Acta Cryst. (1983). A**39**, 225–233

Generalized Intensity Statistics: The Subcentric Distribution and Effects of Dispersion

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* 17 *May* 1982; *accepted* 28 *September* 1982)

Abstract

The well-known acentric and centric distributions apply, asymptotically in the number of atoms in the unit cell, when there is no crystallographic symmetry or centrosymmetry only. Series expansions, involving Laguerre or Hermite polynomials, can be obtained, which take into account paucity of (heterogeneous) atoms and higher space-group symmetries. The asymptotic as well as the generalized distributions are further modified if (i) the crystal exhibits partial (non-space group) symmetry, and (ii) if some atoms exhibit appreciable dispersion. This article deals with the generalization of the asymptotic 'subcentric' distribution of the normalized intensity

$$P(z)\,dz = (\alpha^2 - \beta^2)^{1/2} \exp(-\alpha z)\, I_0(\beta z)\,dz$$

which accommodates both partial (non-crystallographic) centrosymmetry and effects of dispersion. A four-term Gram–Charlier expansion with appropriate orthogonal polynomials has been derived for the subcentric distribution and detailed expressions for the required moments of z have been obtained for the case of dispersion. This generalization, *i.e.* the orthogonal polynomials, the moments of z and the asymptotic subcentric distribution, incorporates the generalized acentric and dispersionless centric expansions as limiting cases. The above derivation has been brought to completion using computer-algebraic techniques, which permit the use of well-established but rarely used mathematical methods in an *ab-initio* generalization of a given asymptotic distribution of intensity.

Introduction

For non-centrosymmetric crystals with sufficiently large numbers of atoms in the unit cell, the probability of a randomly chosen reflection having a normalized intensity between z and $z + dz$ is given by

$$P(z)\,dz = \exp(-z)\,dz \qquad (1)$$

and for centrosymmetric crystals by

$$P(z)\,dz = (2\pi z)^{-1/2} \exp(-z/2)\,dz \qquad (2)$$

(Wilson, 1949), where $z = I/\Sigma$, I is the (corrected) integrated intensity of the reflection and Σ is the sum of the squares of the moduli of the atomic scattering factors (Wilson, 1942, 1978). When the number of atoms in the unit cell is too small, particularly when there is great inhomogeneity of scattering factor, the above asymptotic expressions can be expanded into series based on orthogonal polynomials: Hermite for the centric distribution (2) and Laguerre for the acentric distribution (1) (see, for example, Karle & Hauptman, 1953; Hauptman & Karle, 1953; Bertaut, 1955; Klug, 1958; Shmueli, 1979; Shmueli & Wilson, 1981). Some applications of such symmetry- and composition-dependent expansions were recently illustrated (Shmueli, 1982). When non-crystallographic symmetry is present the asymptotic distributions take other forms, bicentric (Lipson & Woolfson, 1952), hypercentric (Wilson, 1952; Rogers & Wilson, 1953) or sesquicentric (Wilson, 1956). If the structure is intermediate between centrosymmetric and non-centrosymmetric, still other asymptotic distributions, for which the name 'subcentric' might be coined, appear. Examples of subcentric distributions arising from partial structural symmetry are given by Srinivasan & Parthasarathy (1976, ch. 3), and Wilson (1980*a*) has shown that an asymptotic distribution of the same analytic form appears when dispersion produces an effective departure from centrosymmetry. These hypercentric and subcentric distributions can be expressed in closed form in some cases and can be expanded in series in terms of orthogonal polynomials as mentioned above, but, as Rogers & Wilson (1953) pointed out, convergence is slow. This was a serious difficulty in 1953, but with the present-day computing facilities there is no longer a problem in summing many terms of

* Present address: Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW.

0567-7394/83/020225-09$01.50

the series as long as their analytic forms are available. Until recently, the main problem was in constructing the required theoretical expressions for the orthogonal polynomials and problem-dependent moments of intensity or related quantities, even if the mathematical procedures underlying such constructions were known. As will be seen from this paper, the often forbidding algebra which is involved in such derivations (except for a few simple asymptotic distributions) does not present a serious difficulty since all or most of these tedious manipulations can now be performed by symbolic computer programs which perform a variety of algebraic operations and can also output the results so that they can be directly incorporated into programs for relevant numerical computations. A recent version of the *LISP*-based *REDUCE* system (Hearn, 1973) was employed in this work for such algebraic manipulations and 'program writing'.

The present paper is concerned with the generalization of the asymptotic subcentric distribution, for which an approximating expansion based on appropriate orthogonal polynomials will be derived, and with a theoretical treatment of the effects of space-group symmetry, atomic heterogeneity and dispersion on the distribution of the normalized intensity.

Mathematical background

Although the mathematical background of our derivation is amply documented in the literature, it will be summarized here both for the sake of completeness and because the general relationships to be shown were actually used as they stand.

Given an experimental distribution $P_{\rm exp}(z)\,dz$ which departs from a known asymptotic distribution $P^{(0)}(z)\,dz$, where $P^{(0)}(z)$ accounts for the basic properties of the experiment that gave rise to $P_{\rm exp}(z)\,dz$, we can try to approximate the experimental distribution by a generalized expansion of the form

$$P_g(z)\,dz = \sum_k g_k f_k(z) P^{(0)}(z)\,dz, \tag{3}$$

where $\{f_k(z)\}$ is a set of orthogonal polynomials obeying the relationship

$$\int_a^b f_k(z) f_l(z) P^{(0)}(z)\,dz = \delta_{kl}; \tag{4}$$

δ_{kl} is the Kronecker symbol and $[a,b]$ is the range of existence of all the functions of z involved. Making use of (3) and (4), the expansion coefficients g_k in (3) are obtained as

$$g_k = \int_a^b f_k(z) P_g(z)\,dz \tag{5}$$

(Szegö, 1939; Cramér, 1951).

It can be shown that if the distribution

$$N^{(0)}(z) = \int_0^z P^{(0)}(z)\,dz \tag{6}$$

is non-decreasing and if all the moments

$$c_k = \int_a^b z^k P^{(0)}(z)\,dz \tag{7}$$

of the asymptotic distribution exist, the polynomials $f_k(z)$ are uniquely associated with $P^{(0)}(z)$ [often termed the weight function of the orthogonality relationship (4)] and the coefficients of the powers of z, in these polynomials, are expressible in terms of the moments c_k of the asymptotic distribution (*cf.* Szegö, 1939; Cramér, 1951). Moreover, it is seen from (5) that the expansion coefficients g_k are given by averages of the corresponding polynomials with respect to the generalized distribution, and hence these coefficients can be determined by replacing the powers z^l, in the polynomials, with their mean values or moments $\langle z^l \rangle$, which take into account the problem that caused the departure of $P_{\rm exp}(z)$ from $P^{(0)}(z)$ and hence called for a generalization of the latter.

The explicit expression for the polynomial of degree n is given by

$$f_n(z) = K_n \begin{vmatrix} z^n & z^{n-1} & \cdots & z & 1 \\ c_{2n-1} & c_{2n-2} & \cdots & c_n & c_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n+1} & c_n & \cdots & c_2 & c_1 \\ c_n & c_{n-1} & \cdots & c_1 & c_0 \end{vmatrix} \tag{8}$$

(*cf.* Szegö, 1939), where the c_k's are given by (7) and the normalization constant is given by $K_n = (D_{n-1} D_n)^{-1/2}$, where the determinant D_n can be obtained from that in (8) by replacing the powers $z^n, z^{n-1}, \ldots, z^0$ with the moments $c_{2n}, c_{2n-1}, \ldots, c_n$, respectively (*cf.* Szegö, 1939). It is, however, more convenient to rewrite (8) as

$$f_n(z) = K_n \sum_{k=0}^{n} d_{nk} z^k, \tag{9}$$

where d_{nk} is the signed minor corresponding to the element z^k in the first row of the determinant in (8). Then, making use of the requirement

$$\int_a^b [f_n(z)]^2 P^{(0)}(z)\,dz = 1 \tag{10}$$

and equation (7), we obtain

$$K_n = \left(\sum_{k=0}^{n} \sum_{l=0}^{n} d_{nk} d_{nl} c_{k+l} \right)^{-1/2}. \tag{11}$$

Since $f_0(z)$ must equal unity, the generalized expansion (3) can thus be written as

$$P_g(z) = P^{(0)}(z)\left[1 + \sum_{k=1}^{\infty}\left(\sum_{p=0}^{k} d_{kp}\langle z^p\rangle\right) \times \left(\sum_{q=0}^{k} d_{kq} z^q\right) K_k^2\right]. \quad (12)$$

Of course, for the above equations to be exactly obeyed, the asymptotic distribution must be normalized to unity and it can be shown that the same holds for $P_g(z)$ even if (12) is terminated after a finite number of terms.

Hence, given an asymptotic distribution which satisfies (6) and (7), the quantities required for its generalization are its moments c_k, the minors d_{nk} and the problem-dependent mean values (moments) $\langle z^k\rangle$.

The subcentric distribution

The general form of the subcentric distribution, applicable to partial symmetry (Srinivasan & Parthasarathy, 1976, ch. III) and dispersion (Wilson, 1980*a*), is

$$P^{(0)}(z)\,dz = \gamma \exp(-\alpha z) I_0(\beta z)\,dz, \quad (13)$$

where I_0 is a modified Bessel function of the first kind (Abramowitz & Stegun, 1972, ch. 9). The parameters α, β and γ are connected by two relations:

$$\gamma^2 = \alpha^2 - \beta^2 = \alpha; \quad (14)$$

the first arises because any probability distribution has unit area, and the second arises because in the present applications the intensity has been normalized so that $\langle z\rangle = 1$. In the rest of this paper, therefore, γ is replaced by $\alpha^{1/2}$.

In the case of dispersion, we have

$$\alpha = \frac{\Sigma^2}{\Sigma^2 - S^2} \quad (15)$$

and

$$\beta = \frac{S\Sigma}{\Sigma^2 - S^2} \quad (16)$$

(Wilson, 1980*a*), where S is the modulus of the sum of the squares of the complex scattering factors (Wilson, 1978, 1980*a*). Also, a rearrangement of the distribution parameters in equation (3.84) of Srinivasan & Parthasarathy (1976) shows that their asymptotic distribution, for the case of several centrosymmetric atomic groups being present in the unit cell of the space group $P1$, reduces to (13) with S in (15) and (16) replaced with Σ_c, the sum of the squares of the scattering factors of atoms which belong to the centrosymmetric groups (Srinivasan & Parthasarathy, 1976). It is interesting to note that α tends to unity if S (for dispersion) or Σ_c (for partial centrosymmetry) tend to zero and, since β must then tend to zero, (13) reduces in this limit to (1), the asymptotic acentric distribution. This is obvious in the case of partial centrosymmetry, and in the case of dispersion this corresponds to uncorrelated real and imaginary parts of the structure factor – strictly valid in the acentric case (Wilson, 1980*a*).

In the other limiting situation, *i.e.* when S or Σ_c tend to Σ, α tends to infinity. In the case of dispersion this corresponds to a tendency to the ideal centric distribution (Wilson, 1980*a*), while in the case of partial symmetry this means that the unit cell of $P1$ is occupied (in the limit) by centrosymmetric atomic groups only. As will be seen below, the limiting asymptotic distributions must be the ideal centric one, given by (2), in both applications.

The moments of (13) can be found making use of the known definite integral

$$\int_0^{\infty} t^{\mu} \exp(-t\cosh\theta) I_{\nu}(t\sinh\theta)\,dt = \Gamma(\mu+\nu+1) P_{\mu}^{-\nu}(\cosh\theta) \quad (17)$$

(Watson, 1922; Gradshteyn & Ryzhik, 1980), where $P_{\mu}^{-\nu}$ is a Legendre function. With $\nu = 0$ and μ an integer n, the nth moment of (13) is then given by

$$c_n = \alpha^{1/2}\int_0^{\infty} z^n \exp(-\alpha z) I_0(\beta z)\,dz = n!\,\alpha^{-n/2} P_n(\alpha^{1/2}), \quad (18)$$

where $P_n(x)$ are the usual Legendre polynomials [$P_0 = 1$, $P_1 = x$, $P_2 = \frac{1}{2}(3x^2 - 1)$, *etc.*]. This result for the moments of the asymptotic subcentric distribution is much simpler than that given by Srinivasan & Parthasarathy (1976) in terms of hypergeometric functions.

Since all the expressions required for setting up the subcentric version of (12) were computer generated, it was more convenient to rewrite (18) in terms of the Rodriguez formula for Legendre polynomials as

$$c_n = (2x)^{-n}\frac{d^n}{dx^n}(x^2-1)^n, \quad (19)$$

where $x = \alpha^{1/2}$, and to use (19) for the generation of moments up to $n = 8$.

These moments were used in expanding the determinants in (8) for $n = 1$, 2, 3 and 4, whereafter the coefficients d_{nk} of the powers of z were automatically separated and expressions for the corresponding normalization constants were obtained making use of (11). Thus, the expressions for the orthonormal polynomials $f_1(z)$, $f_2(z)$, $f_3(z)$ and $f_4(z)$, associated with the asymptotic subcentric distribution (13), as well as

all the data (apart from the mean values $\langle z^k \rangle$) required for the construction of a four-term generalized expansion (12), have been obtained. Finally, five terms of (12) are available but in the present case the second term vanishes identically since $\langle z \rangle = 1$ (*cf.* Appendix *B*).

Detailed results of this calculation have been deposited* along with a listing of the Algol-like program of *REDUCE* instructions used. It was also found convenient to perform the substitutions, which lead to the limiting distributions, in the same computation and the results of some of these tests are also included in the deposited material. Of course, these reductions to limiting cases also serve as a check on the correctness of the computation. Thus, for $\alpha = 1$, c_n reduced to $n!$ and the normalized polynomials $f_n(z)$ to Laguerre polynomials $L_n(z)$, as expected. It should be noted, however, that polynomials generated by (8) are only orthogonal but not standardized, and thus, for example, the $(-1)^n$ factor standardizing the Laguerre polynomials (*cf.* Abramowitz & Stegun, 1972) is missing. This is of no importance as far as (12) is concerned but should be kept in mind when (8) is used as a generator of orthogonal polynomials for other applications. The other limiting case, $\alpha \to \infty$, yielded $c_n = (2n-1)!!$ and the determinants for $f_k(z)$ reduced to Hermite polynomials thus confirming that this limiting case corresponds to the generalized centric distribution (*cf.* Shmueli & Wilson, 1981). The limiting values of c_n in the latter case can also be obtained directly using the general definition of the Legendre polynomials (*e.g.* Abramowitz & Stegun, 1972, ch. 22). After some rearrangement we obtain

$$\alpha^{-n/2} P_n(\alpha^{1/2}) = \frac{(2n-1)!!}{n!} \times \left[1 - \frac{n(n-1)}{2(2n-1)} \alpha^{-1} + O(\alpha^{-2})\right] \quad (20)$$

and as α tends to infinity, all the terms on the r.h.s. of (20), except the first one, tend to zero and we have from (18)

$$c_n = (2n-1)!! \quad (21)$$

for this limiting case. These moments define the centric distribution (2) and their insertion into (8) is indeed expected to give Hermite polynomials. The first four moments of (13) and the first two polynomials, generated as described above, are given in Appendix *B*.

* This material, entitled *Computer-generated orthogonal expansion for the subcentric distribution*, has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 38149 (17 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

In order to complete the construction of the generalized expansion (12) for the subcentric distribution, we need the theoretical expressions for the moments (mean values) of normalized intensity which will include the dependence on the factors that call for the generalization of (13). In this paper, we shall deal with the case of dispersion only which has been studied by us in some detail. Of course, (12) with the here available polynomials will become a generalization of the asymptotic distribution for the case of partial symmetry, to symmetries other than triclinic and arbitrary compositions, once the corresponding moments of z become available.

Only the terms of (12) with $k \leq 4$ have been generated, in view of our previous experience with such expansions (Shmueli & Wilson, 1981; Shmueli, 1982). However, more terms can be readily obtained if need arises.

Effects of dispersion

Derivations of higher even moments of the structure amplitude $|F|$ have been presented and discussed by several investigators (*e.g.* Karle & Hauptman, 1953; Hauptman & Karle, 1953; Foster & Hargreaves, 1963; Wilson, 1978; Shmueli & Wilson, 1981, 1982; Shmueli, 1982). As the study of intensity statistics progresses and additional factors are taken into account, the need is felt for a more general scheme of derivation and for reducing the labour involved. The present approach will hopefully contribute to the achievement of the first objective while the problem of systematizing the algebra is dealt with in Appendix *A*. Since, however, the notation of the latter Appendix will be used in this subsection, the reader is advised to consult Appendix *A* prior to examining the general expressions for the moments of $|F|$ and z which are derived here.

The $2n$th absolute moment of $|F|$ is given by

$$\langle |F|^{2n} \rangle = \langle (FF^*)^n \rangle \equiv \sum^{2n} C\langle S \rangle, \quad (22)$$

where the symbol $\sum^{2n}$ denotes a $2n$-fold summation which involves indices ranging over the atoms of the asymmetric unit, C is a $2n$-fold product of atomic scattering factors (and their complex conjugates) and S is a $2n$-fold product of the corresponding trigonometric structure factors (*cf.* Shmueli & Wilson, 1981). Thus, C and S reflect the composition and symmetry of the crystal considered respectively. Since, however, the statistical properties of the structure factor and its moments (to be understood here as mean values of the powers of $|F|$) are due to the symmetry terms S, the actual averaging is performed on the latter quantities over a large number of reflections with similar values of $\sin\theta/\lambda$, and hence the composition terms C in (22) are taken as constants in this process.

It has been shown that non-vanishing averages $\langle S \rangle$ must be products of mean values of even powers of moduli of the trigonometric structure factor (Wilson, 1978; Shmueli & Kaldor, 1981), usually referred to as absolute even moments of this factor. Each such product can be formed in a number of different ways depending on whether the space group is centrosymmetric or noncentrosymmetric (Wilson, 1978; Shmueli & Wilson, 1981; Shmueli, 1982). As will be seen below, in the centrosymmetric case with the scattering factors being possibly complex, the composition terms C can assume several forms for the same representations of $\langle S \rangle$. When the multiplicities of the $C\langle S \rangle$ arrangements are determined, (22) can be decomposed into partial summations as

$$\langle |F|^{2n} \rangle = \sum_k m_k \sum\nolimits^I (C\langle S \rangle)_k, \tag{23}$$

where $(C\langle S \rangle)_k$ is a non-vanishing product arrangement of the $C\langle S \rangle$ term with multiplicity m_k and $\sum^I$ denotes an inner partial summation.

Coming to the present application of the above, a typical inner partial summation appearing in the eighth moment $|F|$, with dispersion, is

$$\sum_{i \neq j \neq k} \sum \sum |f_i|^4 f_j^2 f_k^{*2} q_i p_j p_k. \tag{24}$$

Here, p and q are the second and fourth absolute moments of the trigonometric structure factor respectively and f_j is the complex scattering factor of the jth atom. The f-dependent coefficient $C = |f_i|^4 f_j^2 f_k^{*2}$, for this summation, can be schematically written as

$$(ff^* \, ff^*)(ff)(f^* \, f^*), \tag{25}$$

the parentheses containing those f's for which the indices in the original eightfold summation were contracted. The number of different ways of placing two f's and two f^*'s in the first group in (25) is $\binom{4}{2}^2 = 36$, the remaining two groups being formed in one distinct way only. Hence, the multiplicity of (24) in (23) with $n = 4$, is $m = 36$. There are several other possible arrangements of the C term which multiply $q_i p_j p_k$ and the sum of the multiplicities of all such arrangements must of course equal the number of different ways in which the original eight indices could be contracted to form the product of moments $q_i p_j p_k$ (this number is 210 – *cf.* Shmueli & Wilson, 1981).

Table 1 contains the forms of the C and S products and the multiplicities of the corresponding partial summations for the fourth, sixth and eighth moments of $|F|$, with complex scattering factors. The usual designations p, q, r and s for the second, fourth, sixth and eighth absolute moments of the trigonometric structure factor (Wilson, 1978; Shmueli & Wilson, 1981, 1982) are employed.

Each line of the table corresponds to a partial summation such as, for example, (24) and the above three moments of $|F|$ can be readily written down using this table with (23). Moments of $|F|$ for noncentrosymmetric space groups are constructed from those lines only which contain in the fifth column the 'a and c' indicator while all the lines are relevant for the centric even moments of $|F|$. Dispersionless centric moments are obtained by replacing $|f|^{2k}$, f^{2k} and f^{*2k} with (real) f^{2k}, collecting like terms and accumulating the appropriate multiplicities (*cf.* Shmueli & Wilson, 1981).

In order to obtain useful expressions from which the moments of $|F|$ can actually be computed, the partial (multiple) summations have to be decomposed into single summations which can be more readily inspected and evaluated. Methods for doing this were given elsewhere (Shmueli & Wilson, 1981; Shmueli, 1982) and a more general one which is also suitable to a computer-algebraic approach is given in Appendix *A*.

Introducing the abbreviations

$$\mu_2 = p|f|^2, \quad \mu_4 = q|f|^4, \quad \mu_6 = r|f|^6, \quad \mu_8 = s|f|^8,$$
$$\nu_2 = pf^2, \quad \nu_4 = qf^4, \quad \theta_4 = q|f|^2 f^2, \quad \theta_6 = r|f|^4 f^2, \tag{26}$$

using the notation of Appendix *A* and the contents of Table 1 with (23), the fourth, sixth and eighth moments of $|F|$, for the centrosymmetric case with complex scattering factors, can be written as

$$\langle |F|^4 \rangle = 2S_2(\mu_2,\mu_2) + S_2(\nu_2,\nu_2^*) + S_1(\mu_4) \tag{27}$$

$$\begin{aligned}\langle |F|^6 \rangle = {} & 6S_3(\mu_2,\mu_2,\mu_2) + 9S_3(\mu_2,\nu_2,\nu_2^*) \\ & + 9S_2(\mu_4,\mu_2) + 6\,\mathrm{Re}[S_2(\theta_4,\nu_2^*)] \\ & + S_1(\mu_6)\end{aligned} \tag{28}$$

$$\begin{aligned}\langle |F|^8 \rangle = {} & 24S_4(\mu_2,\mu_2,\mu_2,\mu_2) + 72S_4(\mu_2,\mu_2,\nu_2,\nu_2^*) \\ & + 9S_4(\nu_2,\nu_2,\nu_2^*,\nu_2^*) + 72S_3(\mu_4,\mu_2,\mu_2) \\ & + 36S_3(\mu_4,\nu_2,\nu_2^*) + 6\,\mathrm{Re}[S_3(\nu_4,\nu_2^*,\nu_2^*)] \\ & + 96\,\mathrm{Re}[S_3(\theta_4,\mu_2,\nu_2^*)] + 16S_2(\mu_6,\mu_2) \\ & + 12\,\mathrm{Re}[S_2(\theta_6,\nu_2^*)] + 18S_2(\mu_4,\mu_4) \\ & + S_2(\nu_4,\nu_4^*) + 16S_2(\theta_4,\theta_4^*) + S_1(\mu_8)\end{aligned} \tag{29}$$

and the corresponding moments of z, required for constructing the expression coefficients in (12), with $n = 4$, are given by

$$\langle z^k \rangle = \langle |F|^{2k} \rangle / S_1^k(\mu_2), \quad k = 1, \ldots, 4, \tag{30}$$

where $S_1(\mu_2)$ is the average reduced intensity Σ.

The expansion of (27)–(29) into useful combinations and products of single summations has been performed, using the decomposition algorithm given in Appendix *A*, with the *REDUCE* system which was employed above for the construction of the required orthogonal polynomials. The correctness of these readily obtained (but lengthy) expansions was tested in

Table 1. *Detailed structure of the fourth, sixth and eighth absolute moments of F [cf. equation (29)]*

The order of moment, $2n$, appears in the first column and C, $\langle S \rangle$ are the composition and symmetry factors respectively in the partial summations in (23), while m is the corresponding muptiplicity of a partial summation; f_i is the complex scattering factor of the ith atom, $|f_i|$ is its magnitude and f_i^* its complex conjugate; the symbols p_i, q_i, r_i and s_i denote moments of the trigonometric structure factor of the ith atom (see text). The fifth column with 'a and c' shows that the partial summation in this line appears in expressions for both acentric and dispersionless centric moments, while an entry with 'c' belongs to a centric moment only.

$2n$	C	$\langle S \rangle$	m	a and/or c
4	$\|f_i\|^2\|f_j\|^2$	p_ip_j	2	a and c
	$f_i^2f_j^{*2}$	p_ip_j	1	c
	$\|f_i\|^4$	q_i	1	a and c
6	$\|f_i\|^2\|f_j\|^2\|f_k\|^2$	$p_ip_jp_k$	6	a and c
	$\|f_i\|^2f_j^2f_k^{*2}$	$p_ip_jp_k$	9	c
	$\|f_i\|^4\|f_j\|^2$	q_ip_j	9	a and c
	$\|f_i\|^2f_i^2f_j^{*2}$	q_ip_j	3	c
	$\|f_i\|^2f_i^{*2}f_j^2$	q_ip_j	3	c
	$\|f_i\|^6$	r_i	1	a and c
8	$\|f_i\|^2\|f_j\|^2\|f_k\|^2\|f_l\|^2$	$p_ip_jp_kp_l$	24	a and c
	$\|f_i\|^2\|f_j\|^2f_kf_l^{*2}$	$p_ip_jp_kp_l$	72	c
	$f_i^2f_j^2f_k^{*2}f_l^{*2}$	$p_ip_jp_kp_l$	9	c
	$\|f_i\|^4\|f_j\|^2\|f_k\|^2$	$q_ip_jp_k$	72	a and c
	$\|f_i\|^4f_j^2f_k^{*2}$	$q_ip_jp_k$	36	c
	$f_i^4f_j^{*2}f_k^{*2}$	$q_ip_jp_k$	3	c
	$f_i^{*4}f_j^2f_k^2$	$q_ip_jp_k$	3	c
	$(\|f_i\|^2f_i^2)\|f_j\|^2f_k^{*2}$	$q_ip_jp_k$	48	c
	$(\|f_i\|^2f_i^{*2})\|f_j\|^2f_k^2$	$q_ip_jp_k$	48	c
	$\|f_i\|^6\|f_j\|^2$	r_ip_j	16	a and c
	$(\|f_i\|^4f_i^2)f_j^{*2}$	r_ip_j	6	c
	$(\|f_i\|^4f_i^{*2})f_j^2$	r_ip_j	6	c
	$\|f_i\|^4\|f_j\|^4$	q_iq_j	18	a and c
	$f_i^4f_j^{*4}$	q_iq_j	1	c
	$(\|f_i\|^2f_i^2)(\|f_j\|^2f_j^{*2})$	q_iq_j	16	c
	$\|f_i\|^8$	s_i	1	a and c

the same computer run by performing the substitutions which should lead to the acentric and dispersionless centric moments of the normalized intensity. Thus, upon substituting (i) $\nu_2 = \nu_4 = \theta_4 = \theta_6 = 0$ and (ii) $\nu_2 = \mu_2$, $\nu_4 = \theta_4 = \mu_4$, $\theta_6 = \mu_6$, the acentric and centric moments of z (Shmueli & Wilson, 1981) have been recovered, respectively, from the expansions of the general expressions (27)–(29) and (30). The expanded expressions for the above moments of z, as well as the above mentioned tests, form part of the deposited material. The same expressions, output by *REDUCE* in a Fortran-compatible format, were incorporated into a program for a numerical evaluation of the subcentric version of (12), terminated as described above.

Thus, apart from eliminating tedious and error-prone 'hand' manipulations, and performing work which would probably not be attempted to the above extent, computer-algebraic routines can also be used to advantage in an examination of limiting forms of the expressions derived. Obviously, these powerful programs must be judiciously employed and, if this requirement is satisfied, we believe that the danger of 'losing sight of the physics' is about the same as that inherent in many other computer applications (*cf. Computers in the New Laboratory* – A *Nature* Survey, 1981).

It is interesting to discuss the above tendency of the general centric moments to the acentric and dispersionless centric ones by considering the two limiting distributions of phases of the atomic scattering factors:

(i) the phases of f are uniform in the $(0,2\pi)$ interval, and

(ii) the phase is constant.

Let us expand (27) with the aid of $(A3)$ and $(A4)$ and rewrite it in the usual form.

$$\langle |F|^4 \rangle = 2S_1^2(\mu_2) - 2S_1(\mu_2^2) + S_1(\nu_2)\,S(\nu_2^*) - S_1(\nu_2\,\nu_2^*) + S_1(\mu_4) \tag{31}$$

$$= 2\Sigma + |\sum_k p_k f_k^2|^2 + \sum_k (q_k - 3p_k^2)|f_k|^4 \tag{32}$$

(Wilson, 1978). The third term in (31) can be expanded as

$$S_1(\nu_2)\,S_1(\nu_2^*) = \sum_k \sum_l p_k\,p_l f_k^2 f_l^{*2} = \sum_k \sum_l p_k\,p_l |f_k|^2 |f_l|^2 \exp[2i(\varphi_k - \varphi_l)], \tag{33}$$

where $f = |f| \exp(i\varphi)$, and in the case of constant phasse ($\varphi_k = \varphi_l$ = constant for any k and l), (33) reduces to Σ^2 whereupon (32) becomes the dispersionless centric fourth moment of $|F|$ (Wilson, 1978). Equation (33) can also be rewritten as

$$S_1(\nu_2)\,S_1(\nu_2^*) = \sum_k p_k^2 |f_k|^4 + 2\sum_{k>l}\sum p_k\,p_l |f_k|^2 |f_l|^2 \times \cos 2(\varphi_k - \varphi_l). \tag{34}$$

Here, in the case of uniform phase distribution, the second summation in (34) is most likely to be zero and we have

$$S_1(\nu_2)\,S_1(\nu_2^*) = \sum_k p_k |f_k|^4 = S_1(\mu_2^2) \tag{35}$$

in the uniform-phase limit. Substituting (35) into (31), the acentric fourth moment of $|F|$ (Wilson, 1978) is obtained.

The above consideration, apart from being interesting for its own sake, also serves as a useful check of the correctness of the expansions (27)–(29) in terms of single summations.

Conclusion

The above derivation of the generalized subcentric distribution and its detailed expression for the case of dispersion shows this distribution to be the most general formalism, so far published in the crystallographic literature, which accounts for deviations from ideal statistics, caused by factors other than dependence of the various atomic contributions to the structure factor. The generalized subcentric distribution can thus be regarded as the formal answer to problems involving dispersion (Wilson, 1980*a*), but it also constitutes a framework which accommodates various important special and limiting generalized distributions which have been investigated (Shmueli & Wilson, 1981; Shmueli, 1982; Srinivasan & Parthasarathy, 1976). It is possible that in actual applications the more specialized expressions may be more convenient to handle; however, the question of the practical aspects of this new distribution must await numerical tests of its various possible modes of application.

It is believed that the methods employed in this study may be conveniently and, hopefully, profitably applied to other areas of intensity and structure-factor statistics. Several such investigations are now in progress at the authors' laboratories, and one of them is concerned with the problem of combining the effects of counting statistics (Wilson, 1980*b*) and 'structural' intensity statistics (Shmueli & Wilson, 1981; Shmueli, 1982) in a single distribution function of the structure amplitude.

One of us (US) is grateful to Mrs Chana Sofer, Mrs Judith Kovetz, and to other members of the staff of the Tel-Aviv University Computation Center, for valuable assistance and advice concerning the implementation and use of the *REDUCE* system.

APPENDIX *A*
An algorithm for decomposition of summations

A recurring problem in the derivation of moments of sums of random variables is the decomposition of the various partial summations, to which the moments reduce, into manageable single summations (*cf.* Foster & Hargreaves, 1963; Wilson, 1978; Shmueli & Wilson, 1981; Shmueli, 1982). Although such a decomposition merely amounts to adding and subtracting equal terms, it may become very cumbersome where higher-order summations are to be treated and has certainly been a labour-determining factor in various studies. With the increasing availability of computer programs which can handle well defined algebraic manipulations, this tedious and error-prone process can be carried out reliably and most rapidly, and yield a result of analytical as well as computational significance.

For example, let us consider a fivefold summation of the form

$$\sum_{l\neq j\neq k\neq l\neq m}\sum\sum\sum\sum\sum x_i y_j z_k w_l v_m, \qquad (A1)$$

where each index has the same range of values but no two indices can assume the same value in any of the terms of (*A*1).

By adding and subtracting equal terms, the summation (*A*1) can be rewritten as

$$\sum_m \left\{\sum_{i\neq j\neq k\neq l}\sum\sum\sum\sum x_i y_j z_k w_l\right\} v_m - \sum_{i\neq j\neq k\neq m}\sum\sum\sum\sum [(x_i v_i) y_j z_k w_m - x_i (y_j v_j) z_k w_m - x_i y_j (z_k v_k) w_m - x_i y_j z_k (w_m v_m)], \qquad (A2)$$

where the index m in the first term in (*A*2) may be equal to i or j or k or l, while i, j, k and l are unequal. Clearly, the addition takes place in the first term and the subtraction in the remaining four. Each of the latter can be rewritten as above and so on, until the fivefold summation is completely decomposed into single summations.

This simple but lengthy process can be made systematic and concise, as shown below.

Defining an n-tuple summation of the form (*A*1) by

$$S_n(x_1, x_2, \ldots, x_n) \equiv \sum_{k_1\neq k_2}\sum \cdots \sum_{\neq k_n} x_{1k_1} x_{2k_2} \ldots x_{nk_n}, \qquad (A3)$$

where the indices $k_1, k_2, \ldots, k_n$ range over the same values but no two indices can be equal in any of the terms of (*A*3), an algorithm for the decomposition of such summations can be formulated as follows.

$$\begin{aligned} S_2(x_1,x_2) &= S_1(x_1)\, S_1(x_2) - S_1(x_1 x_2) \\ &\equiv \mathscr{S}[S_1(x_1); x_2] \qquad (A4) \end{aligned}$$

$$\begin{aligned} S_3(x_1,x_2,x_3) &= S_2(x_1,x_2)\, S_1(x_3) - S_2(x_1 x_3, x_2) \\ &\quad - S_2(x_1, x_2 x_3) \\ &\equiv \mathscr{S}[S_2(x_1,x_2); x_3] \qquad (A5) \end{aligned}$$

and, in general,

$$\begin{aligned} S_k(x_1, \ldots, x_k) &= S_{k-1}(x_1, \ldots, x_{k-1})\, S_1(x_k) \\ &\quad - S_{k-1}(x_1 x_k, x_2, \ldots, x_{k-1}) \\ &\quad - S_{k-1}(x_1, x_2 x_k, \ldots, x_{k-1}) \\ &\quad - \cdots - S_{k-1}(x_1, x_2, \ldots, x_{k-1} x_k) \\ &\equiv \mathscr{S}[S_{k-1}(x_1, \ldots, x_{k-1}); x_k]. \qquad (A6) \end{aligned}$$

A comparison of (*A*4)–(*A*6) with the more familiar (*A*2) shows immediately that the function of the operator $\mathscr{S}$ is to add and subtract equal terms while lowering the order of the summation involved by one. Quantities which appear as products carry the same index and the presence of k commas within a symbolic

summation means that there are $k + 1$ unequal indices (*cf.* Shmueli, 1982).

For example, the decomposition of (*A*1) into single summations can be symbolically written as follows:

$$\begin{aligned} S_5(x,y,z,w,v) &= \mathscr{S}[S_4(x,y,z,w); v] \\ &= \mathscr{S}[\mathscr{S}[S_3(x,y,z); w]; v] \\ &= \mathscr{S}[\mathscr{S}[\mathscr{S}[S_2(x,y); z]; w]; v] \\ &= \mathscr{S}[\mathscr{S}[\mathscr{S}[\mathscr{S}[S_1(x); y]; z]; w]; v]. \end{aligned} \tag{A7}$$

The above process can form the basis for a straightforward calculation but its real advantage lies in the adaptability to computer evaluation. In order to illustrate this feature, we show the input which is required by the *LISP*-based system *REDUCE* (Hearn, 1973) for the decomposition of summations, up to a fourfold one. The rather self-explanatory code reads:

```
OPERATOR S1, S2, S3, S4;
FOR ALL X,Y LET S2(X,Y) = S1(X) * S1(Y) -
  S1(X * Y);
FOR ALL X,Y,Z LET S3(X,Y,Z) = S2(X,Y) *
  S1(Z) - S2(X * Z,Y) - S2(X,Y * Z);
FOR ALL X,Y,Z,W LET S4(X,Y,Z,W) = S3(X,
  Y,Z) * S1(W) - S3(X * W,Y,Z) - S3(X,Y *
  W,Z) - S3(X,Y,Z * W);                    (A8)
```

Once the above 'operators' are defined, a submission of an expression, *e.g.*

```
24 * S4(A,A,A,A) + 72 * S4(A,A,B,BS)
   + 9 * S4(B,B,BS,BS)                     (A9)
```

results in its expansion in terms of single (S1) summations, the summands being products of A,B and BS and their powers. The expansion can also be obtained in the form of a usable Fortran listing which can serve its numerical evaluation. Of course, if any relationships exist between the quantities which appear in the summations, they can be specified and a further simplification of the result is obtained. Usual simplifications performed by *REDUCE* consist of collecting powers and like terms or their products.

APPENDIX *B*

The basic structure of the expressions derived in the text can be illustrated by the first few moments and polynomials involved. In order to obtain the orthogonal polynomials $f_1(z)$ and $f_2(z)$, associated with the asymptotic subcentric distribution (13), we require the moments c_n with $n \leq 4$ [*cf.* (8) and (19)]. These are given by

$$c_1 = 1 \tag{B1}$$

$$c_2 = (3x^2 - 1)/x^2 \tag{B2}$$

$$c_3 = [3(5x^2 - 3)]/x^2 \tag{B3}$$

$$c_4 = [3(35x^4 - 30x^2 + 3)]/x^4, \tag{B4}$$

where $x = \alpha^{1/2}$ and α is defined by (15). Of course, $c_0 = 1$ since the asymptotic distribution is normalized to unity. From (*B*1)–(*B*4) the required polynomials can be constructed from (8) and (11), but it will suffice to give the coefficients of the powers of z appearing in f_1 and f_2. We obtain

$$d_{10} = -1, \quad d_{11} = 1 \tag{B5}$$

$$d_{20} = (6x^4 - 3x^2 - 1)/x^4, \quad d_{21} = [4(-3x^2 + 2)]/x^2$$

and

$$d_{22} = (2x^2 - 1)/x^2, \tag{B6}$$

and the normalization constants follow, using (*B*1)–(*B*6) and (11), as

$$K_1 = x/(2x^2 - 1)^{1/2} \tag{B7}$$

and

$$K_2 = x^4/[2(24x^8 - 48x^6 + 36x^4 - 13x^2 + 2)^{1/2}]. \tag{B8}$$

It is readily seen that the moments (*B*1)–(*B*4) tend to $n!$ or $(2n-1)!!$, and the general polynomials

$$f_1(z) = K_1(d_{11}z + d_{10}) \tag{B9}$$

$$f_2(z) = K_2(d_{22}z^2 + d_{21}z + d_{20}) \tag{B10}$$

tend to $L_n(z)$ (not standardized) or

$$H_{2n}(z/2)^{1/2}/[2^n(2n)^{1/2}!]$$

according as α (or x) tend to unity or infinity, respectively. These limits of α correspond to the generalized acentric and dispersionless centric distributions (see text).

Regarding the expansion coefficients of the generalized distribution (3) [with $P^{(0)}(z)$ given by (13)], the coefficient g_1 vanishes for any x since $\langle z \rangle = 1$ by definition and $d_{11} = -d_{10} = 1$, while g_2 assumes the forms

$$(\langle z^2 \rangle - 2)/2 \tag{B11}$$

and

$$(\langle z^2 \rangle - 3)/[(4!)^{1/2}] \tag{B12}$$

in the acentric and dispersionless centric limits respectively (*cf.* Shmueli & Wilson, 1981). Analogous results follow from the polynomials of orders 3 and 4 in z (*cf.* deposited material).

References

ABRAMOWITZ, M. & STEGUN, I. (1972). *Handbook of Mathematical Functions*. New York: Dover.
BERTAUT, E. F. (1955). *Acta Cryst.* **8**, 823–832.
Computers in the New Laboratory – A Nature Survey (1981). *Nature*, **290**, 193–200.
CRAMÉR, H. (1951). *Mathematical Methods of Statistics*, section 12.6. Princeton Univ. Press.
FOSTER, F. & HARGREAVES, A. (1963). *Acta Cryst.* **16**, 1124–1133.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Table of Integrals, Series and Products*, entry: 6.682(4). New York: Academic Press.
HAUPTMAN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136–141.
HEARN, A. C. (1973). *REDUCE 2 User's Manual*. Univ. of Utah, Salt Lake City, Utah, USA.
KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131–135.
KLUG, A. (1958). *Acta Cryst.* **11**, 515–543.
LIPSON, H. & WOOLFSON, M. M. (1952). *Acta Cryst.* **5**, 680–682.
ROGERS, D. & WILSON, A. J. C. (1953). *Acta Cryst.* **6**, 439–449.
SHMUELI, U. (1979). *Acta Cryst.* A**35**, 282–286.
SHMUELI, U. (1982). *Acta Cryst.* A**38**, 362–371.
SHMUELI, U. & KALDOR, U. (1981). *Acta Cryst.* A**37**, 76–80.
SHMUELI, U. & WILSON, A. J. C. (1981). *Acta Cryst.* A**37**, 342–353.
SHMUELI, U. & WILSON, A. J. C. (1982). In *Crystallographic Statistics: Progress and Problems*, edited by S. RAMASESHAN, M. F. RICHARDSON & A. J. C. WILSON, pp. 83–97. Bangalore: Indian Academy of Science.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
SZEGÖ, G. (1939). *Orthogonal Polynomials*, sections 2.1–2.5. New York: American Mathematial Society.
WATSON, G. N. (1922). *Theory of Bessel Functions*. Cambridge Univ. Press.
WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151, 152.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
WILSON, A. J. C. (1952). *Research*, **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*a*). *Acta Cryst.* A**36**, 945–946.
WILSON, A. J. C. (1980*b*). *Acta Cryst.* A**36**, 929–936.

Acta Cryst. (1983). A**39**, 233–245

Further Properties of a Gaussian Model of Disorder

BY T. R. WELBERRY AND C. E. CARROLL

Research School of Chemistry, Australian National University, PO Box 4, *Canberra, ACT* 2600, *Australia*

(*Received* 16 *August* 1982; *accepted* 7 *October* 1982)

Abstract

A generalization of a previously described Gaussian growth-disorder model is described. The properties of this general model are discussed in relation to the more restricted but more easily simulated growth-disorder model. Optical diffraction patterns of realizations obtained by Monte Carlo procedures are presented for two possible applications of the model. The extra degree of freedom provided by the generalization enables a greater diversity of diffraction patterns to be achieved. In particular, it is possible to produce realizations having an approximately isotropic correlation field. The relationship between the Gaussian model and the Hosemann paracrystal is discussed.

1. Introduction

In previous papers (Welberry, 1977; Welberry, Miller & Pickard, 1979; Welberry, Miller & Carroll, 1980; Welberry & Carroll, 1982), we have described a series of stochastic models of disorder called 'growth-disorder models', which enable the rapid production of optical diffraction masks representing disordered lattices. Such diffraction masks have been the principal tool in a number of studies of disorder phenomena reported recently, including orientational disorder in molecular crystals (Welberry, Jones & Epstein, 1982), cation framework distortions in materials with ferroelectric properties (Welberry, 1982), and highly disordered lattices known as paracrystals which are used in the study of polymers and amorphous materials (Welberry, Miller & Carroll, 1980).

The efficacy of the growth-disorder models for the purpose of optical diffraction mask making relies on the simple and rapid growth algorithm which enables suitably large realizations of disordered lattices containing predetermined short-range-order properties to be produced. The disordered distributions which may be produced by this means are, however, not the most general possible on a given lattice, and often approximations to the desired distribution must be made. Access to realizations of more general distributions can only be obtained *via* lengthy Monte Carlo iterative procedures and for routine usage such methods are not feasible. In this paper we explore some aspects of the relationship between a growth-disorder model and its

0567-7394/83/020233-13$01.50