crystallites. The second reason is that it is unlikely that the disordered molecules in columns are piled up regularly to give definite molecular images. The image shown in Fig. 11 is obtained with a film of Pt-Phc whose thickness is two molecular layers. The molecular columns are ordered in this film. The well arranged crystal parts increase when the film thickness increases. In these cases, disorder is observed as a bend of the column, which results in the deterioration of image quality as can be seen in Fig. 9, where the molecular columns are projected along the column axis. The disordered states displayed in Fig. 9 and Fig. 10 can be considered to result from the lack of kinetic energy for surface diffusion of the molecule necessary for crystallization on the substrate, because the deposition was carried out at room temperature, at which the phthalocyanine molecule does not have sufficient energy for surface diffusion. The process of crystal growth of phthalocyanine may be considered as follows. The cohesive energy of phthalocyanine molecules due to π-electronic interaction is anisotropic and acts more strongly to build a molecular column. When the column grows to a certain extent, the attractive force between columns becomes strong and they condense side by side and make a crystal, as illustrated in Fig. 12. The contrast differences in each molecular image in Fig. 9 suggest that the length of the column is not constant. The images in Figs. 9 and 10 are obvious proofs for the concept that formation of linear crystal, molecular column, precedes the formation of three-dimensional crystals.

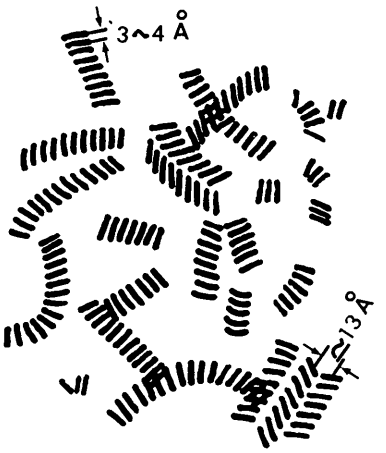

Fig. 12. Schematic representation of the initial state of the crystal formation.

References

BROWN, C. J. (1968). *J. Chem. Soc. A*, pp. 2488–2493, 2494–2498.

FRYER, J. R. (1977). *Inst. Phys. Conf. Ser.* No. 36, pp. 423–424.

FUJIYOSHI, Y., KOBAYASHI, T., ISHIZUKA, K., UYEDA, N., ISHIDA, Y. & HARADA, Y. (1980). *Ultramicroscopy*, **5**, 459–468.

KOBAYASHI, T., FUJIYOSHI, Y., IWATSU, F. & UYEDA, N. (1981). *Acta Cryst.* A**37**, 692–697.

WILLIAMS, J., THOMAS, J. M., WILLIAMS, J. O. & HOBBS, L. W. (1975). *J. Chem. Soc. Faraday Trans.* II, **71**, 138–145.

Acta Cryst. (1982). A**38**, 362–371

A Study of Generalized Intensity Statistics: Extension of the Theory and Practical Examples

BY URI SHMUELI

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

(*Received* 27 *August* 1981; *accepted* 22 *December* 1981)

Abstract

Generalized probability density functions, cumulative distribution functions and moments of the normalized structure amplitude $|E|$, depending on space-group symmetry of the crystal and on the composition of the asymmetric unit, were extended to include the tenth moment of $|E|$ and five-term expansions. The formalism was also simplified and is presented in a concise and unified form. The equations linking the formalism to practical problems, the composition and space-group terms, are discussed from a practical point of view and a convenient implementation of the above statistics in a computer program is indicated. The generalized cumulative distributions of $|E|$ and of the normalized intensity $z = |E|^2$ are compared with corresponding distributions based on five published structures, each containing one outstandingly heavy atom (Pt, Rh and

0567-7394/82/030362-10$01.00

Br) and about twenty light ones in the asymmetric unit, excluding hydrogens. These examples indicate that the above formalism is a valuable tool for resolving space-group ambiguities which cannot be treated by conventional methods because of effects of atomic heterogeneity. $N(z)$ distributions for a structure belonging to the space group *Fddd* show that the theoretical expressions correctly predict the existence of different intensity distributions in reflection subsets with *hkl* all even and *hkl* all odd for this space group.

Introduction

Intensity statistics which allow for an arbitrary composition of the asymmetric unit and, *formally*, any space-group symmetry of the crystal have been extensively investigated in the 1953–1976 period, starting with the studies of Karle & Hauptman (1953) and Hauptman & Karle (1953). A review of these and subsequent investigations of such statistics was given by Srinivasan & Parthasarathy (1976).

The well established mathematical formalisms, on one hand, and the surprising scarcity of their applications, on the other, indicated that these powerful methods could well be exploited, provided (i) the formalisms existing at that time (1976) were made more concise and simpler and (ii) a general treatment of the space-group dependence of these statistics, or results of such a treatment, were available.

An answer to the second requirement was first given by Wilson (1978), who introduced a new approach to the statistics of the trigonometric structure factor and evaluated the fourth absolute moment of this quantity for all the space groups but two, and an attempt at reducing the formalisms to readily applicable forms, albeit for triclinic, monoclinic and orthorhombic symmetries only, was made by the author (Shmueli, 1979) and three-term expansions for generalized cumulative distributions of the normalized structure amplitude $|E|$ were put forward.

The above two contributions were unified and extended to include general expressions for the first four moments of the normalized intensity, $z = |E|^2$, and four-term expansions for the probability density and cumulative distribution functions of z (Shmueli & Wilson, 1981). A considerable conciseness of the various expressions was achieved in the above study, and a unified character of the quantities pertaining to centrosymmetric and noncentrosymmetric space groups was first demonstrated (Shmueli & Wilson, 1981). This reformulation of intensity statistics was usefully supplemented by computer algorithms for the evaluation of absolute even moments of the trigonometric structure factor, for all space groups (Shmueli & Kaldor, 1981). Such algorithms are indispensable for obtaining moments higher than the fourth, for space groups of higher symmetry.

During subsequent tests of these statistics it became apparent that cases of extreme atomic heterogeneity may well require more extensive expansions and the results of Shmueli & Wilson (1981) were extended to the fifth moment of z and five-term expansions for the probability and cumulative distribution functions of this quantity. A reexamination of the extended formalism showed that it could be further unified, and thereby simplified, and the resulting version of these statistics is presented in the next section, along with a discussion which also includes programming considerations. All the mathematical and statistical considerations and derivations, which led to the above extension, are given in the Appendices to this paper.

The article is concluded with an application of these cumulative distribution functions to five sets of $h,k,l,|F_c|$ data which were recalculated from solved crystal structures, each containing just one heavy atom (Pt, Rh or Br) and a relatively small number of light ones in the asymmetric unit. Such compositions occur frequently in inorganic and organometallic chemistry, and are well known to invalidate the conclusions of conventional statistical tests, based on the central limit theorem, in a variety of space groups.

One of the examples concerns a structure belonging to the space group *Fddd* which is predicted to give rise to different intensity distributions in the *hkl*-even and *hkl*-odd subsets (Wilson, 1978; Shmueli & Kaldor, 1981), and these predictions are now put to a practical test.

Generalized distributions and moments

In what follows, a brief but complete summary of the equations of symmetry- and composition-dependent intensity statistics will be presented. The basic approach is the same as that described in detail by Shmueli & Wilson (1981) and the present extension and simplified notation are based on Appendix *A* and Appendix *B* to this paper, which deal with a unified representation of the general forms of the probability density functions of $|E|$ and with the derivation of the tenth moment of this quantity in terms of functions of composition and space-group symmetry, respectively.

The formal validity of these generalized statistics is restricted by the assumptions that (i) all the atoms occupy general positions, (ii) there is no pseudo-symmetry or other dependence in the structure, and (iii) dispersion is negligible. Similarly to any other statistical treatment, a large set of data is desirable in the present one.

After the presentation of the formalism, its links to the atomic composition of the asymmetric unit and space-group symmetry of the crystal will be discussed from a practical point of view and the section will be concluded with a programming note on the implementation of the above in an easy-to-operate routine for

computation of generalized moments and cumulative distributions of the normalized intensity, z, or normalized structure amplitude, $|E|$.

The centrosymmetric case

The expressions for the generalized probability density and cumulative distribution functions of $|E|$ are given by

$$P_c(|E|) = \left(\frac{2}{\pi}\right)^{1/2} \exp\left(-\frac{|E|^2}{2}\right) \times \left[1 + \sum_{k=2}^{n} \frac{A_{2k}}{2^k(2k)!} H_{2k}\left(\frac{|E|}{\sqrt{2}}\right) + \ldots\right] \quad (1)$$

and

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

and the even moments of $|E|$ are related to the expansion coefficients by

$$\langle |E|^{2k} \rangle = \alpha_{k0} + \sum_{p=2}^{k} \alpha_{kp} A_{2p}, \quad (3)^*$$

where

$$\alpha_{kp} = \binom{k}{p} \frac{(2k-1)!!}{(2p-1)!!} \quad (4)$$

with $(2p-1)!! = (2p)/(2^p p!)$.

Equations (1), (2) and (3) are completely general, but the number of terms in (1) and (2) which can be computed depends on the orders of the even moments of $|E|$ which have been related to the symmetry and composition of the crystal. At present, the fourth, sixth, eighth and tenth moments of $|E|$ are available and, hence, at most the first five terms of either (1) or (2) can be computed. *I.e.* $n = 5$ in the present paper.

By comparing the available moments of $|E|$ (Shmueli & Wilson, 1981; Appendix *B* below) with (3), we obtain

$$A_4 = a_4 Q_4, \quad (5)$$

$$A_6 = a_6 Q_6, \quad (6)$$

$$A_8 = a_8 Q_8 + 35(A_4^2 - \gamma_4^2 Q_8), \quad (7)$$

$$A_{10} = a_{10} Q_{10} + 210(A_4 A_6 - \gamma_4 \gamma_6 Q_{10}) + 3150\gamma_4^2 Q_{10} \quad (8)$$

* Equation (3) and its acentric analogue, are obtained by evaluating $\langle |E|^{2k} \rangle = \int_0^\infty |E|^{2k} P(|E|)\, d|E|$, where $P(|E|)$ is given by (1) and (13) respectively.

with

$$a_{2k} = \sum_{p=2}^{k} (-1)^{k-p}(k-p)!\, \alpha_{kp}\, \gamma_{2p} + (-1)^{k-1}(k-1)!\, \alpha_{k0} \quad (9)$$

$$Q_{2k} = \sum_{j=1}^{m} f_j^{2k} \Big/ \left(\sum_{j=1}^{m} f_j^2\right)^k, \quad (10)$$

where m is the number of atoms in the asymmetric unit and $f_j, j = 1, \ldots, m$ are their scattering factors, and

$$\gamma_{2k} = \langle |J|^{2k} \rangle / \langle |J|^2 \rangle^k, \quad (11)$$

where

$$J(\mathbf{h}) = \sum_s \exp[2\pi i \mathbf{h}^T (P_s\, \mathbf{r} + \mathbf{t}_s)] \quad (12)$$

is the trigonometric structure factor (Wilson, 1978; Shmueli & Wilson, 1981; Shmueli & Kaldor, 1981).

Equations (1) and (2) with $n = 4$, except for a change of the variable from z to $|E|$, are equivalent to equations (1) and (5) of Shmueli & Wilson (1981). The main features of the present extension are the derivation of the tenth moment of $|E|$ which leads to (8) and, consequently, to another term in each of (1) and (2), and the introduction of the α_{kp} constants (equation 4) which considerably simplify the formalism.

The noncentrosymmetric case

The expressions for the generalized probability density and cumulative distribution functions of $|E|$ are given for this case by

$$P_a(|E|) = 2|E| \exp(-|E|^2) \times \left[1 + \sum_{k=2}^{n} \frac{(-1)^k B_{2k}}{k!} L_k(|E|^2) + \ldots\right] \quad (13)$$

and

$$N_a(|E|) = 1 - \exp(-|E|^2) + \exp(-|E|^2) \times \left\{\sum_{k=2}^{n} \frac{(-1)^k B_{2k}}{k!} [L_{k-1}(|E|^2) - L_k(|E|^2)] + \ldots\right\}, \quad (14)$$

respectively, and are equivalent to equations (6) and (10) of Shmueli & Wilson (1981) for $n = 4$.

Equations (3)–(12) apply to the noncentrosymmetric case subject to the following replacements: the coefficients B_{2k} replace A_{2k} in (3), (5), (6), (7) and (8), the constants α_{kp} appearing in (3) and (9) are redefined as

$$\alpha_{kp} = \binom{k}{p} \frac{k!}{p!} \quad (15)$$

for the noncentrosymmetric case, and the constants 35, 210 and 3150 appearing in (7) and (8) are replaced by 18, 100 and 900 respectively.

The above equations for probability density and cumulative distribution functions can be reexpressed in terms of the normalized intensity z by making use of the relation

$$P(|E|) = 2|E|P(z) \tag{16}$$

(*e.g.* Shmueli & Wilson, 1981) and noting that $z = |E|^2$. In order to obtain the corresponding expressions for $N(z)$, it is enough to replace $|E|$ by $\sqrt{z}$, throughout the right hand side of (2) and (14).

The Hermite and Laguerre polynomials appearing in (1), (2) and (13), (14) respectively are defined and tabulated in mathematical handbooks (*e.g.* Abramowitz & Stegun, 1972). Their explicit forms are also reconsidered in Appendix *A* which leads to the introduction of (4) and (15), resulting in the unified presentation given above. It can be noted that (4) and (15) can also be unified to

$$\alpha_{kp} = \binom{k}{p} \langle |E|^{2k} \rangle^{(0)} / \langle |E|^{2p} \rangle^{(0)} \tag{17}$$

where $\langle |E|^{2k} \rangle^{(0)}$, the ideal $2k$th moment of $|E|$, equals $(2k-1)!! = 1 \times 3 \times \ldots \times (2k-1)$ and $k!$ for the centric and acentric case respectively (*cf.* Appendix *A*).

The identical functional forms of the coefficients A_{2k} and B_{2k}, pointed out by Shmueli & Wilson (1981) for $k = 2$, 3 and 4, and confirmed in Appendix *B* for $k = 5$, are thus a general feature of these expansions, at least as far as their dependence on the even moments of $|E|$ is concerned.

In order to compute these statistics, their appropriate dependence on atomic composition and space-group symmetry must be evaluated from (10) and (11) respectively. Considering the composition terms, it must be pointed out that their dependence on $\sin\theta/\lambda$ should be accounted for when X-ray or electron diffraction data are being treated. This may seem an unnecessary precaution, since the angular dependence of Q_{2k} (10) is often rather weak (it is absent when all the atoms are the same). However, numerical tests show that, in cases of strong heterogeneity, a replacement of scattering factors by corresponding atomic numbers may lead to poor approximation.

A convenient way of evaluating the Q_{2k} term is to take its weighted average over the $\sin^2\theta/\lambda^2$ ranges, used in a previous construction of the Wilson plot, the weight of a range being related to the number of reflections this range (or shell) contains.

It was pointed out to the author by Professor Wilson (Wilson, 1980) that the neglect of dispersion, which leads to the simple form of the composition terms [equation (10); equation (18) of Shmueli & Wilson, 1981], may be less justified for electron diffraction than for X-ray data. For neutron diffraction data, (10) can be evaluated as it stands, with the f_j's replaced by the appropriate neutron scattering cross sections of the atoms.

The space-group term (11) requires the knowledge of even absolute moments of the trigonometric structure factor for the space group under consideration. In order to evaluate (1), (2), (13) and (14) with all the five terms for which moments of $|E|$ are available, the ratios γ_4, γ_6, γ_8 and γ_{10} are required. For triclinic, monoclinic and orthorhombic space groups, except *Fdd*2 and *Fddd*, the trigonometric structure factors are simple enough to give concise closed expressions from which γ_{2k}, for any k, can be evaluated.

For example, the trigonometric structure factor for space groups based on the point group *mm*2 has the form

$$J = A + iB \tag{18}$$

where

$$A = 4l' \cos 2\pi hx \cos 2\pi ky \cos 2\pi lz \tag{19}$$

$$B = 4l' \cos 2\pi hx \cos 2\pi ky \sin 2\pi lz \tag{20}$$

(*International Tables for X-ray Crystallography*, 1952) and l' takes on the values 1, 2 or 4 according as the unit cell is primitive, base or body centered or face centered respectively.

We thus have

$$\langle |J|^{2k} \rangle = (4l')^{2k} \langle \cos^{2k} 2\pi hx \cos^{2k} 2\pi ky \rangle \tag{21}$$

since $\langle |J|^{2k} \rangle = \langle (A^2 + B^2)^k \rangle$. For the case of atoms in general positions and a large set of *hkl*'s, *hx* and *ky* or, more accurately, their fractional parts are uniformly distributed in the (0,1) range and since independence was also assumed, we can replace (21) by

$$\langle |J|^{2k} \rangle = (4l')^{2k} \langle \cos^{2k} \alpha \rangle^2 \tag{22}$$

$$= (4l')^{2k} \left[\frac{(2k-1)!!}{(2k)!!} \right]^2, \tag{23}$$

where $(2k)!! = 2^k k!$ (Abramowitz & Stegun, 1972). Since the second absolute moment of $|J|$ is $4l'$ (*cf.* Wilson, 1978) and since only the fraction $1/l'$ of the reflections has non-zero J's, we have

$$\gamma_{2k} = l'^{k-1} \left[\frac{(2k-1)!!}{k!} \right]^2 \tag{24}$$

for these space groups. The γ_{2k}'s for the above mentioned space groups, derived as above, are presented in Table 1.

Direct averaging by the above or a related method becomes much too cumbersome when the number of triple sine/cosine products in A or B exceeds two, and the moments of $|J|$ can then be obtained with the aid of computer algorithms. Such methods were developed by Shmueli & Kaldor (1981) and were applied to the

Table 1. *Expressions for* γ_{2k} (14) *for the triclinic monoclinic and orthorhombic space groups except Fdd2 and Fddd*

The moment ratios γ_{2k} are expressed in terms of M_k, where

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

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

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

computation of γ_4 and γ_6 for all the space groups and all the *hkl* subsets giving rise to different functional forms of J within the same space group. Only general *hkl* subsets were treated. It should be noted that γ_4 and γ_6 are denoted by q/p^2 and r/p^3 respectively in Table 1 of Shmueli & Kaldor (1981). Similar computations of the eighth moment of $|J|$ are in progress and will be reported elsewhere.

The equations of generalized intensity statistics, presented in this section, were implemented in a Fortran program which computes the fourth and sixth moments of $|E|$ and evaluates cumulative distributions for a given composition of the asymmetric unit and two possible space groups of the crystal. Most of the input to this program can be copied directly from a routine which scales the data and computes $|E|$ values. This can be made more expedient if the program is linked, *via* a file transfer, to the scaling and normalizing routine. The local version of program *NORMAL* (*MULTAN* 78; Main, Hull, Lessinger, Germain, Declercq & Woolfson, 1978) was modified to provide for such a transfer, which also includes some relevant experimental statistics computed by *NORMAL*. The problem-specific input to the intensity statistics program *INSTAT* was thus reduced to two space-group labels and the corresponding values of γ_{2k} which are available or required. The examples presented in the next section were evaluated with the aid of the above procedure.

Applications to solved structures

The formalism summarized in the previous section was applied to several hypothetical as well as real examples. The former were based on a computer simulation of the heavy-atom problem, which afforded a possibility of examining various aspects of the theory such as dependence of the result on the number of expansion terms used, a comparison of Gram–Charlier and Edgeworth arrangements of the expansion (*cf.* Shmueli & Wilson, 1981) and more. These computer experiments were briefly mentioned elsewhere (Shmueli, 1981; Shmueli, Kaldor & Wilson, 1981) and will be published separately. In this section cumulative distributions of $|E|$ and z, recalculated from published crystal data and atomic parameters, will be presented along with the relevant theoretical distributions. For each of the examples to be given, an independent set of $h,k,l,|F_c|$ data was generated, with $\sin\theta/\lambda \leq 0{\cdot}65$, starting from published unit-cell dimensions, Laue group and the atomic parameters, positional and thermal. Such a set was then input to program *NORMAL* and was treated as a routine F_o data set for which the $|E|$ values and the experimental statistics [moments of $|E|$, $N(z)$ and $N(|E|)$ distributions] are to be computed. The relevant results were then transferred to the input of the program, mentioned at the end of the previous section and, unless otherwise stated, five-term expansions for $N(|E|)$ or $N(z)$ were evaluated using (2) and (14), and the definitions of the expansion coefficients given above. These expansions were then compared with the cumulative distributions recalculated from the published structures and with the ideal centric and acentric $N(z)$ or $N(|E|)$ (Howells, Phillips & Rogers, 1950; Srinivasan & Parthasarathy, 1976).

The main criteria for the choice of examples were (i) outstanding heterogeneity of atomic composition of the asymmetric unit (*i.e.* just one heavy atom and not too many light ones), (ii) all the atoms in general positions, and (iii) no obvious hypersymmetry in the structure. Five such examples were readily located in two or three subsequent issues of the *Journal of Inorganic Chemistry* and, in fact, many more good examples were omitted. In one case at least, criterion (iii) was not satisfied but the effect of the hypersymmetry present did not preclude a meaningful test of the theory.

These cumulative distributions are presented in Figs. 1, 2 and 3. Fig. 1 shows three examples for solved $P\bar{1}$ structures, Fig. 2 deals with a $C2/c$ *vs* Cc example ($C2/c$ being the correct space group) and Fig. 3 displays a property which is characteristic of some space groups of higher symmetry, namely different intensity distributions for two or more *hkl* subsets within the same space group. Many space groups lead to different functional forms of the trigonometric structure factor for the various *hkl* subsets, but only some of these lead to different numerical values of the moments of $|J|$ which determine the probability distribution and related statistics. The example in Fig. 3 concerns the space group *Fddd* for which different distributions of the even and odd *hkl* subsets are predicted (*cf.* Wilson, 1978; Shmueli & Kaldor, 1981).

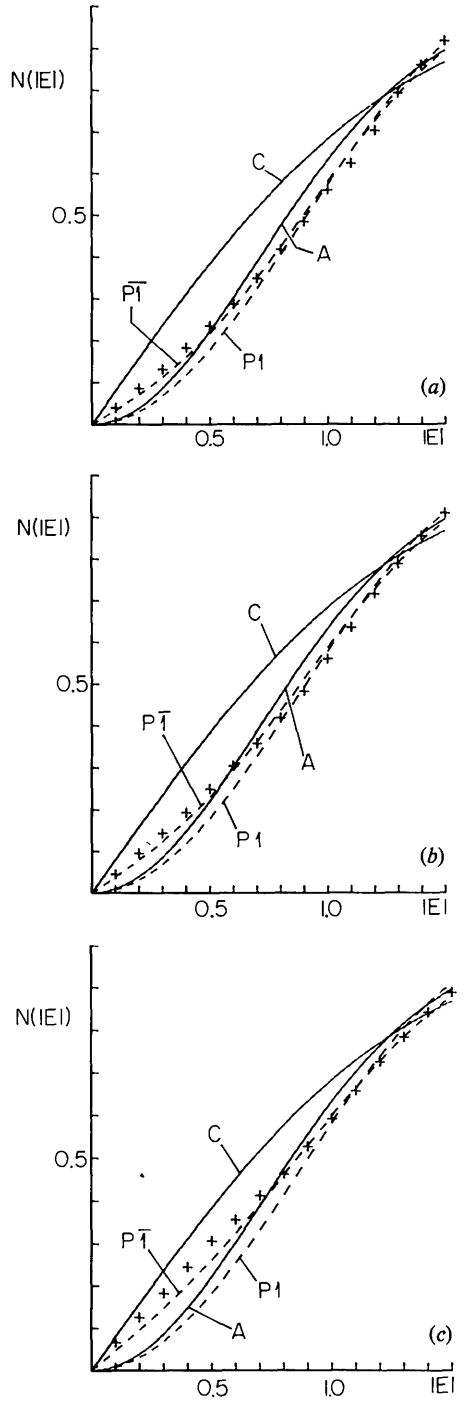

Fig. 1. $N(|E|)$ cumulative distributions for solved triclinic heterogeneous structures. The solid curves are the ideal centric (C) and acentric (A) distributions, the dashed curves are five-term expansions evaluated from equations (2) and (14) for the space group indicated and the crosses represent the $N(|E|)$ values recalculated from the published structures as described in the text. (*a*) Asymmetric unit: $C_7H_{13}ClN_4O_2Pt$, space group $P\bar{1}$ with $Z = 2$ (Faggiani, Lippert & Lock, 1980). (*b*) Asymmetric unit: $C_6H_{18}Cl_2N_4O_4Pt$, space group $P\bar{1}$ with $Z = 2$ (Faggiani *et al.*, 1980). (*c*) Asymmetric unit: $C_{10}H_{20}O_5Rh$, space group $P\bar{1}$ with $Z = 2$ (Cotton & Felthouse, 1980).

The chemical formulae of the asymmetric units and the references to the articles from which the structures were taken appear in the figure legends. The positions of hydrogen atoms were not available for any of these structures and their neglect here is probably quite unimportant. However, the H's are included in the formulae, as given in the original papers.

The $N(|E|)$ tests for the two centrosymmetric triclinic platinum complexes [Figs. 1(*a*) and (*b*)] show a good agreement between the statistics based on the published structures and the centrosymmetric $N(|E|)$ curve computed from (2). However, this agreement is of some use only in the range of $|E|$ values not exceeding 0·7, since the separation between the $P\bar{1}$ and $P1$ theoretical curves becomes too small for higher values of $|E|$. In general, the deviation of the $P1$ curve from the ideal acentric $N(|E|)$ is bound to be smaller than the deviation of the theoretical $P\bar{1}$ from the ideal centric cumulative distribution. This is because the asymmetric unit of the noncentrosymmetric space group is necessarily twice as large as that for the centrosymmetric one, when the test is to indicate the presence or absence of a center of symmetry only. Thus, the asymmetric unit of $P1$ in Figs. 1(*a*) and (*b*) contains *two* platinum atoms and twice the number of light atoms, the heterogeneity is thereby decreased and, consequently, the departure from the ideal statistics becomes smaller. It is interesting to note that the $P\bar{1}$–$P1$ separation is somewhat greater in Fig. 1(*b*) than in Fig. 1(*a*), showing the effect of an additional chlorine on the decrease of heterogeneity.

The results for the triclinic rhodium compound (Fig. 1*c*) show a much better discrimination between $P1$ and $P\bar{1}$ cumulative distributions for this level of hetero-

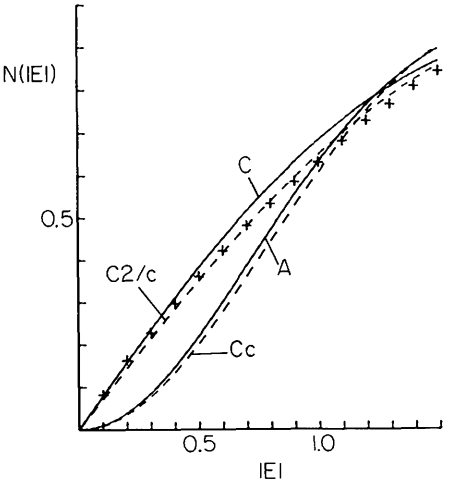

Fig. 2. A generalized $N(|E|)$ test for a solved monoclinic heterogeneous structure. The meaning of graphs and symbols is the same as in Fig. 1. Asymmetric unit: $C_{22}H_{18}O_3PPt$, space group $C2/c$ with $Z = 8$ (Koie, Shinoda, Saito, Fitzgerald & Pierpont, 1980).

geneity. The agreement between the recalculated $N(|E|)$ and that obtained for $P\bar{1}$ is worse than that for the two platinum compounds above but not harmfully so, since the recalculated $N(|E|)$ is displaced to the centric side of the $P\bar{1}$ curve. This may be due to some hypersymmetry which I failed to recognize or to the neglect of the Rh–Rh correlations (in the chemical molecule an Rh–Rh bond is bisected by a center of symmetry of $P\bar{1}$) which are also suggested by the somewhat wavy Wilson plot (*cf.* Wilson, 1981).

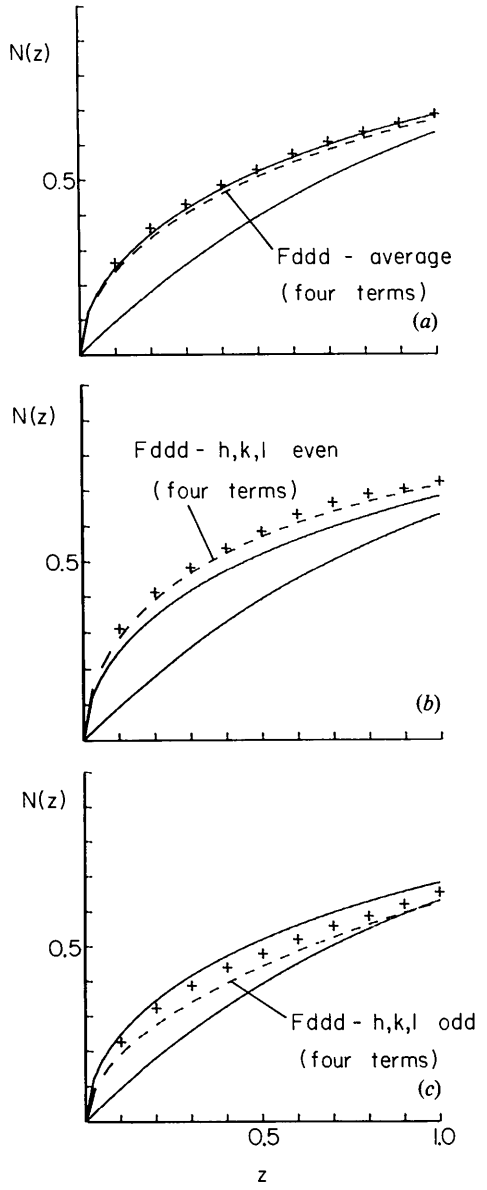

Fig. 3. $N(z)$ distributions in the space group *Fddd*. The solid curves are the conventional $N(z)$ distributions, the dashed curves are four-term expansions, calculated from equation (2) and the crosses represent the $N(z)$ values recalculated from the structure as described in the text. The asymmetric unit is $C_4H_{14}B_9BrO_2$ and the space group is *Fddd* with $Z = 32$ (Leonowicz & Scholer, 1980). (*a*) All recalculated $|F_c|$'s; (*b*) $|F_c|$'s with h, k and l even only; and (*c*) $|F_c|$'s with h, k and l odd only.

A much milder space-group dependence, and a good overall agreement is seen in the $N(|E|)$ test for the monoclinic platinum compound (Fig. 2).

Three $N(z)$ distributions are presented for the *Fddd* example (Fig. 3). Only four-term $N(z)$ expansions were evaluated since the available algorithms were not yet programmed for a general computation of the tenth moment of $|J|$, when this paper was written. However, this accuracy of the expansion should most probably suffice for the moderate heterogeneity of the structure examined (one bromine and fifteen light non-H atoms). Fig. 3(*a*) is based on the full independent set of $|F_c(hkl)|$'s and the theoretical distribution was computed using average values of γ_4, γ_6 and γ_8 $\{\bar{\gamma}_{2k} = \frac{1}{2}[\gamma_{2k}(\text{even}) + \gamma_{2k}(\text{odd})]\}$. The recalculated $N(z)$ in Fig. 3(*b*) rests on the $|F_c|$'s for which h, k and l are even while that in Fig. 3(*c*) was obtained from the $|F_c|$'s for which h, k and l are odd.

The existence of two different intensity distributions, in the recalculated diffraction pattern for this example, is evident. The recalculated $N(z)$ distributions follow the theoretical ones quite well but in all three cases (Figs. 3*a*, *b* and *c*) the recalculated distributions are displaced to the centric or hypercentric side, possibly because of the non-crystallographic mirror plane in the molecule (Leonowicz & Scholer, 1980), and the fact that bromine is located in this pseudo mirror plane or very close to it. The actual agreement between the recalculated and predicted $N(z)$ distributions is quite good for the *hkl*-even and all-data sets, the discrepancy being largest for the *hkl*-odd subset.

The values of γ_{2k}, used in the above examples, are listed in Table 2.

It is important to point out that the correct values of γ_{2k} as published (Shmueli & Kaldor, 1981) or given in Table 1 must be divided, for the purpose of comparison with conventional ideal distributions, by l'^{k-1}, where l' is the lattice multiplicity (*cf.* Table 1 and derivation above).

The examples presented in this section clearly indicate that generalized intensity statistics, such as those derived by Shmueli & Wilson (1981) and further extended in this paper, are of a definite practical importance and may often reduce the amount of trial

Table 2. *Values of γ_{2k} used in examples*

	γ_4	γ_6	γ_8	γ_{10}	
*P*1	1	1	1	1	
P$\bar{1}$	1·50	2·50	4·375	7·875	
Cc	1·50	2·50	4·375	7·875	
*C*2/*c*	2·25	6·25	19·141	62·015	
Fddd	3·375	15·625	83·740	–	(*hkl* even)
	1·875	4·375	11·553	–	(*hkl* odd)
	2·625	10	47·647	–	(average)

and error involved in the preliminary but crucially important stage of structure determination, the determination of the space group.

Other statistics, such as $P(|E|)$ functions (equations 1 and 13) and even moments of $|E|$ can also be readily computed and may provide complementary or supporting evidence to that obtained from the cumulative distributions. Of these two, the $P(|E|)$ functions are certainly superior and probably the most meaningful of all statistical tests, as they contain all the information on the distributions. However, the $P(|E|)$ curves often have a rather irregular appearance and it is thus easier (not safer!) to examine the cumulative distributions in which the irregularities in the observed probability density function are efficiently smoothed out by the integration which leads from $P(|E|)$ to $N(|E|)$.

I wish to thank Professor A. J. C. Wilson for his continuing interest in my work on intensity statistics and for the interesting, stimulating and encouraging correspondence. I am also grateful to him for letting me read his paper on intensity statistics and stereochemistry prior to its publication.

APPENDIX *A*
Expansion coefficients and moments of $|E|$

A formal derivation of the functional forms of the generalized probability density functions of $|E|$ (equations 1 and 13) can be found, subject to change of variable and notation, in Cramér's introduction to the theory of such expansions (Cramér, 1951, § 12.6). In fact, the first two examples to the above named section show that the 'ideal' Wilson-type (Wilson, 1949) probability functions for the centric and acentric distributions serve as weight functions in the orthogonality relationships involving Hermite and Laguerre polynomials, and this leads directly to the functional form of the equations given by Shmueli & Wilson (1981) and other authors who rederived these equations (Karle & Hauptman, 1953; Hauptman & Karle, 1953; Bertaut, 1955; Klug, 1958; Srinivasan & Parthasarathy, 1976). Another derivation seems unnecessary but the functional forms of the problem-dependent expansion coefficients appearing in (1) and (13) deserve a reconsideration in view of the common statistical significance of these equations.

Making use of Cramér's examples and his equation (12.6.3) (Cramér, 1951), the formal expressions for the probability density functions of $|E|$ can be written as

$$P_c(|E|) = P_c^{(0)}(|E|)\left[1 + \sum_{k=2}^{\infty} \langle He_{2k}(|E|)\rangle \times He_{2k}(|E|)/(2k)!\right] \quad (A1)$$

(*cf.* Klug, 1958) and

$$P_a(|E|) = P_a^{(0)}(|E|)\left[1 + \sum_{k=2}^{\infty} \langle L_k(|E|^2)\rangle L_k(|E|^2)\right] \quad (A2)$$

for the centrosymmetric and noncentrosymmetric space groups respectively, where $P_c^{((0)}(|E|)$ and $P_a^{(0)}(|E|)$ are the ideal centric and acentric probability density functions of $|E|$ respectively, based on Wilson's (1949) application of the central limit theorem, $He_{2k}(|E|)$ is an even-order Hermite polynomial and $L_k(|E|^2)$ is a Laguerre polynomial. The dependence on space-group symmetry and atomic composition is contained in the averages $\langle He_{2k}\rangle$ and $\langle L_k\rangle$ and these, in turn, are linear combinations of even moments of $|E|$, for the first few of which this dependence has been explicitly worked out (*cf.* Shmueli & Wilson, 1981; Appendix *B* below). The numerical coefficients of these linear combinations must depend on the ideal moments of $|E|$ (*cf.* Cramér, 1951), *i.e.* moments of the ideal centric and acentric distributions, and a reexamination of the explicit expressions for $He_{2k}(x)$ and $L_k(x^2)$, which contain these coefficients by definition, may reveal what is left of the rather complicated dependence of these polynomials on the ideal moments of their variables [Cramér, 1951, equation (12.6.1)].

In Table 22.3 of Abramowitz & Stegun (1972), we find

$$He_{2k}(x) = \sum_{m=0}^{k} (-1)^m \frac{(2k)!}{m!\, 2^m (2k-2m)!} x^{2(k-m)}. \quad (A3)$$

Now, $\langle |E|^{2p}\rangle^{(0)} = (2p-1)!!$ is the value of the $2p$th ideal centric moment of $|E|$ and the coefficients in (*A*3) can be rewritten in terms of such double factorials by noting that

$$(2k)! = [1 \times 3 \times \ldots \times (2k-1)]\, 2^k\, k!$$
$$= (2k-1)!!\, 2^k\, k!, \quad (A4)$$

with the understanding that $(2k-1)!!$ equals unity for $k = 0$. We thus obtain

$$He_{2k}(x) = \sum_{m=0}^{k} (-1)^m \binom{k}{m} \frac{(2k-1)!!}{(2k-2m-1)!!} x^{2(k-m)} \quad (A5)$$

and upon changing the index of summation from m to $p = k - m$, the explicit expression for $He_{2k}(x)$ becomes

$$He_{2k}(x) = \sum_{p=0}^{k} (-1)^{k-p} \binom{k}{p} \frac{(2k-1)!!}{(2p-1)!!} x^{2p}. \quad (A6)$$

The required Laguerre polynomials are given by

$$L_k(x^2) = \sum_{p=0}^{k} (-1)^p \binom{k}{k-p} \frac{1}{p!} x^{2p} \quad (A7)$$

(*e.g.* Abramowitz & Stegun, 1972). Upon dividing and multiplying (*A*7) by $(-1)^k k!$, and noting that $(-1)^{k+p} = (-1)^{k-p}$, we have

$$L_k(x^2) = \frac{1}{(-1)^k k!} \sum_{p=0}^{k} (-1)^{k-p} \binom{k}{p} \frac{k!}{p!} x^{2p}. \quad (A8)$$

Since the $2k$th ideal acentric moment of $|E|$ is given by $\langle |E|^{2k} \rangle^{(0)} = k!$, the summations in (*A*6) and (*A*8) have identical functional forms, as far as their dependence on the *ideal* moments of $|E|$ is concerned. The averages of He_{2k} and L_k in (*A*1) and (*A*2) can thus be written as

$$\langle He_{2k}(|E|) \rangle = \sum_{p=0}^{k} (-1)^{k-p} \alpha_{kp} \langle |E|^{2p} \rangle \quad (A9)$$

and

$$\langle L_k(|E|^2) \rangle = \frac{1}{(-1)^k k!} \sum_{p=0}^{k} (-1)^{k-p} \alpha_{kp} \langle |E|^{2p} \rangle, \quad (A10)$$

where

$$\alpha_{kp} = \binom{k}{p} \langle |E|^{2k} \rangle^{(0)} / \langle |E|^{2p} \rangle^{((0)} \quad (A11)$$

and

$$\langle |E|^{2k} \rangle^{(0)} = \begin{cases} (2k-1)!!, & \text{if centric} \\ k!, & \text{if acentric.} \end{cases} \quad (A12)$$

For computational purposes, the Hermite polynomials $H_n(x)$ related to $He_{2k}(x)$ by

$$H_{2k}\left(\frac{x}{\sqrt{2}}\right) = 2^k He_{2k}(x) \quad (A13)$$

(Abramowitz & Stegun, 1972) were found to be more convenient since more extensive tabulations exist for $H_n(x)$.

Substitution of (*A*9) and (*A*13) into (*A*1) leads to (1), and (13) is obtained by inserting (*A*10) into (*A*2).

The quantities α_{kp}, given by (*A*11) and (*A*12), also happen (?) to appear in the expressions for the non-ideal moments of $|E|$ in terms of the space-group symmetry and composition of the crystal (see Appendix *B*). Their introduction here may or may not be of theoretical interest but it certainly is of practical value: the number of numerical constants which had to be given along with the first five terms of (1) and (13) was reduced from 26 to six only, by making α_{kp} a part of the formalism.

APPENDIX *B*
Extension of the generalized statistics

In this Appendix the derivation of the tenth moment of $|E|$ will be described and some considerations which contribute to the simplification of the formalism will be presented. Full details of the derivations of the fourth, sixth and eighth moments were given by Shmueli & Wilson (1981) and the present derivation follows similar lines. However, the algebra required for an additional even moment is more tedious than that needed for all the available ones and the detailed derivation, which was simplified to some extent by a summation convention similar to that of Foster & Hargreaves (1963), has been deposited.†

The tenth moment of the structure amplitude $|F|$ (or the fifth moment of the reduced intensity FF^*) is given by

$$\langle |F|^{10} \rangle = \sum_{n_1} \sum_{n_2} \ldots \sum_{n_9} \sum_{n_{10}} \langle w_{n_1} w_{n_2}^* \ldots w_{n_9} w_{n_{10}}^* \rangle, \quad (B1)$$

where

$$w_n = f_n J_n, \quad (B2)$$

f_n and J_n being the scattering factor and the trigonometric structure factor of the nth atom respectively (*cf.* Shmueli & Wilson, 1981). According to the statistical properties of the trigonometric structure factor (Wilson, 1978), the non-vanishing terms in (*B*1) must be products of even moments of $|w|$ (*e.g.* $\langle |w|^4 \rangle \langle |w|^2 \rangle \langle |w|^2 \rangle \langle |w|^2 \rangle$, and $\langle |w|^6 \rangle \langle |w|^2 \rangle \langle |w|^2 \rangle$, *etc.*). (*B*1) thus splits up into a series of multiple summations, each involving one kind of such product, and a single sum of the tenth moments of $|w|$. The multiplicities of these partial summations differ for the centrosymmetric and noncentrosymmetric space groups and can be evaluated by direct counting, by elementary combinatorics, or by solving the corresponding partition problem of combinatorial analysis (*e.g.* Abramowitz & Stegun, 1972). The second approach, adopted here, is illustrated below.

Suppose we wish to find the number of different ways in which $2n$ indices [$n = 5$ in (*B*1)] can be contracted into a group of $2k$ indices and $n - k$ groups of two indices each, to give the product

$$\langle |w|^{2k} \rangle \underbrace{\langle |w|^2 \rangle \ldots \langle |w|^2 \rangle}_{n-k \text{ terms}}. \quad (B3)$$

The number of different ways in which $2n - 2k$ indices can be contracted into $n - k$ pairs is $(2n - 2k - 1)!!$ and $(n - k)!$ in the centrosymmetric and noncentrosymmetric case respectively (Shmueli & Wilson, 1981). We further recall that any even-numbered group of indices can be contracted to give a non-zero moment of $|w|$ in the centrosymmetric case, since then $J = J^*$ (Wilson, 1978), while in the noncentrosymmetric case each such group must contain equal numbers of indices

† This material, entitled *The Derivation of the Tenth Moment* has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 36631 (13 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

appended to w's and to w^*'s in ($B1$) or analogous expressions. It follows that the multiplicity of a summation of products given by ($B3$) is

$$\binom{2n}{2k}(2n-2k-1)!! = \binom{n}{k}\frac{(2n-1)!!}{(2k-1)!!} \qquad (B4)$$

and

$$\binom{n}{k}^2 (n-k)! = \binom{n}{k}\frac{n!}{k!} \qquad (B5)$$

for centrosymmetric and noncentrosymmetric space groups respectively. The products $\langle|w|^4\rangle\langle|w|^4\rangle\langle|w|^2\rangle$ and $\langle|w|^6\rangle\langle|w|^4\rangle$ are treated directly by similar considerations.

It is interesting to note that these multiplicities have the same forms as the coefficients α_{nk} in the Hermite and Laguerre polynomials arrived at in Appendix A [*cf.* ($A6$) and ($A7$)] and this fact accounts for the functional form of equations (5)–(9).

Next, the various partial summations, to which appropriate multiplicities have been assigned, are decomposed into a combination of single summations and their products. This process, rather simple for low-order moments, becomes the labour-determining step in the derivation of higher moments, and simplified summation conventions are almost essential.

Assuming that all the atoms are located in general positions and dispersion is negligible, these single summations are simplified exactly as shown elsewhere (Shmueli & Wilson, 1981) and the tenth moment of $|E|$ (or any other even moment) is obtained from the relation

$$\langle|E|^{2k}\rangle = \langle|F|^{2k}\rangle/\langle|F|^2\rangle^k. \qquad (B6)$$

The expansion coefficients A_{2k} or B_{2k} [*cf.* (1) and (13)] as functions of moments of the trigonometric structure factor and atomic composition are then obtained by comparing (3), or the acentric analogue, with the moments given by ($B6$), the derivation of which was outlined above and is described in detail in the deposited material.

The result for the new coefficients is

$$A_{10} \text{ or } B_{10} = hQ_{10} + W(\gamma_4 - \alpha_{20})(\gamma_6 - \alpha_{32}\gamma_4 + 2\alpha_{30}) \times Q_4 Q_6, \qquad (B7)$$

where

$$h = \gamma_{10} - \alpha_{54}\gamma_8 + 2\alpha_{53}\gamma_6 - 6\alpha_{42}\gamma_4 + 24\alpha_{50} - W\gamma_4\gamma_6 + 2V\gamma_4^2, \qquad (B8)$$

W and V being 210 and 1575 or 100 and 450 for centrosymmetric or noncentrosymmetric space groups respectively, and Q_{2k} and γ_{2k} are defined by (10) and (11).

References

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

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

COTTON, F. A. & FELTHOUSE, T. R. (1980). *Inorg. Chem.* **19**, 323–328.

CRAMÉR, H. (1951). *Mathematical Methods of Statistics*. Princeton Univ. Press.

FAGGIANI, R., LIPPERT, B. & LOCK, C. J. L. (1980). *Inorg. Chem.* **19**, 295–300.

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

HAUPTMAN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136–141.

HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210–214.

International Tables for X-ray Crystallography (1952). Vol. I. Birmingham: Kynoch Press.

KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131–135.

KLUG, A. (1958). *Acta Cryst.* **11**, 515–543.

KOIE, Y., SHINODA, S., SAITO, Y., FITZGERALD, B. J. & PIERPONT, C. G. (1980). *Inorg. Chem.* **19**, 770–773.

LEONOWICZ, M. E. & SCHOLER, F. R. (1980). *Inorg. Chem.* **19**, 122–124.

MAIN, P., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J. P. & WOOLFSON, M. M. (1978). *MULTAN 78. A System of Computer Programs for Solution of Crystal Structures by Direct Methods*. Univ. of York, England.

SHMUELI, U. (1979). *Acta Cryst.* A**35**, 282–286.

SHMUELI, U. (1981). *Acta Cryst.* A**37**, C321.

SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* A**37**, 76–80.

SHMUELI, U., KALDOR, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, C332.

SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 342–353.

SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

WILSON, A. J. C. (1978). *Acta Cryst.* A**34**, 986–994.

WILSON, A. J. C. (1980). Private communication.

WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 808–810.