $N(|E|)$ (Shmueli, 1979) are needed for a meaningful representation of the function and thus a generalization of the sixth moment of $|E|$ is also required. The approach used by Wilson (1978) for the derivation of $\langle|F|^4\rangle$ is here extended to $\langle|F|^6\rangle$ and also to $\langle|F|^8\rangle$ wherefrom the related moments follow readily. The underlying probability density functions are presented in considerably simplified forms which lend themselves readily to an interpretation in terms of space-group symmetry. Finally, numerical values required for the calculation of the moments, and hence of $N(|E|)$, are presented for a wide range of symmetries and the expected deviations from the limiting values of the moments (Wilson, 1949) and from the limiting distributions (Howells, Phillips & Rogers, 1950) are illustrated.

Distribution functions

Expansions for the probability density function of the structure amplitude $|F|$, which account for deviations from the Wilson (1949) statistics due to symmetry and atomic heterogeneity, were derived for centrosymmetric (Karle & Hauptmann, 1953) and non-centrosymmetric (Hauptmann & Karle, 1953) crystals. It was shown by Rogers & Wilson (1953) that the expansion for the centrosymmetric case is a Gram–Charlier type *A* asymptotic expansion, and much attention was given to this expansion by other authors (Bertaut, 1955; Klug, 1958; Mitra & Belgaumkar, 1973; Srinivasan & Parthasarathy, 1976). The expansion for the non-centrosymmetric case was rederived by Srinivasan & Parthasarathy (1976), who also represented both expansions in terms of the normalized structure amplitude and normalized intensity. While the centrosymmetric expansion can be conveniently expressed in terms of Hermite polynomials and moments of the variable (Shmueli, 1979), the expressions so far published for the non-centrosymmetric one appear rather cumbersome. A simplified representation of the latter has been achieved during this work and will be shown below.

We summarize, in what follows, the expansions for the probability density functions of the normalized intensity $z = |F|^2/\langle|F|^2\rangle$, their dependence on the moments of z and the cumulative distribution functions, derived from these expansions.

The centrosymmetric case

The probability density function is given by

$$P_c(z) = (2\pi z)^{-1/2}\exp\left(-\frac{z}{2}\right)\left\{1 + \frac{A_4}{4!2^2}H_4\left[\left(\frac{z}{2}\right)^{1/2}\right] + \frac{A_6}{6!2^3}H_6\left[\left(\frac{z}{2}\right)^{1/2}\right] + \frac{A_8}{8!2^4}H_8\left[\left(\frac{z}{2}\right)^{1/2}\right] + \ldots\right\}, \quad (1)$$

where $H_n(x)$ is a Hermite polynomial of order n and the coefficients A_4, A_6 and A_8 are linear combinations of the moments of z:

$$A_4 = \langle z^2 \rangle - 3, \tag{2}$$

$$A_6 = \langle z^3 \rangle - 15\langle z^2 \rangle + 30, \tag{3}$$

$$A_8 = \langle z^4 \rangle - 28\langle z^3 \rangle + 210\langle z^2 \rangle - 315. \tag{4}$$

The cumulative distribution function of z is obtained by evaluating the integral $\int_0^z P_c(z')dz'$ which leads to

$$N_c(z) = \text{erf}\left[\left(\frac{z}{2}\right)^{1/2}\right] - \frac{2}{\sqrt{\pi}}\exp\left(-\frac{z}{2}\right)\left\{\frac{A_4}{4!2^2}H_3\left[\left(\frac{z}{2}\right)^{1/2}\right] + \frac{A_6}{6!2^3}H_5\left[\left(\frac{z}{2}\right)^{1/2}\right] + \frac{A_8}{8!2^4}H_7\left[\left(\frac{z}{2}\right)^{1/2}\right] + \ldots\right\}, \tag{5}$$

where the first term is the limiting distribution function of Howells, Phillips & Rogers (1950).

In Appendix A, we show the definite integral which permits the derivation of (2)–(4) from (1) and the integral leading from (1) to (5).

The non-centrosymmetric case

The probability density function is given by

$$P_a(z) = \exp(-z)\left\{1 + \frac{B_4}{2!}L_2(z) - \frac{B_6}{3!}L_3(z) + \frac{B_8}{4!}L_4(z) - \ldots\right\}, \tag{6}$$

where

$$L_n(z) = \frac{1}{n!}\exp(z)\frac{d^n}{dz^n}[z^n \exp(-z)]$$

is a Laguerre polynomial of order n (Abramowitz & Stegun, 1972) and the coefficients B_4, B_6 and B_8 are related to the moments of z by the following equations:

$$B_4 = \langle z^2 \rangle - 2, \tag{7}$$

$$B_6 = \langle z^3 \rangle - 9\langle z^2 \rangle + 12, \tag{8}$$

$$B_8 = \langle z^4 \rangle - 16\langle z^3 \rangle + 72\langle z^2 \rangle - 72. \tag{9}$$

The cumulative distribution function of z, obtained by integrating (6), is

$$N_a(z) = 1 - \exp(-z) + \exp(-z) \times \sum_{k=2}^{4} \frac{(-1)^k B_{2k}}{k!}[L_{k-1}(z) - L_k(z)] + \ldots. \tag{10}$$

As above, the first two terms on the r.h.s. of (10) are just the acentric $N(z)$ distribution function of Howells, Phillips & Rogers (1950).

The integrals which permit the derivation of (7)–(9) and (10), from (6), are given in Appendix A.

Equations (1) and (6) are equivalent to other published expansions for the probability density functions, except for the explicit introduction of orthogonal polynomials instead of writing them in full. This representation results not only in conciseness, important for coding the expressions, but also renders the derivations much more transparent. For example, it is clear from the integrals ($A1$) and ($A2$) in Appendix A, that only the first n terms of the expansion are relevant for the calculation of the nth moment of z. The expressions for the expansion coefficients in terms of the moments of z, except those depending on $\langle z^4 \rangle$, and the first three terms of either the cumulative distribution functions, are equivalent to those given by Shmueli (1979).

It should be added that the cumulative distribution functions of the normalized structure amplitude $|E|$ are obtained by simply replacing z with $|E|^2$ throughout (5) and (10). Furthermore, the probability density functions for $|F|$ and $|E|$ are related to $P(z)$ as follows:

$$P(|F|) = \frac{2z}{|F|}P(z), \quad P(|E|) = 2|E|P(z). \tag{11}$$

The normalized intensity is then reexpressed in terms of $|F|$ or $|E|$, as required.

A comment on the functional form of (6) may be in order. The limiting probability density function $P^{(0)}(z) = \exp(-z)$ is a special form of a Pearson type III frequency function (Cramér, 1945). As shown by Cramér (1945), the orthogonal polynomials which are associated with an expansion of a frequency function in terms of special forms of the above, including $\exp(-z)$, are the Laguerre polynomials. The functional form of (6) is consistent with the above remark and its formal rederivation from these considerations appears to be possible. Such a derivation is, however, outside the scope of this paper.

Convergence of the $N(z)$ asymptotic expansions

The convergence of formal expansions of the type just considered has been treated by Cramér (1928 and, more accessibly, 1945, pp. 213–231, 1970, chapter, VII). Roughly, the conditions amount to (i) all moments of the distribution must be finite; (ii) at least the first three absolute moments must be finite; and (iii) $P(z)$ must approach zero fast enough for the integral

$$\int_{-\infty}^{\infty} \exp(z^2/4)P(z)dz \tag{12}$$

to be finite. These conditions are clearly satisfied for the intensity distributions since $P(z)$ is identically zero for $z > \Phi/\Sigma$, where

$$\Phi = \left| \sum_{i=1}^{n} f_i \right|^2 \quad \text{and} \quad \Sigma = \sum_{i=1}^{n} f_i^2 \qquad (13)$$

and dispersion has been neglected. Considered as functions of z, then, expansions like (1) and (5) will converge, and when the terms are properly arranged (the so-called Edgeworth form of the series), the series is asymptotic in n, the number of atoms in the unit cell. The error in stopping at any term is then of lower order than the last term kept. However, for practical purposes, the important question is whether the available terms (the first three for all space groups) give a reasonable representation of $P(z)$ and $N(z)$. This is investigated numerically for the space groups $P1$ and $P\bar{1}$; as will be seen later (Table 2 and Figs. 1 and 2) these are particularly unfavorable cases. Moreover, it is important to find out what is the minimum number of expansion terms which may give such a representation. All four terms may, if required, be evaluated for triclinic, monoclinic and orthorhombic space groups, from the first four moments of z, given by Foster & Hargreaves (1963*b*) and Srinivasan & Parthasarathy (1976), and we shall do so in the present calculation.

The cumulative distribution functions will be expressed for this purpose as

$$N(z) = N^{(0)}(z) + N_2 + N_3 + N_4 \qquad (14)$$

where $N^{(0)}(z)$ is the distribution based on the Wilson (1949) statistics and N_k is an expansion term depending on moments of z up to $\langle z^k \rangle$ [*cf.* (5) and (10)]. The importance of the successive terms can be assessed by evaluating the ratios $\rho_2 = N_2/N(z)$, $\rho_3 = N_3/N(z)$ and $\rho_4 = N_4/N(z)$ which are shown in Table 1. The values of ρ_k in Table 1 are based on (5) and (10) for $P\bar{1}$ and $P1$ respectively, the required moments of z being estimated as shown by Shmueli (1979) for a $C_{20}X$ asymmetric unit. The ratios Z_X/Z_C between the atomic numbers of the heavy atom X and a C atom, cover a wide range of atomic heterogeneity (Z_X = 30 to 78) and seem to be sufficient for our purpose.

It is seen from Table 1 that the successive percentage contributions ρ_2, ρ_3 and ρ_4 usually decrease for given z, except for high values of z and high heterogeneities in the non-centrosymmetric case. The apparent convergence, in the range of not too high z values, of course improves with decreasing heterogeneity since then, as will be seen from later parts of the paper, the asymptotic nature of the expansion becomes increasingly well-defined. The contribution of the N_2 term in (14) is the most important one; however, it appears that the inclusion of this term alone would underestimate the effect quite considerably. It is therefore believed that the simplest expansion to be of practical value is a three-term expansion, and the omission of the fourth term, N_4, does not appear to be harmful in calculations not requiring a very high accuracy. Such three-term cumulative distributions were given by Shmueli (1979) for triclinic, monoclinic and orthorhombic space groups, except for *Fdd*2 and *Fddd*, and their extension to higher symmetries is now possible in view of the present results and the availability of numerical values for the fourth and sixth moments of the trigonometric structure factor (Wilson, 1978; Shmueli & Kaldor, 1981). As will be seen in the next section, the inclusion of the fourth term N_4 is also possible, if need arises.

Calculation of higher moments of $|E|$

In this section and in Appendix *B* we extend Wilson's (1978) derivation of the fourth moment of $|F|$ to the sixth moment, shown to be required for a reasonably

Table 1. *Relative contributions of the successive expansion terms in* (5) *and* (10) *to* $N(z)$, *based on a* $C_{20}X$ *asymmetric unit and four-term expansions* ($\times 100$)

All the symbols are defined in the text.

	$Z_X/Z_C = 5$			$Z_X/Z_C = 9$			$Z_X/Z_C = 13$			
z	ρ_2	ρ_3	ρ_4	ρ_2	ρ_3	ρ_4	ρ_2	ρ_3	ρ_4	
0·25	−5·0	−2·8	−1·1	−10·1	−8·3	−4·9	−12·6	−11·5	−7·6	
0·50	−4·2	−2·1	−0·7	−8·5	−6·2	−3·3	−10·6	−8·6	−5·1	
0·75	−3·5	−1·5	−0·4	−7·0	−4·5	−2·0	−8·7	−6·2	−3·0	Space group:
1·00	−2·8	−1·0	−0·2	−5·7	−3·1	−1·0	−7·1	−4·2	−1·5	$P\bar{1}$
1·25	−2·3	−0·6	−0·0	−4·6	−1·9	−0·2	−5·7	−2·6	−0·3	
1·50	−1·8	−0·3	0·1	−3·6	−1·0	0·3	−4·4	−1·3	0·5	
0·25	−12·3	−7·7	−3·2	−24·8	−23·0	−14·8	−30·8	−31·9	−23·0	
0·50	−9·2	−4·8	−1·6	−18·6	−14·4	−7·3	−23·1	−19·9	−11·4	
0·75	−6·7	−2·6	−0·5	−13·5	−7·9	−2·4	−16·8	−11·0	−3·7	Space group:
1·00	−4·6	−1·1	0·1	−9·4	−3·3	0·6	−11·6	−4·6	1·0	$P1$
1·25	−3·0	−0·1	0·5	−6·1	−0·2	2·3	−7·5	−0·3	3·5	
1·50	−1·7	0·6	0·6	−3·5	1·9	2·9	−4·3	2·6	4·6	

accurate representation of the asymptotic expansions, and to the eighth moment which is needed if an evaluation of all the available terms is planned. The essential difference between the present derivation and that of Foster & Hargreaves (1963*a*) is that the latter authors express the averages in terms of mixed moments of the real and imaginary parts of the trigonometric structure factor J [Appendix B, equation ($B2$)] while, in the present approach, powers of the modulus of J are averaged directly. The latter approach leads to appreciably simpler expressions for moments *and* distributions, and was also found to be readily applicable to numerical computations of the moments of $|J|$ (Shmueli & Kaldor, 1981).

The actual derivation, preceded by an outline of Wilson's (1978) approach, is presented in Appendix B. Making use of ($B17$)–($B19$) and ($B20$)–($B23$) we obtain the expressions for the moments of $|E|$ (or z), relevant to this study.

$$\langle |E|^4 \rangle = \langle z^2 \rangle = L + a\frac{S_4}{S_2^2} \tag{15}$$

$$\langle |E|^6 \rangle = \langle z^3 \rangle = M + Na\frac{S_4}{S_2^2} + c\frac{S_6}{S_2^3} \tag{16}$$

$$\langle |E|^8 \rangle = \langle z^4 \rangle = R + Sa\frac{S_4}{S_2^2} + Tc\frac{S_6}{S_2^3} + Ua^2\frac{S_4^2}{S_2^4} + g\frac{S_8}{S_2^4}, \tag{17}$$

where

$$S_n = \sum_j f_j^n, \tag{18}$$

$$a = \frac{q}{p^2} - L, \tag{19}$$

$$c = \frac{r}{p^3} - \frac{Nq}{p^2} + 2M, \tag{20}$$

$$g = \frac{s}{p^4} - \frac{Uq^2}{p^4} - \frac{Tr}{p^3} + \frac{2Sq}{p^2} - 6R, \tag{21}$$

the symbols p, q, r and s denote the second, fourth, sixth and eighth moments of the trigonometric structure factor modulus $|J|$ and the constants L, M, N, R, S, T and U, derived in the Appendix, are given below:

L	M	N	R	S	T	U	
3	15	15	105	210	28	35	centrosymmetric
2	6	9	24	72	16	18	non-centrosymmetric.

The values of p, q and r in (19) and (20) are now available for all the space groups (*cf.* Table 2; Wilson, 1978; Shmueli & Kaldor, 1981) and the eighth moment s in (21) can also be evaluated if needed (see below).

Before proceeding to possible applications of the above results, it is of interest to discuss the relationship they bear to the expansion coefficients of the probability density functions (1) and (6). It is now seen that (2)–(4) and (7)–(9) can be unified as follows:

$$A_4 \text{ or } B_4 = \langle z^2 \rangle - L, \tag{22}$$

$$A_6 \text{ or } B_6 = \langle z^3 \rangle - N\langle z^2 \rangle + 2M, \tag{23}$$

$$A_8 \text{ or } B_8 = \langle z^4 \rangle - T\langle z^3 \rangle + S\langle z^2 \rangle - 3R. \tag{24}$$

If, further, (15), (16), (17) are substituted into (22)–(24), a large number of cancellations occurs and a remarkably simple result is obtained:

$$A_4 \text{ or } B_4 = a\frac{S_4}{S_2^2}, \tag{25}$$

$$A_6 \text{ or } B_6 = c\frac{S_6}{S_2^3}, \tag{26}$$

$$A_8 \text{ or } B_8 = g\frac{S_8}{S_2^4} + Ua^2\frac{S_4^2}{S_2^4}, \tag{27}$$

which, when substituted into the asymptotic expansions (1), (5), (6) and (10), makes their interpretation in terms of symmetry, composition and the number of atoms in the asymmetric unit much easier than in any of the equivalent expansions so far published.

Practical applications of the above results to a computation of moments and/or cumulative distribution functions of $|E|$ call for the values of p, q, r and s or the symmetry-dependent coefficients a, c and g. Until recently, these values were available for the triclinic, monoclinic and orthorhombic space groups only, with the exception of *Fdd*2 and *Fddd* (Foster & Hargreaves, 1963*b*; Srinivasan & Parthasarathy, 1976). The less-trivial space groups were first treated by Wilson (1978) whose compilation of q ($=\langle |J|^4 \rangle$) values encompasses all the space groups except *Fd*3*m* and *Fd*3*c*. During the course of this work, some attempts were made at a calculation of r ($=\langle |J|^6 \rangle$) values but it soon became clear that this cannot be done efficiently by straightforward algebra and trigonometry and some computer-oriented approach is desirable. Two algorithms for the computation of even moments of $|J|$ were developed and the values of q/p^2 and r/p^3 were obtained for all the space groups and all the *hkl* subsets leading to different values of the moments of $|J|$ (Shmueli & Kaldor, 1981). These algorithms are also capable of dealing with the eighth moment of $|J|$ but no extensive computations have as yet been made of the latter quantity. The problem of evaluating three-term asymptotic expansions as well as $\langle z^2 \rangle$ and $\langle z^3 \rangle$ for any space group is thus solved and four-term expansions can also be treated if need arises.

Table 2. *Second, fourth and sixth moments of the trigonometric structure factor and coefficients for the calculation of generalized fourth and sixth moments of the normalized structure amplitude* $|E|$

All symmorphic space groups with *P*-type Bravais lattices are included in the table. The calculation of the moments is reported elsewhere (Wilson, 1978; Shmueli & Kaldor, 1981). The coefficients *a* and *c* are defined by (19) and (20) respectively.

Space group	*P*	*q*	*r*	*a*	*c*
$P\bar{1}$	2	6	20	−3/2	10
*P*2/*m*	4	36	400	−3/4	5/2
Pmmm	8	216	8000	3/8	−5
*P*4/*m*	8	216	8000	3/8	−5
*P*4/*mmm*	16	1008	102400	15/16	−65/16
$P\bar{3}$	6	90	1860	−1/2	10/9
$P\bar{3}m$	12	396	21360	−1/4	10/9
*P*6/*m*	12	540	40800	3/4	−97/36
*P*6/*mmm*	24	2376	463200	9/8	235/144
*Pm*3	24	1800	225600	1/8	−5/9
*Pm*3*m*	48	8784	3463680	13/16	595/144
*P*1	1	1	1	−1	4
*P*2	2	6	20	−1/2	1
*P*222	4	28	256	−1/4	1/4
*P*4	4	36	400	1/4	−2
*P*422	8	136	3392	1/8	−1/2
*P*3	3	15	93	−1/3	4/9
*P*32	6	66	996	−1/6	1/9
*P*6	6	90	2040	1/2	19/18
*P*622	12	324	13800	1/4	−19/72
*P*23	12	276	9120	−1/12	1/36
*P*432	24	1272	111552	5/24	14/72
Pm	2	6	20	−1/2	1
*Pmm*2	4	36	400	1/4	−2
$P\bar{4}$	4	28	256	−1/4	1/4
*P*4*mm*	8	168	5120	5/8	−15/8
$P\bar{4}2m$	8	136	3392	1/8	−1/2
*P*3*m*	6	66	1068	−1/6	4/9
$P\bar{6}$	6	90	1860	1/2	−10/9
*P*6*mm*	12	396	23160	3/4	47/72
$P\bar{6}2m$	12	396	21360	3/4	−7/18
$P\bar{4}3m$	24	1272	123072	5/24	37/36

Table 2 shows the second, fourth and sixth moments of the trigonometric structure factor, computed with the above-mentioned algorithms (Shmueli & Kaldor, 1981), for all the symmorphic space groups with primitive lattices. The values of *a* and *c* to be used with (19) and (20) and/or (25) and (26) are also included. The *q* values agreed exactly with those found by Wilson (1978) and the *a* and *c* values for the triclinic, monoclinic and orthorhombic space groups included in Table 2, agree with those obtainable from the results of Foster & Hargreaves (1963*b*). The results given in Table 2 will be used in the illustrative examples given below.

Illustration of the effects of symmetry and composition

We first consider the two extreme cases, *i.e.* the equal-atom case and that of a single very heavy scatterer and *m* equal light atoms in the asymmetric unit. It is convenient to rewrite, for this purpose, (15) and (16) as

$$\langle |E|^4 \rangle = L\left(1 - \frac{S_4}{S_2^2}\right) + \frac{q}{p^2}\frac{S_4}{S_2^2}, \tag{28}$$

$$\langle |E|^6 \rangle = M\left(1 - 3\frac{S_4}{S_2^2} + 2\frac{S_6}{S_2^3}\right) + \frac{Nq}{p^2}\left(\frac{S_4}{S_2^2} - \frac{S_6}{S_2^3}\right) + \frac{r}{p^3}\frac{S_6}{S_2^3}. \tag{29}$$

The composition-dependent ratios are now given by

$$\frac{S_{2n}}{S_2^n} = \frac{m + \rho^{2n}}{(m + \rho^2)^n}, \tag{30}$$

where ρ is the ratio of the scattering factor of the heavy atom to that of a light atom.

In the equal-atom case, we have $S_4/S_2^2 = 1/(m + 1)$ and $S_6/S_2^3 = 1/(m + 1)^2$. For moderately large *m*, the effects of symmetry become very small and the fourth and sixth moments of $|E|$ are close to *L* and *M* respectively, as predicted by Wilson's (1949) statistics.

If, on the other hand, the ratio ρ becomes very large, both S_4/S_2^2 and S_6/S_2^3 tend to the value of unity and the fourth and sixth moments of $|E|$ approach q/p^2 and r/p^3 respectively. It is readily seen from Table 2 that different space groups may be affected by atomic heterogeneity in entirely different ways. Thus, for example for $P\bar{1}$, $q/p^2 = 1\frac{1}{2}$ (*cf.* Bertaut, 1955) and $r/p^3 = 2\frac{1}{2}$, indicating a very strong influence of the heavy scatterer, and for *Pmmm*, $q/p^2 = 3\frac{3}{8}$ and $r/p^3 = 15\frac{5}{8}$, suggesting a small effect of the atomic heterogeneity on the moments of $|E|$. Of course, the above consideration, like the rest of this paper, is valid only if the heavy scatterer is located in a general position of the space group in question.

The effects of both space-group symmetry and composition of the asymmetric unit on intensity statistics can be concisely illustrated by plotting the fourth moment of $|E|$, as given by (15) and with the coefficient *a* taken from Table 2, for all the space groups appearing in the table and for a wide range of compositions. Such an illustration of these effects is shown in Fig. 1. The asymmetric unit chosen for these calculations contains twenty equal light atoms and one heavy atom so that the composition dependence can still be expressed as a function of ρ [*cf.* (28)]. The relevant fourth moments of the trigonometric structure factor, shown in Table 2, have been obtained by a direct averaging of $|J|^4$ (Wilson, 1978) and by programmed structure-factor algebraic and symbol-handling procedures (Shmueli & Kaldor, 1981).

The deviations of $\langle |E|^4 \rangle$ curves from the horizontal lines in Fig. 1, which represent the limiting values of the

moments based on the Wilson (1949) statistics, are proportional to the coefficients a in (19), and by comparing Fig. 1 with Table 2 most space groups can be identified in this overall illustration. A similar illustration can be readily prepared for the sixth moment of $|E|$; it shows a much stronger dependence on composition, but is qualitatively similar to Fig. 1.

It is seen that, in the above example, the limiting values of $\langle|E|^4\rangle$ are good indicators of the presence of a center of symmetry for all space groups and for $\rho < 3$, and remain so for some space groups and all the range of ρ shown. However, the space groups based on the point groups 1, $2/m$, $4mm$, $6mm$ and a few others, produce moments of $|E|$ which are not comparable with the limiting moments and are thus the most likely candidates for (avoidable) problems with their recognition.

It may also be pointed out that the space groups $P4/mmm$ and $P6/mmm$ are associated with an $\langle|E|^4\rangle$ *vs* ρ dependence which, if it were observed, would be indicative of the presence of hypersymmetry in the structure.

An illustration of the effect of space-group symmetry on the cumulative distribution function of $|E|$ is given in Fig. 2. The distributions shown were evaluated from (5) and (10), neglecting the highest-order terms which depend on the eighth moment of $|E|$. The composition chosen for Fig. 2 corresponds to a $C_{20}I$ asymmetric unit ($\rho = 8{\cdot}83$) and all the space groups appearing in Table 2 are included in the generalized cumulative distribution functions shown in the figure.

The deviations of the generalized $N(|E|)$ distributions (dashed lines) from the limiting distributions based on the Wilson (1949) statistics (full lines) are

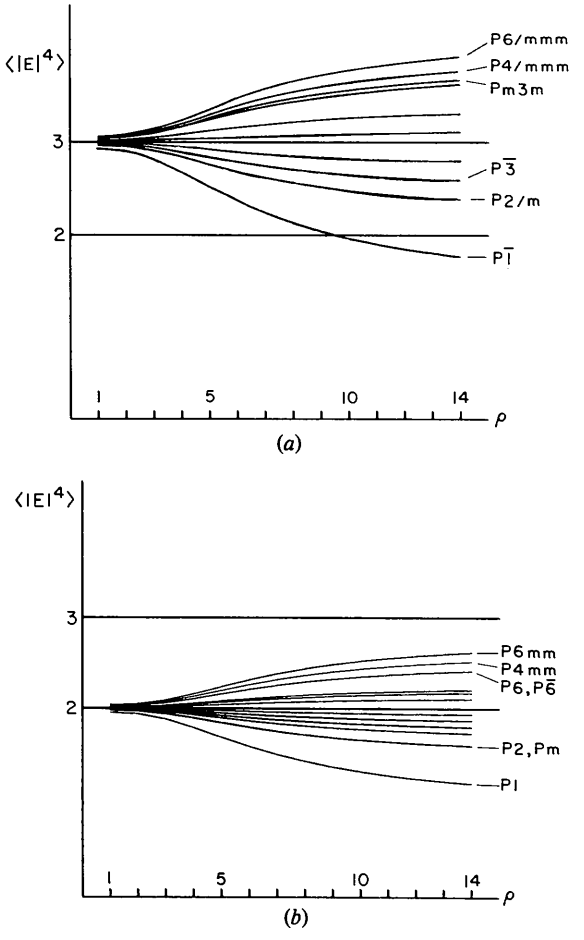

Fig. 1. Dependence of the fourth moment of $|E|$ on space-group symmetry and atomic heterogeneity. The moments were evaluated as described in the text for (*a*) centrosymmetric and (*b*) non-centrosymmetric space groups. The horizontal lines at $\langle|E|^4\rangle = 2$ and 3 correspond to the limiting distributions for non-centrosymmetric and centrosymmetric space groups respectively. For the definition of the abscissa see text.

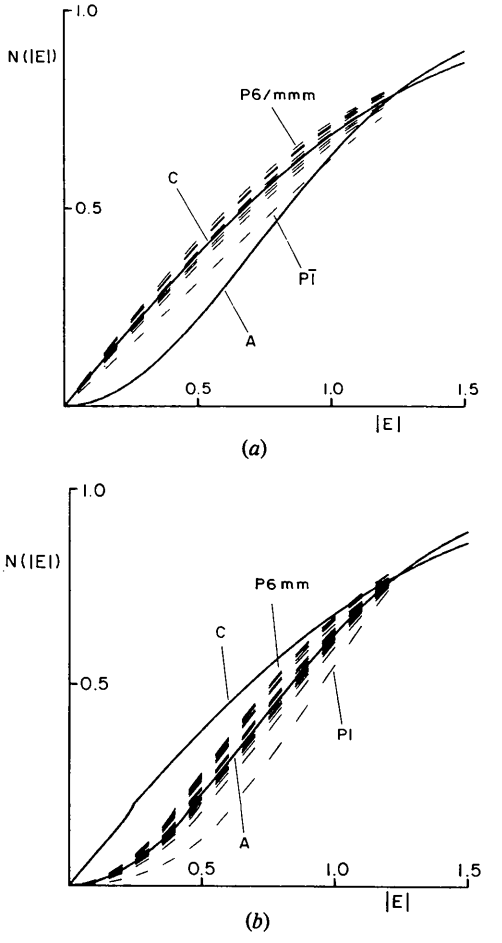

Fig. 2. Generalized cumulative distribution functions of $|E|$, for all the space groups given in Table 2. The $N(|E|)$ functions were evaluated as described in the text for (*a*) centrosymmetric and (*b*) non-centrosymmetric space groups. The full curves denoted by C and A correspond to the limiting cumulative distribution functions for centrosymmetric and non-centrosymmetric space groups respectively.

consistent with those of the higher moments of $|E|$, discussed above, regarding the applicability of the limiting distributions to the resolution of possible space-group ambiguities.

Two limitations of the above illustrations must be pointed out. Firstly, a fixed number of light atoms and one heavy atom were chosen for a convenient and consistent illustration of the effect. Although this is a rather typical example, it cannot be taken as a generally representative one. Furthermore, the ratio $\rho = f_X/f_L$ was assumed to be independent of the scattering angle, for the construction of the $N(|E|)$ distributions, and was set equal to the ratio of the corresponding atomic numbers. This is a convenient reference for comparison but also a rather poor approximation, resulting in a considerable underestimation of the effect. Hence, the above illustrations should be regarded as being of a primarily qualitative nature. Both limitations are clearly non-existent in any practical application of the present results, as the atomic composition and angular dependence of the S_{2n}/S_2^n sums are then properly evaluated for the crystal under consideration.

Discussion

The approach adopted above, whereby a non-ideal probability distribution is explicitly represented as the product of (i) the ideal distribution to which it approximates, and (ii) a series of polynomials orthogonal with respect to the ideal distribution, has advantages of clarity and conciseness. Here it is applied only to the acentric and centric distributions, but it could obviously be used for deviations from the hypercentric distribution (Lipson & Woolfson, 1952; Wilson, 1952; Rogers & Wilson, 1953); arising from additional non-crystallographic symmetry; the sesquicentric distributions (Wilson, 1956), and the subcentric distribution, arising from partial symmetry (Parthasarathy, 1966; Srinivasan & Parthasarathy, 1976, chapter 3) or dispersion (Wilson, 1980). The orthogonal polynomials required for these distributions are less familiar than the Laguerre and Hermite, and it will be interesting to see if there is any analogue to the formal equality of the coefficients of successive terms of the series, see (22)–(27). We hope to treat the subcentric case in a later paper.

The most straightforward application of the present results, moments as well as distributions of the normalized structure amplitudes, is their use in the verification of space-group assignments. Existing computer programs already evaluate $\langle |E|^n \rangle$ and S_n (or their equivalents) as a function of $\sin\theta/\lambda$, so incorporating the present results into current routines should present no difficulties.

Many of the hitherto proposed direct methods for phase determination depend explicitly or implicitly on probability distribution functions of the normalized structure amplitude(s). Of particular interest are the joint probability distributions of several structure factors which, like the one-dimensional distributions discussed here, depend on the space group (*e.g.* Naya, Nitta & Oda, 1964) and on atomic heterogeneity (*e.g.* Karle & Gilardi, 1973; Giacovazzo, 1976), and are presented as Gram–Charlier or Edgeworth type expansions (Cramér, 1945). It is hoped that the approach adopted in this paper will be of some use in further simplifications of these, still rather cumbersome, joint distribution functions and promote their applicability to structure determination.

Finally, we wish to comment on the advantage of the Edgeworth arrangement over the Gram–Charlier arrangement of series like (1), (5), (6) and (10). The distinction between the two arrangements has already been mentioned in passing. In the Gram–Charlier form the *polynomials*, say p_n, occur in what appears to be the logical order of increasing n, whereas in the Edgeworth form the terms are rearranged so that the *coefficients* of the polynomials are ordered in accordance with decreasing powers of some parameter N – in the present application it would be the number of atoms in the asymmetric unit. Obviously S_n, as defined by (18), is proportional to N, and the quantities a, c, g, L, ..., U, depend only on the space group, not on N. The coefficients A_4 and B_4 [(25)] are thus of the order of N^{-1}, and A_6 and B_6 [(26)] are of the order of N^{-2}. One would therefore expect that A_8 and B_8 [(27)] would be of the order of N^{-3}, but inspection shows that this is true only of one of the two components:

$$A_8'' = B_8'' = gS_8/S_2^4 \tag{31}$$

is indeed of the order of N^{-3}, but

$$A_8' = B_8' = Ua^2 S_4^2/S_2^4 \tag{32}$$

is of the order of N^{-2}, like A_6 and B_6, and it is logical to include it if series correct to terms in N^{-2} are desired. Equation (1) would thus be better written in the Edgeworth form:

$$P_c(z) = (2\pi z)^{-1/2} \exp\left(-\frac{z}{2}\right)\{1 + A_4 H_4/2^2 4! + [A_6 H_6/2^3 6! + A_8' H_8/2^4 8!] + A_8'' H_8/2^4 8! + \ldots\}, \tag{33}$$

and similarly for (5), (6) and (10), the argument of the polynomials being $(\frac{1}{2}z)^{1/2}$ as before. The point is of practical as well as of theoretical importance. The coefficient A_8' is available for all space groups, the necessary values of a being available from the work of Wilson (1978) and Shmueli & Kaldor (1981), and it may often be more important than the coefficient A_8'', available for comparatively few space groups as yet. For sufficiently complex structures the additional factor

of N^{-2} is bound to make A_8'' unimportant, but this is not necessarily true for simpler structures.

We are indebted to Professor E. F. Bertaut for a helpful conversation and provision of a duplicated table of linearized powers of structure factors for certain space groups; these were used for some preliminary calculations of r before the computer programs for the evaluation of this quantity were developed.

APPENDIX *A*

The calculation of the moments of z in terms of the expansion coefficients A_k and B_n, appearing in (1) and (6), can be readily done with the aid of the definite integrals

$$\int_{-\infty}^{\infty} \exp(-x^2)x^m H_n(x)\mathrm{d}x = \begin{cases} 0, & n > m \text{ or } n-m = \text{odd} \\ \dfrac{m!\sqrt{\pi}}{4^p p!}, & m \geq n, p = \dfrac{m-n}{2} \end{cases} \quad (A1)$$

and

$$\int_0^{\infty} \exp(-x)x^m L_n(x)\mathrm{d}x = \begin{cases} 0, & n > m \\ \dfrac{(-1)^n (m!)^2}{n!(m-n)!}, & m \geq n \end{cases} \quad (A2)$$

(Gröbner & Hofreiter, 1966).

The results can be readily checked by substituting the appropriate Rodrigues's formulae

$$H_n(x) = (-1)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n}(e^{-x^2}),$$

$$L_n(x) = \frac{1}{n!} e^x \frac{\mathrm{d}^n}{\mathrm{d}x^n}(x^n e^{-x})$$

into (*A*1) and (*A*2), and integrating by parts.

The integrals leading to the cumulative distributions (5) and (10) are:

$$\int_0^x \exp(-t^2) H_n(t)\mathrm{d}t = [H_{n-1}(0) - H_{n-1}(x)] \exp(-x^2) \quad (A3)$$

and

$$\int_0^x \exp(-t)L_n(t)\mathrm{d}t = [L_{n-1}(x) - L_n(x)] \exp(-x) \quad (A4)$$

(Abramowitz & Stegun, 1972).

Note that the result given in (*A*2) does not contain the $1/n!$ factor in the reference quoted, because of a different definition of $L_n(x)$. Also, (*A*4) follows from the integral $\int_x^{\infty} \exp(-t)L_n(t)\mathrm{d}t$ given by Abramowitz & Stegun (1972) when we note that $\int_0^{\infty} \exp(-t)L_n(t)\,\mathrm{d}t = 0$ for $n \neq 0$.

APPENDIX *B*

Derivation of generalized moments of $|F|$

As pointed out in the text, the derivation of the required moments is based on Wilson's (1978) derivation of the generalized fourth moment of $|F|$. For the sake of completeness and also because of some differences in notation, Wilson's derivation will be outlined here.

The structure factor is given by

$$F(\mathbf{h}) = \sum_a f_a J_a \quad (B1)$$

with

$$J_a = \sum_s \exp[2\pi i \mathbf{h}^T (P_s \mathbf{r}_a + \mathbf{t}_s)]. \quad (B2)$$

The summation in (*B*1) extends over all the atoms in the asymmetric unit, while that in (*B*2) ranges over all the equivalent positions of the Wyckoff position in which the ath atom is located; f_a is the atomic scattering factor, $\mathbf{h}$ is the diffraction vector hkl, and $(P_s|\mathbf{t}_s)$ is a space-group operation corresponding to the sth equivalent position in the set of positions occupied by the ath atom.

It can be verified that the average $\langle J_a J_a^* \rangle = \langle |J_a|^2 \rangle$, taken over a large set of hkl values, equals the multiplicity of the Wyckoff position which we denote by p_a (Wilson, 1978). Furthermore, the averages $\langle J_a J_b \rangle$ and $\langle J_a J_b^* \rangle$ with $a \neq b$, vanish for centrosymmetric as well as for non-centrosymmetric space groups.

We also have (Wilson, 1978)

$$\langle J_a J_a \rangle = \begin{cases} p_a & \text{if centrosymmetric} \\ 0 & \text{if non-centrosymmetric} \end{cases} \quad (B3)$$

since $J_a = J_a^*$ for centrosymmetric space groups. As will be seen below, (*B*3) and its higher analogs are the only sources of difference between the moments obtained for the centrosymmetric and non-centrosymmetric cases.

In what follows the first four even moments of the trigonometric structure factor will be denoted by

$$p = \langle |J|^2 \rangle, \quad q = \langle |J|^4 \rangle, \quad r = \langle |J|^6 \rangle, \quad s = \langle |J|^8 \rangle. \quad (B4)$$

Other abbreviations to be used are

$$w_a = f_a J_a, \quad \alpha_a = \langle |w_a|^2 \rangle, \quad \beta_a = \langle |w_a|^4 \rangle,$$
$$\gamma_a = \langle |w_a|^6 \rangle, \quad \delta_a = \langle |w_a|^8 \rangle. \quad (B5)$$

In view of the above, the average (corrected) intensity is given by

$$\langle |F|^2 \rangle = \sum_a \sum_b \langle w_a w_b^* \rangle$$

$$= \sum_a \sum_b f_a f_b^* \langle J_a J_b^* \rangle$$

$$= \sum_a \alpha_a$$

$$= \sum_a |f_a|^2 \langle |J_a|^2 \rangle$$

$$= \sum_a f_a^2 p_a. \qquad (B6)$$

The last line of (*B*6) was obtained by neglecting dispersion.

The required higher moments of $|F|$ can thus be written as

$$\langle |F|^4 \rangle = \sum_a \sum_b \sum_c \sum_d \langle w_a w_b^* w_c w_d^* \rangle, \qquad (B7)$$

$$\langle |F|^6 \rangle = \sum_a \sum_b \sum_c \sum_d \sum_e \sum_f \langle w_a w_b^* w_c w_d^* w_e w_f^* \rangle, \qquad (B8)$$

$$\langle |F|^8 \rangle = \sum_a \sum_b \sum_c \sum_d \sum_e \sum_f \sum_g \sum_h \langle w_a w_b^* w_c w_d^* w_e w_f^* w_g w_h^* \rangle. \qquad (B9)$$

A surviving term in any of the above summations can be either an even moment of w [β, γ or δ from (*B*5)] or a product of even moments only. Considering all the possible combinations of even moments of w, we can, using (*B*5), rewrite (*B*7) – (*B*9) as follows:

$$\langle |F|^4 \rangle = L \sum_{a \neq c} \sum \alpha_a \alpha_c + \sum_a \beta_a, \qquad (B10)$$

$$\langle |F|^6 \rangle = M \sum_{a \neq c \neq e} \sum \sum \alpha_a \alpha_c \alpha_e + N \sum_{a \neq c} \sum \alpha_a \beta_c + \sum_a \gamma_a, \qquad (B11)$$

$$\langle |F|^8 \rangle = R \sum_{a \neq c \neq e \neq g} \sum \sum \sum \alpha_a \alpha_c \alpha_e \alpha_g + S \sum_{a \neq c \neq e} \sum \sum \alpha_a \alpha_c \beta_c + T \sum_{a \neq c} \sum \alpha_a \gamma_c + U \sum_{a \neq e} \sum \beta_a \beta_e + \sum_a \delta_a, \qquad (B12)$$

with the understanding that no pair of equal subscripts appears in any of the summations in (*B*11) and (*B*12). Thus, $a \neq c \neq e$ includes the inequality $a \neq e$ *etc.* The integers L, M, N, R, S, T and U denote the multiplicities of the various double, triple and quadruple summations which appear in (*B*10)–(*B*12). These are, of course, different for centrosymmetric and non-centrosymmetric crystals and their evaluation can be performed by considering all the permissible combinations of indices, as will be shown below.

Equations (*B*10)–(*B*12) can be brought to a manageable form by converting the multiple summations to combinations of products of single summations, through addition and subtraction of equal terms. Thus, for example, a double summation in (*B*11) can be written as

$$\sum_{a \neq c} \sum \alpha_a \beta_c = \left(\sum_a \alpha_a \right) \left(\sum_c \beta_c \right) - \sum_a \alpha_a \beta_a. \qquad (B13)$$

More care is needed for higher-order summations. Thus, for example,

$$\sum_{a \neq c \neq e} \sum \sum \alpha_a \alpha_c \alpha_e = \left(\sum_a \alpha_a \right)^3 - \sum_a \alpha_a^3 - 3 \sum_{a, \neq c} \sum \alpha_a \alpha_c^2. \qquad (B14)$$

The last summation in (*B*14) results from those combinations of indices for which $a = c \neq e$ or $a \neq c = e$ or $a = e \neq c$ and hence the multiplier 3. This double summation is further simplified as in (*B*13). Similarly, a fourfold summation in (*B*12) can be decomposed as

$$\sum_{a \neq c \neq e \neq g} \sum \sum \sum \alpha_a \alpha_c \alpha_e \alpha_g = \left(\sum_a \alpha_a \right)^4 - 6 \sum_{a \neq c \neq e} \sum \sum \alpha_a \alpha_c \alpha_e^2 - 4 \sum_{a \neq c} \sum \alpha_a \alpha_c^3 - 3 \sum_{a \neq e} \sum \alpha_e^2 \alpha_e^2 - \sum_a \alpha_a^4 \qquad (B15)$$

and upon a repeated decomposition of the triple and double summations in (*B*15) we have

$$\sum_{a \neq c \neq e \neq g} \sum \sum \sum \alpha_a \alpha_c \alpha_e \alpha_g = \left(\sum_a \alpha_a \right)^4 - 6 \left(\sum_a \alpha_e \right)^2 \left(\sum_e \alpha_e^2 \right) + 8 \left(\sum_a \alpha_a \right) \left(\sum_c \alpha_c^3 \right) + 3 \left(\sum_a \alpha_a^2 \right)^2 - 6 \sum_a \alpha_a^4. \qquad (B16)$$

Converting all the summations in (*B*10)–(*B*12) to the form of (*B*13) or (*B*16), completely general expressions for the moments of $|F|$ are obtained. In what follows we shall neglect dispersion and assume that all the atoms are located in general positions. These general expressions can now be simplified as shown below. For example,

$$\sum_a \alpha_a \beta_a = \sum_a \langle |w_a|^2 \rangle \langle |w_a|^4 \rangle$$

$$= \sum_a \langle |f_a J_a|^2 \rangle \langle |f_a J_a|^4 \rangle$$

$$= \sum_a p_a q_a f_a^6 \quad \text{(dispersion neglected)}$$

$$= pqS_6 \quad \text{(general positions also assumed)}$$

where $S_n = \sum_a f_a^n$, and abbreviations given in (*B*4) and (*B*5) were used.

Proceeding as outlined above and substituting $\langle |F|^2 \rangle$ for $\sum_a p_a f_a^2 = pS_2$, from (*B*6), we obtain the required expressions for the moments of $|F|$.

$$\langle |F|^4 \rangle = \langle |F|^2 \rangle^2 \left\{ L + \left(\frac{q}{p^2} - L \right) \frac{S_4}{S_2^2} \right\}, \qquad (B17)$$

$$\langle |F|^6 \rangle = \langle |F|^2 \rangle^3 \left\{ M + \left(\frac{Nq}{p^2} - 3M \right) \frac{S_4}{S_2^2} + \left(\frac{r}{p^3} - \frac{Nq}{p^2} + 2M \right) \frac{S_6}{S_2^3} \right\}, \qquad (B18)$$

$$\langle |F|^8 \rangle = \langle |F|^2 \rangle^4 \left\{ R + \left(\frac{Sq}{p^2} - 6R \right) \frac{S_4}{S_2^2} + \left(\frac{Tr}{p^3} - \frac{2Sq}{p^2} + 8R \right) \frac{S_6}{S_2^3} + \left(\frac{s}{p^4} - \frac{Uq^2}{p^4} - \frac{Tr}{p^3} + \frac{2Sq}{p^2} - 6R \right) \frac{S_8}{S_2^4} + \left(\frac{Uq^2}{p^4} - \frac{Sq}{p^2} + 3R \right) \frac{S_4^2}{S_2^4} \right\}. \qquad (B19)$$

The expressions in the curled parentheses on the r.h.s of (*B*17)–(*B*19) are just the corresponding moments of the normalized structure amplitude $|E|$ discussed in the paper.

A detailed derivation of $\langle |F|^4 \rangle$, including the effects of dispersion on the moment, is given by Wilson (1978).

The constants L, M, N, R, S, T and U, appearing in (*B*17)–(*B*19), can be evaluated from simple considerations which differ for the centrosymmetric and non-centrosymmetric cases according to (*B*3). Thus, any even-numbered group of indices can be contracted to give a non-vanishing average in the centrosymmetric case while in the non-centrosymmetric case, each group of indices to be so contracted must contain equal numbers of indices from the sets a, c and b, d in (*B*7), a, c, e and b, d, f in (*B*8) and a, c, e, g and b, d, f, h in (*B*9). Clearly, only ordered groups of different indices are non-equivalent for the purpose of contraction.

An inspection of (*B*7)–(*B*9) shows that the number of different combinations of n pairs of indices (out of $2n$ indices) can be found as follows. In the centrosymmetric case, keeping one index of the first pair fixed, this pair can be chosen in $2n - 1$ ways, the second pair in $2n - 3$ ways, *etc.* Hence, this number equals $(2n - 1)!! = (2n - 1) \times (2n - 3) \times \ldots \times 1$ for this case. In the non-centrosymmetric case, the first pair can be chosen in n ways only, the second pair in $n - 1$ ways, *etc.* Hence the number of different combinations of pairs is $n!$ in the latter case. The constants L, M and R are thereby defined.

It can be seen from (*B*11) that N must equal the number of ordered pairs that can be formed from six indices and this number is 15 or 9 for centrosymmetric or non-centrosymmetric cases respectively. For eight indices, the number (T) of different ordered pairs (or sextets) is 28 or 16 and that of different ordered quartets is 70 or 36 for centrosymmetric or non-centrosymmetric space groups respectively. In order to find the values of S we observe that once a quartet is chosen, the remaining two pairs can be chosen in L ways, and the values of U follow from the fact that the number of different pairs of quartets must be one half the number of quartets. The various multiplicities are summarized below.

L	M	N	R	S	T	U	
3	15	15	105	210	28	35	centrosymmetric
2	6	9	24	72	16	18	non-centrosymmetric.

Several simplifications emerge from the above. An inspection of (*B*17)–(*B*19) leads to the following relationships:

$$\frac{Nq}{p^2} - 3M = N\left(\frac{q}{p^2} - L \right), \qquad (B20)$$

$$\frac{Sq}{p^2} - 6R = S\left(\frac{q}{p^2} - L \right), \qquad (B21)$$

$$\frac{Uq^2}{p^4} - \frac{Sq}{p^2} + 3R = U\left(\frac{q}{p^2} - L \right)^2, \qquad (B22)$$

$$\frac{Tr}{p^3} - \frac{2Sq}{p^2} + 8R = T\left(\frac{r}{p^3} - \frac{Nq}{p^2} + 2M \right), \qquad (B23)$$

which usefully simplify the representation of the moments (see text). Other simplifications, related to the expansion coefficients A_k and B_k [(1) and (6)], are treated in the paper.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover.

BERTAUT, E. F. (1955). *Acta Cryst.* **8**, 823–832.

CRAMÉR, H. (1928). *Skand. Aktuarietidskr.* **6**, 13–74, 141–180.

CRAMÉR, H. (1945). *Mathematical Methods of Statistics*. Uppsala: Almqvist and Wiksells.

CRAMÉR, H. (1970). *Random Variables and Probability Distributions*, 3rd ed. Cambridge Univ. Press.

FOSTER, F. & HARGREAVES, A. (1963*a*). *Acta Cryst.* **16**, 1124–1133.

FOSTER, F. & HARGREAVES, A. (1963*b*). *Acta Cryst.* **16**, 1133–1139.

GIACOVAZZO, C. (1976). *Acta Cryst.* A**32**, 74–82.
GRÖBNER, W. & HOFREITER, N. (1966). *Integraltafel* 2. *Bestimmte Integrale.* New York: Springer.
HAUPTMANN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136–141.
HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210–214.
KARLE, J. & GILARDI, R. D. (1973). *Acta Cryst.* A**29**, 401–407.
KARLE, J. & HAUPTMANN, H. (1953). *Acta Cryst.* **6**, 131–135.
KLUG, A. (1958). *Acta Cryst.* **8**, 515–543.
LIPSON, H. & WOOLFSON, M. M. (1952). *Acta Cryst.* **5**, 680–682.
MAIN, P., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J. P. & WOOLFSON, M. M. (1978). *A System of Computer Programs for Solution of Crystal Structures by Direct Methods.* Univ. of York, England.
MITRA, G. B. & BELGAUMKAR, J. (1973). *Proc. Indian Natl Sci. Acad. Part A*, **39**, 95–100.
NAYA, S., NITTA, I. & ODA, T. (1964). *Acta Cryst.* **17**, 421–433.
PARTHASARATHY, S. (1966). *Z. Kristallogr.* **123**, 27–50.
ROGERS, D. & WILSON, A. J. C. (1953). *Acta Cryst.* **6**, 439–449.
SHMUELI, U. (1979). *Acta Cryst.* A**35**, 282–286.
SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* A**37**, 76–80.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.
WILSON, A. J. C. (1949). **2**, 318–321.
WILSON, A. J. C. (1952). *Research (London)*, **5**, 589–590.
WILSON, A. J. C. (1956). *Acta Cryst.* **9**, 143–144.
WILSON, A. J. C. (1978). *Acta Cryst.* A**34**, 986–994.
WILSON, A. J. C. (1980). *Acta Cryst.* A**36**, 945–946.

Acta Cryst. (1981). A**37**, 353–360

The Behaviour of Residuals with Respect to the Elimination of Low-Intensity Data

BY G. H. PETIT, A. T. H. LENSTRA AND J. F. VAN LOOCK

Department of Chemistry, University of Antwerp (UIA), Universiteitsplein 1, B-2610 Wilrijk, Belgium

(*Received* 14 *July* 1979; *accepted* 19 *November* 1980)

Abstract

Theoretical expressions for R_2 and R_2^N, indicator functions used in automated structure analysis, are given for space groups $P1$ and $P\bar{1}$ as a function of $\sigma_1^2 = \sum_{j=1}^{P} f_j^2 / \sum_{j=1}^{N} f_j^2$ (P is the number of atoms in the tentative structural model) and of a threshold a on the intensity data. It appears that R_2 is only slightly dependent, R_2^N, however, is strongly dependent on a. If low-intensity reflections are eliminated from the data set, R_2, and not R_2^N, thus remains a useful control function in testing the reliability of atomic positions Theory and experiment are shown to be in good agreement.

1. Introduction

In automated structure analysis, one obviously needs criteria to discriminate between correct and incorrect structure models. As such a criterion we prefer the mathematical indicator function R_2 above chemical criteria. R_2 is defined as

$$R_2 = \frac{\sum_H (I_N - I_P)^2}{\sum_H I_N^2}, \qquad (1)$$

where I_N represents the observed intensity of the N-atom structure looked for and I_P is the calculated intensity of a proposed structure model containing P atoms.

The functional behaviour of R_2 is evaluated theoretically in many papers (Wilson, 1969, 1974, 1976; Lenstra, 1973, 1974, 1975, 1979; Parthasarathy & Parthasarathi, 1972; Parthasarathy, 1975; Van de Mieroop & Lenstra, 1978). Applying R_2 in realistic structure determinations, Van de Mieroop (1979) obtained encouraging results. In the procedure he tried, the structure is determined by adding the atoms one by one to the model. Each time the reliability of the enlarged model is checked, comparing its experimental R_2 value with the expected one. Clearly, the reliability check requires substantial amounts of computer time.

An acceleration of the reliability check was sought in a reduction of the numbers of reflections used to calculate R_2. Following the practice with rotation functions (Tollin & Rossmann, 1966), we decided to eliminate intensities for which $E^2 < a$ and to investigate the consequences on the path of R_2.

In this paper we report the theoretical derivation of R_2 as a function of the threshold a for the space groups $P1$ and $P\bar{1}$ and compare the results with corresponding experimental values. Parthasarathi & Parthasarathy (1975) have stated that the normalized residual R_2^N

0567-7394/81/030353-08$01.00