assistance provided by the members of the Electron Diffraction Section, in particular Mr A. F. Moodie and Dr D. F. Lynch. Many thanks are due to Mr J. G. Allpress and Dr J. V. Sanders for their helpful discussions and for critical reading of the manuscript.

References

ALLPRESS, J. G. (1969*a*). *J. Solid State Chem.* **1**, 66–81.
ALLPRESS, J. G. (1969*b*). *Mater. Res. Bull.* **4**, 707–720.
ALLPRESS, J. G. (1970). *J. Solid State Chem.* **2**, 78–93.
ALLPRESS, J. G., HEWAT, E. A., MOODIE, A. F. & SANDERS, J. V. (1972). *Acta Cryst.* A**28**, 528–536.
ALLPRESS, J. G., SANDERS, J. V. & WADSLEY, A. D. (1969). *Acta Cryst.* B**25**, 1156–1164.
ANDERSON, J. S., BROWNE, J. M. & HUTCHINSON, J. L. (1972*a*). *Nature, Lond.* **237**, 151–153.
ANDERSON, J. S., BROWNE, J. M. & HUTCHINSON, J. L. (1972*b*). *J. Solid State Chem.* **5**, 419–431.
ANDERSSON, S., MUMME, W. G. & WADSLEY, A. D. (1966). *Acta Cryst.* **21**, 802–808.
ANSTIS, G. R., LYNCH, D. F., MOODIE, A. F. & O'KEEFE, M. A. (1973). *Acta Cryst.* A**29**, 138–147.
COWLEY, J. M. & IIJIMA, S. (1972*a*). *Proc. 30th Ann. Electron Microsc. Soc. Amer.* Edited by C. J. ARCENEAUX, pp. 550–551.
COWLEY, J. M. & IIJIMA, S. (1972*b*). *Z. Naturforsch.* **27a**, 445–451.
COWLEY, J. M. & MOODIE, A. F. (1957*a*). *Acta Cryst.* **10**, 609–619.
COWLEY, J. M. & MOODIE, A. F. (1957*b*). *Proc. Phys. Soc.* B**70**, 486–496.
COWLEY, J. M. & MOODIE, A. F. (1960). *Proc. Phys. Soc.* **76**, 378–384.
COWLEY, J. M. & MOODIE, A. F. (1962), *J. Phys. Soc. Japan, Suppl. B*-II, **17**, 86–91.
EISENHANDLER, C. B. & SIEGEL, B. M. (1966). *J. Appl. Phys.* **37**, 1613–1620.
HEAD, A. K. (1967). *Aust. J. Phys.* **20**, 557–566.
HEWAT (née CHIDZEY), E. A. (1970). M. Sc. Thesis, Univ. of Melbourne.
HUTCHINSON, J. L. & ANDERSON, J. S. (1972). *Phys. Stat. Sol.* (*a*) **9**, 207–213.
IIJIMA, S. (1971). *J. Appl. Phys.* **42**, 5891–5893.
IIJIMA, S. (1973). *Acta Cryst.* A**29**, 18–24.
IIJIMA, S. & O'KEEFE, M. A. (1973). In preparation.
LYNCH, D. F. & O'KEEFE, M. A. (1972). *Acta Cryst.* A**28**, 536–548.
MOODIE, A. F. (1973). To be published.
ROTH, R. S. & WADSLEY, A. D. (1965). *Acta Cryst.* **18**, 724–730.
SCHERZER, O. (1949). *J. Appl. Phys.* **20**, 20–29.
UYEDA, N., KOBAYASHI, T., SUITO, F., HARADA, Y. & WATANABE, M. (1970). *Microsc. Electron.* **1**, 23–24.
VAN LANDUYT, J., VOCHTEN, R. & AMELINCKX, S. (1970). *Mater. Res. Bull.* **5**, 275–280.
WADSLEY, A. D. (1961). *Acta Cryst.* **14**, 664–670.

Acta Cryst. (1973). A**29**, 401

Exponential Form of Joint Probability Distribution; Unequal Atoms

BY J. KARLE AND R. D. GILARDI

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington D.C. 20375, *U.S.A.*

(*Received* 17 *October* 1972; *accepted* 8 *February* 1973)

The joint probability distributions for three structure factors whose subscripts add to zero, in the general case of unequal atoms, are expressed in an exponential form for space groups $P1$ and $P\bar{1}$. The latter are representative of noncentrosymmetric and centrosymmetric space groups respectively. The exponential form possesses considerably improved convergence properties over those of the standard asymptotic series, although it too remains asymptotic. With the range of values ordinarily obtained for the normalized structure-factor magnitudes, the exponential forms are quite accurate. However, the accuracy deteriorates somewhat as these magnitudes approach their largest possible values. By altering the exponential form with the use of a result from the inequality theory, joint probability distributions are obtained which are accurate over the entire range of values for the structure-factor magnitudes and are most accurate at the largest values. Several probability measures of interest are derived from the joint distribution functions such as expected values, variances and the probability that a structure factor has a positive sign. Numerical tests indicate that the derived probability measures are very reliable and that their validity extends to higher space groups than $P1$ and $P\bar{1}$.

Introduction

The exponential form of the joint probability distribution for three structure factors ($\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$) in the equal-atom case has been derived for space groups $P1$ and $P\bar{1}$ (Karle, 1972). The virtue of this form is the considerably improved convergence properties of the exponential series over those for the ordinary series expansion (Bertaut 1960*a*, *b*; Karle, 1972). It is of interest to consider whether exponential series possessing the improved convergence properties can be obtained for the joint probability distributions in the case of unequal atoms. It is found that this is the case and, as occurs for equal atoms, terms of the order $N^{-m/2}$ in the exponen-

tial series, where N is the number of atoms in the unit cell, are associated with polynomials in the structure-factor magnitudes whose highest degree is $m+2$. This reduces the degree of the polynomial, as compared to the standard series expansion, by the factor $|E|^{2m-2}$, where $|E|$ represents a normalized structure-factor magnitude. The difference between the joint distributions for the equal and unequal-atom cases appears only in terms of functions of the atomic numbers which replace N, the number of atoms in the unit cell, in the formulas.

A variety of formalisms exist for expressing the joint probability distributions of complex-valued structure factors for the noncentrosymmetric case in terms of series expansions employing the associated Laguerre polynomials (Hauptman & Karle, 1953*b*; Bertaut, 1956; Karle & Hauptman, 1956; Naya, Nitta & Oda, 1965). Similarly, formalisms exist for forming the joint probability distributions of structure factors for the centrosymmetric case in terms of Hermite polynomials (Hauptman & Karle, 1953*a*; Bertaut, 1955; Klug, 1958; Naya, Nitta & Oda, 1964). The series representation of the joint distribution is written so that successive terms have increasing reciprocal powers of $N^{1/2}$. Naya, Nitta & Oda have carried the series out to the $N^{-5/2}$ term. These expressions form a convenient starting point for transforming the joint distributions to the exponential form. The transformation is accomplished by using a simple, but somewhat tedious, algorithm which assures that the exponential form, when expanded, results in the original series composed of associated Laguerre or Hermite polynomials, as the case may be. Considerable reduction in the tedium of the manipulations was accomplished by use of a computer program for multiplying and combining terms in the polynomial expressions.

In a previous paper concerning the exponential form for the equal-atom case (Karle, 1972), the results were modified by the application of inequality theory which assured accuracy and convergence no matter how closely the normalized structure-factor magnitudes approached their maximum value [$N^{1/2}$ for the equal atom case, $\sigma_1/\sigma_2^{1/2}$ for unequal atoms; see, for example, equations (15) and (32) of the previous reference]. Similar considerations and test calculations will be presented in this paper.

Exponential form for joint distribution

Noncentrosymmetric reflections ($\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$)

The joint probability distribution function for three complex-valued normalized structure factors, $E_1=E_{\mathbf{h}_1}$, $E_2=E_{\mathbf{h}_2}$, $E_3=E_{\mathbf{h}_3}$, $(\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0)$, on the assumption that the atomic coordinates are random variables which are uniformly and independently distributed, is considered for the general case of unequal atoms in space group $P1$. The joint distribution function can be written in the exponential form, valid to terms of the order $N^{-5/2}$,

$$
\begin{aligned}
&P(|E_1|,|E_2|,|E_3|,\varphi_1,\varphi_2,\varphi_3)\\
&=(1/\pi^3)|E_1E_2E_3|\exp(-|E_1|^2-|E_2|^2-|E_3|^2)\\
&\times\exp\left(|E_1E_2E_3|\cos(\varphi_1+\varphi_2+\varphi_3)\right.\\
&\times\{2Q_3+(|E_1|^2+\text{cyc.})[-Q_5+Q_3Q_4+2Q_3^3]\\
&+[6Q_5-6Q_3Q_4-6Q_3^3]\\
&+(|E_1|^4+\text{cyc.})[(\tfrac{2}{3})Q_7-(\tfrac{2}{3})Q_3Q_6-(\tfrac{3}{2})Q_4Q_5-3Q_3^2Q_5\\
&+(\tfrac{3}{2})Q_3Q_4^2+3Q_3^3Q_4+2Q_3^5]\\
&+(|E_1E_2|^2+\text{cyc.})[(\tfrac{1}{2})Q_7+(\tfrac{7}{2})Q_3Q_6-Q_4Q_5-17Q_3^2Q_5\\
&+(\tfrac{1}{2})Q_3Q_4^2+9Q_3^3Q_4+10Q_3^5]\\
&+(|E_1|^2+\text{cyc.})[-6Q_7-5Q_3Q_6+(\tfrac{21}{2})Q_4Q_5+49Q_3^2Q_5\\
&-(\tfrac{17}{2})Q_3Q_4^2-29Q_3^3Q_4-20Q_3^5]\\
&+[18Q_7+8Q_3Q_6-27Q_4Q_5-90Q_3^2Q_5+21Q_3Q_4^2\\
&\left.+48Q_3^3Q_4+26Q_3^5]\}\right.\\
&+|E_1E_2E_3|^2\cos 2(\varphi_1+\varphi_2+\varphi_3)[-(\tfrac{3}{4})Q_6+3Q_3Q_5\\
&-(\tfrac{3}{2})Q_3^2Q_4-2Q_3^4]-(|E_1|^4+\text{cyc.})[(\tfrac{1}{4})Q_4]-(|E_1E_2|^2\\
&+\text{cyc.})Q_3^2+(|E_1|^2+\text{cyc.})[Q_4+Q_3^2]-[(\tfrac{3}{2})Q_4+Q_3^2]\\
&+(|E_1|^6+\text{cyc.})[(\tfrac{1}{6})Q_6-(\tfrac{1}{4})Q_4^2]+(|E_1|^4|E_2|^2+|E_1|^2|E_2|^4\\
&+\text{cyc.})[Q_3Q_5-Q_3^2Q_4-Q_3^4]\\
&+|E_1E_2E_3|^2[-Q_6+6Q_3Q_5-3Q_3^2Q_4-6Q_3^4]\\
&+(|E_1|^4+\text{cyc.})[-Q_6-Q_3Q_5+(\tfrac{13}{8})Q_4^2+Q_3^2Q_4+(\tfrac{1}{2})Q_3^4]\\
&+(|E_1E_2|^2+\text{cyc.})[Q_6-10Q_3Q_5+7Q_3^2Q_4+6Q_3^4]\\
&+(|E_1|^2+\text{cyc.})[Q_6+8Q_3Q_5-(\tfrac{5}{2})Q_4^2-5Q_3^2Q_4-3Q_3^4]\\
&-[Q_6+6Q_3Q_5-(\tfrac{15}{8})Q_4^2-3Q_3^2Q_4-(\tfrac{3}{2})Q_3^4]\\
&+\cdots) \qquad (1)
\end{aligned}
$$

where

$$E_{\mathbf{h}}=|E_{\mathbf{h}}|\exp(i\varphi_{\mathbf{h}}) \qquad (2)$$

$$Q_n=\sigma_n/\sigma_2^{n/2} \qquad (3)$$

$$\sigma_n=\sum_{j=1}^{N}Z_j^n \qquad (4)$$

and Z_j is the atomic number of the jth atom.

By expanding the second exponential on the right side of equation (1), a series expansion is obtained which is identical to equation (67) of Naya, Nitta & Oda (1965) if Q_n is substituted for Z_n (the coefficient of the entire Z_4Z_5 term should be $\frac{1}{4}$ and that for the second expression of the $Z_3^2Z_5$ term should be $\frac{3}{2}$). The initial factor $|E_1E_2E_3|$ of equation (1) is understood to be inserted into the latter equation. In this notation the Q_n replace the Z_n of Naya *et al.* to avoid confusion with the atomic numbers appearing in equations (3) and (4). Although equation (4) is a common approximation, alternative definitions of the Q_n might be considered, since normalized structure factors for experimental data are derived from structure factors having variable scattering factors. One alternative definition is,

$$Q_n=\langle Q_n(\mathbf{h})\rangle_{\mathbf{h}}=(\tfrac{1}{3})\sum_{\mathbf{h}}[s_n(\mathbf{h})/s_2^{n/2}(\mathbf{h})] \qquad (5)$$

where $\mathbf{h}$ ranges over the three values $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$,

$$s_n(\mathbf{h})=\sum_{j=1}^{N}f_j^n(\mathbf{h}) \qquad (6)$$

and $f_j(\mathbf{h})$ is the atomic scattering factor for the jth atom. In the case of equal atoms, $Q_n = N^{-(n-2)/2}$. The substitution of these quantities into equation (1) will lead to the joint probability distribution derived previously [Karle, 1972, equation (1)].

Several results appearing in the latter reference have correspondences in the unequal atom case which may be obtained by use of equation (1). The polynomial appearing under the braces in (1) is represented by p_u, giving,

$$p_u = p_u(|E_1|,|E_2|,|E_3|) = \{2Q_3 + (|E_1|^2 + \text{cyc.})[-Q_5 + Q_3Q_4 + 2Q_3^3] + \ldots + [18Q_7 + 8Q_3Q_6 - 27Q_4Q_5 - 90Q_3^2Q_5 + 21Q_3Q_4^2 + 48Q_3^3Q_4 + 26Q_3^5]\} . \quad (7)$$

With this definition, the probability distribution of the sum

$$\Phi_{123} = \varphi_1 + \varphi_2 + \varphi_3 \quad (8)$$

given the values of $|E_1|$, $|E_2|$ and $|E_3|$ is

$$P_1(\Phi_{123};\ |E_1|,|E_2|,|E_3|) \simeq K_1 \exp\left[|E_1E_2E_3| p_u \cos \Phi_{123} + |E_1E_2E_3|^2 v \cos 2\Phi_{123}\right] \quad (9)$$

where

$$v = -(\tfrac{3}{4})Q_6 + 3Q_3Q_5 - (\tfrac{3}{2})Q_3^2Q_4 - 2Q_3^4 , \quad (10)$$

the normalizing constant K_1 is

$$K_1 = [2\pi I_0(|E_1E_2E_3| p_u) + 2\pi |E_1E_2E_3|^2 v I_2(|E_1E_2E_3| p_u)]^{-1} \quad (11)$$

and the I_n are Bessel functions of imaginary argument. The expected value for $\cos \Phi_{123}$, determined from equation (9) is

$$\langle \cos \Phi_{123} \rangle \simeq \frac{(1 + \tfrac{1}{2}|E_1E_2E_3|^2 v)\dfrac{I_1(w_u)}{I_0(w_u)} + \tfrac{1}{2}|E_1E_2E_3|^2 v \dfrac{I_3(w_u)}{I_0(w_u)}}{1 + |E_1E_2E_3|^2 v \dfrac{I_2(w_u)}{I_0(w_u)}} \quad (12)$$

where

$$w_u = |E_1E_2E_3| p_u . \quad (13)$$

The probability distribution function (9) should afford accurate values for the expected values of the cosine invariants, $\cos(\varphi_1 + \varphi_2 + \varphi_3)$, for the magnitudes of the normalized structure factors usually obtained from experiment. It will be seen, in fact, that optimal results are for the most part obtained by using p_u out to the $N^{-3/2}$ term and ignoring the $N^{-5/2}$ term. Nevertheless, the maximum value for $|E|$ is $\sigma_1/\sigma_2^{1/2}$ and although this large value is rarely approached, at such an extreme value the accuracy of the distributions (1) and (9) deteriorates. In the paper concerning the equal-atom case (Karle, 1972), use was made of inequality theory to alter the distribution (9) with a function which very closely approximates (9) in the range of the normally observed values for the $|E|$ and which, in addition, is accurate at the largest values for $|E|$. A similar modification can be applied to the case of unequal atoms. In a purely empirical fashion, and in analogy with the form of the corresponding joint probability distribution for equal atoms [Karle, 1972, equation (15)], the suggested probability distribution, applicable over the full range of $|E|$ values and, in particular, the largest ones, is

$$P_2(\Phi_{123};|E_1|,|E_2|,|E_3|) \simeq K_2 \exp\left[\frac{2\sigma_3|E_1E_2E_3|}{\sigma_2^{3/2}q_1}\cos \Phi_{123}\right] \quad (14)$$

where

$$K_2 = \left[2\pi I_0\left(\frac{2\sigma_3|E_1E_2E_3|}{\sigma_2^{3/2}q_1}\right)\right]^{-1} \quad (15)$$

$$q_1 = 1 + 2|U_1U_2U_3| - |U_1|^2 - |U_2|^2 - |U_3|^2 \quad (16)$$

and the unitary structure factor, U, is defined by

$$U = E\sigma_2^{1/2}/\sigma_1 . \quad (17)$$

For equal atoms, $Q_3 = \sigma_3/\sigma_2^{3/2} = N^{-1/2}$ and also $\sigma_2^{1/2}/\sigma_1 = N^{-1/2}$.

The expected value of the cosine invariant may be obtained from P_2 giving

$$\langle \cos(\varphi_1 + \varphi_2 + \varphi_3) \rangle \simeq \frac{I_1\left(\dfrac{2\sigma_3|E_1E_2E_3|}{\sigma_2^{3/2}q_1}\right)}{I_0\left(\dfrac{2\sigma_3|E_1E_2E_3|}{\sigma_2^{3/2}q_1}\right)} . \quad (18)$$

As discussed previously (Karle, 1972), the following probability distributions are equal,

$$P_2(\Phi_{123};\ |E_1|,|E_2|,|E_3|) = P_3(\varphi_1;\ \varphi_2,\varphi_3,|E_1|,|E_2|,|E_3|) = P_4(\varphi_1+\varphi_2;\ \varphi_3,|E_1|,|E_2|,|E_3|) \quad (19)$$

each having K_2 as a normalizing constant. The probability distribution for φ_1, given many sets of $\varphi_2, \varphi_3, |E_2|, |E_3|$ satisfying $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$ can be obtained by multiplying the individual distributions of type P_3 together and renormalizing. The resulting distribution is estimated to be a good approximation, but it is not an exact result since the individual distributions P_3 are not completely independent. Equations (3.25) and (3.33)* of Karle & Karle (1966) and additional mathematical expressions in §§ 3·2, 3·3 and 3·4 can be altered to be consistent with the results of this paper by redefining the symbol κ of equation (3·20) by dividing by q_1 to give

$$\kappa(\mathbf{h},\mathbf{k}) = 2\sigma_3\sigma_2^{-3/2}|E_\mathbf{h}E_\mathbf{k}E_{\mathbf{h}-\mathbf{k}}|/q_1 \quad (20)$$

where $\mathbf{h}_1 = \mathbf{h}$, $\mathbf{h}_2 = -\mathbf{k}$ and $\mathbf{h}_3 = -\mathbf{h} + \mathbf{k}$. The above referenced equation (3·33) expresses the variance of φ_1 given fixed, specified sets of $\varphi_2, \varphi_3, |E_2|, |E_3|$ satisfying $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$.

An expression for the variance of $\Phi_{123}(-\pi < \Phi_{123} \leq \pi)$ can be readily obtained from equation (14). The expected value of Φ_{123} is zero and the variance is given by

$$\text{Variance of } \Phi_{123} = \frac{\pi^2}{3} + [I_0(\kappa)]^{-1} \sum_{n=1}^{\infty} \frac{I_{2n}(\kappa)}{n^2} - 4[I_0(\kappa)]^{-1} \sum_{n=0}^{\infty} \frac{I_{2n+1}(\kappa)}{(2n+1)^2} . \quad (21)$$

* Sign before last term of equation (3.33) should be minus.

This formula is the same as the variance formula for φ_1, when there is only one member in the given set of fixed, specified $\varphi_2, \varphi_3, |E_2|, |E_3|$.

Another variance of interest is the variance of the cosine invariant, $\cos \Phi_{123}$, which can be readily obtained from equation (14). The result is

$$\text{Variance of } \cos \Phi_{123} = 1 - \frac{I_1(\kappa)}{\kappa I_0(\kappa)} - \frac{I_1^2(\kappa)}{I_0^2(\kappa)} \qquad (22)$$

where $\kappa = \kappa(\mathbf{h}, \mathbf{k})$ as defined in equation (20) or, alternatively, q_1^{-1} may be replaced in equation (20) by p_u defined by equation (7). The numerical tests will show that p_u carried out to the $N^{-3/2}$ term, as given by equation (30), affords very accurate results. The form of this variance formula is the same as one derived previously (Fisher, Hancock & Hauptman, 1970) for the case of equal atoms. Equation (22) differs from the latter expression in that q_1 replaces unity and $\sigma_3/\sigma^{3/2}$ replaces $N^{-1/2}$ for equal atoms.

Centrosymmetric reflections ($\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$)

The joint probability distribution function for three normalized structure factors, $E_1 = E_{\mathbf{h}_1}$, $E_2 = E_{\mathbf{h}_2}$, $E_3 = E_{\mathbf{h}_3}$, ($\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$), on the assumption that the atomic coordinates are random variables which are uniformly and independently distributed, is considered for the general case of unequal atoms in space group $P\bar{1}$. This function can be written in the exponential form, valid to terms of the order $N^{-5/2}$,

$$\begin{aligned}
P(E_1, E_2, E_3) &= (\tfrac{1}{2}\pi)^{3/2} \exp\left[-(\tfrac{1}{2})\,(E_1^2 + E_2^2 + E_3^2)\right] \\
&\times \exp\,(E_1E_2E_3\{Q_3 + (E_1^2 + \text{cyc.})[-(\tfrac{1}{2})Q_5 + (\tfrac{1}{2})Q_3Q_4 \\
&+ Q_3^3] + [(\tfrac{9}{2})Q_5 - (\tfrac{9}{2})Q_3Q_4 - 4Q_3^3] \\
&+ (E_1^4 + \text{cyc.})[(\tfrac{1}{3})Q_7 - (\tfrac{1}{3})Q_3Q_6 - (\tfrac{3}{4})Q_4Q_5 - (\tfrac{3}{2})Q_3^2Q_5 \\
&+ (\tfrac{3}{4})Q_3Q_4^2 + (\tfrac{3}{2})Q_3^3Q_4 + Q_3^5] \\
&+ (E_1^2E_2^2 + \text{cyc.})[(\tfrac{1}{4})Q_7 + (\tfrac{7}{4})Q_3Q_6 - (\tfrac{1}{2})Q_4Q_5 - (\tfrac{17}{2})Q_3^2Q_5 \\
&+ (\tfrac{1}{4})Q_3Q_4^2 + (\tfrac{9}{2})Q_3^3Q_4 + 5Q_3^5] \\
&+ (E_1^2 + \text{cyc.})[-(\tfrac{29}{6})Q_7 - (\tfrac{11}{3})Q_3Q_6 + (\tfrac{33}{4})Q_4Q_5 \\
&+ (\tfrac{73}{2})Q_3^2Q_5 - (\tfrac{27}{4})Q_3Q_4^2 - (\tfrac{43}{2})Q_3^3Q_4 - 14Q_3^5] \\
&+ [(\tfrac{87}{4})Q_7 + (\tfrac{31}{4})Q_3Q_6 - (\tfrac{63}{2})Q_4Q_5 - (\tfrac{189}{2})Q_3^2Q_5 \\
&+ (\tfrac{99}{4})Q_3Q_4^2 + (\tfrac{99}{2})Q_3^3Q_4 + 25Q_3^5]\} \\
&- (E_1^4 + \text{cyc.})[(\tfrac{1}{8})Q_4] - (E_1^2E_2^2 + \text{cyc.})[(\tfrac{1}{2})Q_3^2] \\
&+ (E_1^2 + \text{cyc.})[(\tfrac{3}{4})Q_4 + (\tfrac{1}{2})Q_3^2] - [(\tfrac{9}{8})Q_4 + (\tfrac{1}{2})Q_3^2] \\
&+ (E_1^6 + \text{cyc.})[(\tfrac{1}{18})Q_6 - (\tfrac{1}{8})Q_4^2] \\
&+ (E_1^4E_2^2 + E_1^2E_2^4 + \text{cyc.})[(\tfrac{1}{2})Q_3Q_5 - (\tfrac{1}{2})Q_3^2Q_4 - (\tfrac{1}{2})Q_3^4] \\
&+ E_1^2E_2^2E_3^2[-(\tfrac{7}{8})Q_6 + (\tfrac{9}{2})Q_3Q_5 - (\tfrac{9}{4})Q_3^2Q_4 - 4Q_3^4] \\
&+ (E_1^4 + \text{cyc.})[-(\tfrac{5}{6})Q_6 - (\tfrac{1}{2})Q_3Q_5 + (\tfrac{21}{16})Q_4^2 + (\tfrac{1}{2})Q_3^2Q_4 \\
&+ (\tfrac{1}{4})Q_3^4] + (E_1^2E_2^2 + \text{cyc.})[(\tfrac{7}{8})Q_6 - (\tfrac{15}{2})Q_3Q_5 + (\tfrac{21}{4})Q_3^2Q_4 \\
&+ 4Q_3^4] + (E_1^2 + \text{cyc.})[(\tfrac{13}{8})Q_6 + 6Q_3Q_5 - 3Q_4^2 - (\tfrac{15}{4})Q_3^2Q_4 \\
&- 2Q_3^4] - [(\tfrac{13}{8})Q_6 + (\tfrac{9}{2})Q_3Q_5 - (\tfrac{9}{4})Q_4^2 - (\tfrac{9}{4})Q_3^2Q_4 - Q_3^4] \\
&+ \ldots). \qquad (23)
\end{aligned}$$

Expansion of the second exponential in equation (23) gives a series expansion that is identical with equation (III-2) of Naya, Nitta & Oda (1964). The substitution of $Q_n = N^{-(n-2)/2}$ into (23) for the case of equal atoms gives the joint probability distribution derived previously [Karle, 1972, equation (22)]. From equation (23), and a procedure followed in the latter reference, it is possible to derive the probability that the sign of E_1 is positive given $|E_1|$, E_2 and E_3, obtaining

$$P_+(E_1; |E_1|, E_2, E_3) = \tfrac{1}{2} + \tfrac{1}{2} \tanh p_{1u} |E_1| E_2 E_3 \qquad (24)$$

where p_{1u} is the polynomial appearing within the braces in (23),

$$\begin{aligned}
p_{1u} = \{&Q_3 + (E_1^2 + \text{cyc.})[-(\tfrac{1}{2})Q_5 + (\tfrac{1}{2})Q_3Q_4 + Q_3^3] \\
&+ \ldots + [(\tfrac{87}{4})Q_7 + \ldots + (\tfrac{99}{2})Q_3^3Q_4 + 25Q_3^5]\}. \qquad (25)
\end{aligned}$$

As with the polynomial p_u for noncentrosymmetric crystals, optimal results are probably obtained from p_{1u} when this polynomial is carried to terms associated with $N^{-3/2}$.

It is suggested that good accuracy may be obtained over the full range of possible values for the $|E|$ if $Q_3 q_1^{-1}$, defined by equation (16), were to replace p_{1u} in equation (24), giving

$$P_+(E_1; |E_1|, E_2, E_3) = \tfrac{1}{2} + \tfrac{1}{2} \tanh \frac{\sigma_3 |E_1| E_2 E_3}{\sigma_2^{3/2} q_1}. \qquad (26)$$

Insight into the basis for this alteration may be obtained by referring to the previous paper concerning probability distributions for the case of equal atoms (Karle, 1972).

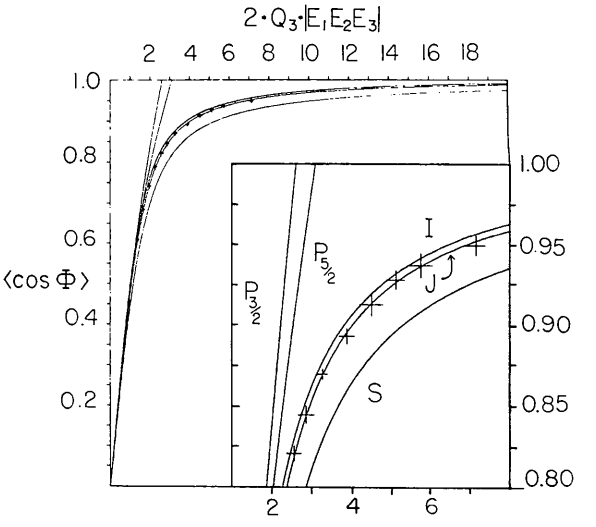

Fig. 1. Variation of expected values of the cosine invariants with $2|E_1E_2E_3|\sigma_3/\sigma_2^{3/2}$ for *N*-acetyl neuraminic acid (space group $P2_1$). The crosses represent values computed experimentally from a large number of invariants and, in the enlarged portions, the arms of the crosses measured from the crossover point represent three standard deviations. The solid curves represent the variation of several theoretical formulas obtained from the exponential form of the joint probability distribution carried to the $N^{-1/2}$ term (*S*), equation (29), the $N^{-3/2}$ term (*J*), equation (30), the modification by means of inequality theory (*I*), equation (31) and also from the standard series form of the joint distribution carried to the $N^{-3/2}$ term ($P_{3/2}$) and to the $N^{-5/2}$ term ($P_{5/2}$).

When there are several sets of known $E_{\mathbf{h}_2}$ and $E_{\mathbf{h}_3}$ such that $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$, and it is assumed that the individual measures of the probability are independent to a good approximation, equation (24) becomes

$$P_+(E_{\mathbf{h}})=\tfrac{1}{2}+\tfrac{1}{2}\tanh\frac{\sigma_3|E_{\mathbf{h}}|\sum_{\mathbf{k}}[E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}/q_1(\mathbf{h},\mathbf{k})]}{\sigma_2^{3/2}} \quad (27)$$

where $\mathbf{h}_1=\mathbf{h}$, $\mathbf{h}_2=\mathbf{k}$ and $\mathbf{h}_3=\mathbf{h}-\mathbf{k}$. The hyperbolic tangent formula (27) may be compared to the one commonly used (Woolfson, 1954; Cochran & Woolfson 1955). If q_1 is replaced by one, the latter is obtained. Since $0\le q_1\le 1$ equations (26) and (27) would lead to higher measures of the probability, particularly when the formulas employ the larger $|E|$ values.

Numerical tests

A statistical quantity of particular interest is the expected value of a cosine invariant, $\langle\cos(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{\mathbf{h}_3})\rangle$ where $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$. The joint probability distributions presented in this paper offer several measures of this quantity. For example, several formulas accrue from equation (12) depending upon the order of the term retained for p_u and whether the polynomial v is included or not. An additional relation formulated by use of inequality theory is given by equation (18).

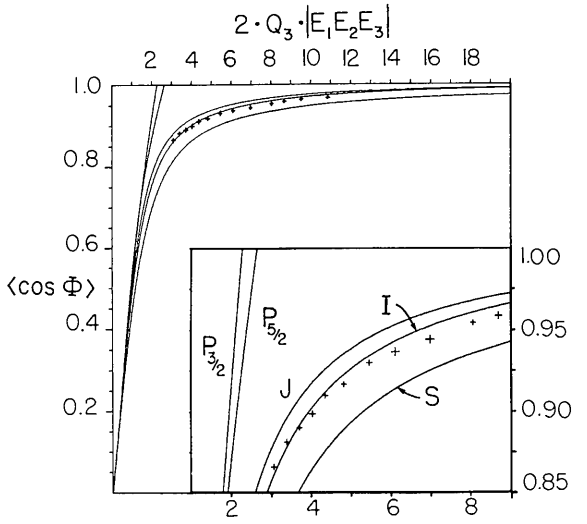

Fig. 2. Variation of expected values of the cosine invariants with $2|E_1E_2E_3|\sigma_3/\sigma_2^{3/2}$ for a bromine substitution of *N*-acetyl neuraminic acid (space group $P2_1$). The crosses represent values computed experimentally from a large number of invariants and, in the enlarged portions, the arms of the crosses measured from the crossover point represent three standard deviations. The solid curves represent the variation of several theoretical formulas obtained from the exponential form of the joint probability distribution carried to the $N^{-1/2}$ term (*S*), equation (29), the $N^{-3/2}$ term (*J*), equation (30), the modification by means of inequality theory (*I*), equation (31) and also from the standard series form of the joint distribution carried to the $N^{-3/2}$ term ($P_{3/2}$) and to the $N^{-5/2}$ term ($P_{5/2}$).

Owing to the statistical character of an expected value, an adequate evaluation of the theory requires comparison with a large amount of crystallographic data. To assure an adequate sample, a very large number of structure factors were computed for the crystal models which were employed. Of interest is the determination of the best form of the probability theory to be used to compute the expected values and its applicability to the higher space groups.

Numerical tests were performed in space groups $P2_1$, $P2_12_12_1$ and $P4_12_12$. For space group $P2_1$, the structure of *N*-acetyl neuraminic acid (Flippen, 1973*a*) was used. The structure contains non-hydrogen atoms of almost equal atomic number ($C_{11}H_{19}NO_9.2H_2O$). In order to test the case of unequal atoms, one of the oxygen atoms in the *N*-acetyl neuraminic acid structure was replaced by a sulfur atom and in another case a hydroxyl group was replaced by a bromine atom. For space group $P2_12_12_1$, the structure of *cis*-thymine glycol (Flippen, 1973*b*) was used as a test structure. This compound has the molecular formula $C_5H_8N_2O_4$. For space group $P4_12_12$, the structure of 3,4-dehydroproline anhydride ($C_{10}H_{10}N_2O_2$) was used (Karle, 1973). In an additional test, the oxygen atoms were replaced by sulfur atoms. Figs. 1 and 2 and Table 1 characterize the results.

Figs. 1 and 2 show how various theoretical measures of the expected values $\langle\cos\Phi_{123}\rangle$ vary with $2Q_3|E_1E_2E_3|$ where $Q_3=\sigma_3/\sigma_2^{3/2}$. The calculations include expected values based on the standard series form of the joint distribution carried out to the $N^{-3/2}$ term and the $N^{-5/2}$ term, the exponential form carried out to the $N^{-1/2}$ term (*S*) and the $N^{-3/2}$ term (*J*) and also the form based on inequality theory (*I*), equation (18). The latter three can be summarized as

$$\langle\cos\Phi_{123}\rangle=I_1(t|E_1E_2E_3|)/I_0(t|E_1E_2E_3|) \quad (28)$$

where

$$t=2Q_3 \quad (29)$$

or

$$t=2Q_3+(|E_1|^2+\text{cyc.})(-Q_5+Q_3Q_4+2Q_3^3)+6(Q_5-Q_3Q_4-Q_3^3) \quad (30)$$

or

$$t=2Q_3/q_1 \quad (31)$$

respectively. Fig. 1 concerns the essentially equal-atom case of *N*-acetyl neuraminic acid and Fig. 2 concerns the case in which a hydroxyl group is replaced with a bromine atom. Experimental values computed from large numbers of invariants (see Table 1) are indicated by the crosses. It was convenient in calculating the functions illustrated in Figs. 1 and 2 to impose a constraint on the relative values of $|E_1|$, $|E_2|$, and $|E_3|$, namely, $|E_1|=|E_2|=|E_3|=|E_1E_2E_3|^{1/3}$. It can be shown however, that this constraint leads to results that are consistent with a realistic sample taken from an experimental distribution of $|E|$ magnitudes. The statistical quantities $\langle\cos\varphi\rangle_J$ and $\langle\cos\varphi\rangle_I$ in Table 1 are averages of many individual calculations on such $|E_1|$, $|E_2|$, $|E_3|$ triples; no constraint was imposed, yet all of these

values agree with the values of the curves J and I of Figs. 1 and 2 to better than 1%.

Several results are apparent. The series form of the joint distribution is suitable only at small values of the argument and deteriorates rapidly as the magnitudes of the $|E|$ increase. The exponential form of the joint distribution carried out to the $N^{-3/2}$ term as given by (30) affords a very good fit for the equal-atom case and the inequality theory is only slightly high. The often employed exponential form carried out only to the $N^{-1/2}$ term, as given by (29), results in expected values which are several percentages too low. The same results were obtained for the sulfur-substituted *N*-acetyl neuraminic acid. Essentially the same results were also found for the bromine-substituted *N*-acetyl neuraminic acid, Fig. 2. In this case the inequality form (31) is somewhat closer to the experimental calculations than the exponential form out to the $N^{-3/2}$ term (30). However the exponential form out to the $N^{-1/2}$ term is still significantly too low. It also turns out that including the $N^{-5/2}$ term in the exponential form does not improve the accuracy of the agreement and, in fact, causes a severe divergence as the magnitude of the $|E|$ increase.

Examples of the details of the numerical comparisons, some of which are included in Figs. 1 and 2, are shown in Table 1. The expected values for the cosine invariants and the corresponding variances, equation (22), to be associated with an individual invariant were computed from the crystal models and from several theoretical formulas. The subscript Exp implies that values were obtained directly from the large number of cosine invariants computed from the crystal data and the subscripts S, J and I labeling the expected values refer to equation (28) associated with equations (29), (30) and (31) respectively. If the argument κ of the variance formula (22) were to be replaced by $t|E_1E_2E_3|$, the definitions (29), (30) and (31) would again correspond to subscripts S, J and I respectively, as applied to the calculations of the variances. The ranges of values for the function $2Q_3|E_1E_2E_3|$ covered by each calculation are shown in the first row of Table 1 and the number of invariants entering into each such range is shown in the second row. The last row presents measures of the standard deviations to be associated with the expected values obtained from the large numbers of the cosine invariants. It is seen that in many cases agreement between the experimental test calculation and theory is within two or three standard deviations. The bromine substituent shows greater deviations. The good agreement with the empirical calculations and equations (30) and (31) shown in Figs. 1 and 2 is thus further substantiated by Table 1. As might be expected, the variances of individual cosine invariants show the same pattern of agreement.

Extensive tests have not been made on the theoretical expressions obtained in this paper for centrosymmetric crystals. However, some test calculations were performed for a molecule containing two moderately heavy atoms ($C_{31}H_{25}ClN_3O_4P$) which crystallizes in space group $P\bar{1}$ (Gilardi & Karle, 1972). Experimental values of $P_+(E_1\,;|E_1|,|E_2|,|E_3|)$ were obtained by examining over 50000 structure invariants and sorting them into intervals of similar $|E_1E_2E_3|$ values. The statistical values were then compared with those predicted from the theoretical expressions (24) and (26) and also with the commonly used hyperbolic tangent formula, equation (26) with q_1 replaced by unity.

The agreement found is similar to that obtained in the numerical tests with noncentrosymmetric crystals. It was found that the commonly used hyperbolic tangent formula consistently underestimates P_+, whereas the predictions from equations (24) and (26) fit the experimental values quite well over the entire range of values for $|E_1E_2E_3|$. Equation (26), for example, will yield higher estimates for P_+ because of $q_1 < 1$ when the $|E|$ of interest differ from zero. Additional evidence has been presented by Tsoucaris (1970, p. 496) indicating that the standard formula for P_+ (q_1 replaced by unity) underestimates this probability measure and a correction factor was suggested.

Table 1. *Expected values* $\langle\cos\Phi\rangle_{\rm Exp}$ *for the cosine invariants computed for N-acetyl neuraminic acid and substitutions from a large number of invariants are compared with theoretical estimates*

The theoretical estimates were obtained from the exponential form of the joint probability distribution carried to the $N^{-1/2}$ term (S), equation (29), the $N^{-3/2}$ term (J), equation (30), and the modification by means of inequality theory (I), equation (31). The range of values for $2|E_1E_2E_3|\sigma_3/\sigma_2^{3/2}$ is shown in the first row and the number of invariants included in this range is shown in the second row. The variances of individual cosine invariants are also listed as well as the value of one standard deviation (last row) of the calculation of $\langle\cos\Phi\rangle_{\rm Exp}$.

	$2(C_{11}H_{19}NO_9.2H_2O)$				$2(C_{11}H_{19}NO_8S.2H_2O)$				$2(C_{11}H_{18}NO_8Br.2H_2O)$			
$2Q_3\|E_1E_2E_3\|$	1·21–1·52	2·43–2·73	3·64–4·25	5·46–6·06	1·55–1·72	2·41–2·75	3·44–4·13	> 6·87	2·91–3·23	3·88–4·20	5·81–6·46	9·04–10·34
No. of Invariants	91588	26469	8114	1171	39318	35241	32943	1449	86926	47830	9760	5182
$\langle\cos\Phi\rangle_{\rm Exp}$	0·6084	0·8208	0·8927	0·9367	0·6787	0·8204	0·8881	0·9514	0·8659	0·8985	0·9364	0·9634
$\langle\cos\Phi\rangle_J$	0·6067	0·8205	0·8972	0·9389	0·6766	0·8203	0·8903	0·9577	0·8812	0·9189	0·9541	0·9750
$\langle\cos\Phi\rangle_I$	0·6260	0·8328	0·9057	0·9472	0·6914	0·8308	0·8980	0·9637	0·8616	0·9041	0·9449	0·9704
$\langle\cos\Phi\rangle_S$	0·5641	0·7714	0·8588	0·9079	0·6274	0·7713	0·8511	0·9332	0·8147	0·8645	0·9138	0·9459
(Var. $\cos\Phi)_{\rm Exp}$	0·235	0·064	0·022	0·008	0·1806	0·0655	0·0244	0·0053	0·0365	0·0207	0·0082	0·0024
(Var. $\cos\Phi)_J$	0·239	0·066	0·022	0·008	0·1819	0·0666	0·0248	0·0039	0·0291	0·0133	0·0044	0·0016
(Var. $\cos\Phi)_I$	0·223	0·058	0·018	0·006	0·1696	0·0593	0·0214	0·0030	0·0397	0·0187	0·0063	0·0019
(Var. $\cos\Phi)_S$	0·272	0·104	0·041	0·017	0·2223	0·1038	0·0460	0·0094	0·0705	0·0380	0·0151	0·0061
(Var. $\langle\cos\Phi\rangle_{\rm Exp})^{1/2}$	0·0016	0·0016	0·0016	0·0026	0·0021	0·0014	0·0009	0·0019	0·0007	0·0007	0·0009	0·0007

Concluding remarks

The exponential form of the joint probability distribution for $\cos(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{\mathbf{h}_3})$ has been shown in the test examples to be very accurate when carried out to the $N^{-3/2}$ term. In the almost equal-atom test cases this form gave excellent agreement throughout the test range. It was only superseded by the modification based on inequality theory as the relative weight of a substituted heavy atom began to increase significantly. There are extreme circumstances conceivable involving the substitution of a very heavy atom in which only the form modified by inequality theory would afford valid measures of the expected values and variances.

There are two remarkable points of interest. The first is the rather broad insensitivity of the statistical measures, such as expected values and variances, to the symmetry of the space group. Very accurate agreement was obtained between theory and calculated values for space groups ranging from the monoclinic to the tetragonal systems in spite of the fact that the theory was derived for the triclinic space group $P1$. In fact, the good agreement was not affected by including in the calculations those invariants occurring within the sphere of data which were composed solely of pure-real and pure-imaginary structure factors.

The second point worthy of notice is the manner in which the theoretical results were derived as compared to the manner in which the comparison calculations are carried out. In developing the theory, the random variables are chosen to be the $\mathbf{r}_j$, representing the coordinates of the atoms, rather than the $\mathbf{h}$. However, for any crystal the coordinates are fixed and it is pertinent to develop the theories in which the $\mathbf{h}$ are the random variables. Clearly, the comparison calculations for expected values, for example, are performed by taking averages over a very large number of $\mathbf{h}$ vectors for some given crystal structure. In fact, the theory of interest is precisely the one for which the crystal structure would be fixed and the reciprocal vectors are the random variables. The very good agreement between theory and calculation shown in the test examples suggests therefore that the probability distribution obtained from holding $\mathbf{h}$ constant and employing the $\mathbf{r}_j$ as random variables may be equal to or a very good approximation to the desired distribution in which the $\mathbf{r}_j$ are held constant and the $\mathbf{h}$ are the random variables. Certainly the statistical quantities derived from the theory are very accurate. Using the $\mathbf{r}_j$ as the random variables plays the role of a mathematical strategem designed to obtain a result whose interpretation and application are different from the basis on which it was derived. It would be possible to carry through the development of the joint probability distribution with the choice of $\mathbf{h}$ as the random variable and it might be of interest to do so for comparison purposes. However, it is convenient to follow the path presented here, namely, to employ the $\mathbf{r}_j$ as random variables, and the results justify this mode of operation.

The question arises concerning why the $\mathbf{r}_j$ can replace the $\mathbf{h}$ as random variables. The answer appears to be connected with the fact that these quantities occur as the product $\mathbf{h}.\mathbf{r}_j$ in the arguments of the trigonometric terms defining the structure factors. It is these trigonometric functions which are manipulated in the algorithms for producing the joint probability distributions. The fractional part of $\mathbf{h}.\mathbf{r}_j$ is uniformly distributed in the interval $(0,1)$ if $\mathbf{h}$ is fixed and the coordinates of $\mathbf{r}_j$ are uniformly and independently distributed in the interval $(0,1)$. The important point is that this is equally true about $\mathbf{h}.\mathbf{r}_j$ if the coordinates of $\mathbf{r}_j$ are fixed and rationally independent and the components of $\mathbf{h}$ are uniformly and independently distributed over all the integers (Hauptman & Karle, 1953*a*, p. 83). As a practical matter this circumstance is fairly well realized if the components of $\mathbf{h}$ cover a large number of integers and a coordinate of $\mathbf{r}_j$ is a rational number significantly different from zero and having a fairly large denominator. Atoms in fixed, special positions could cause difficulties. Thus, because of the special mathematical properties of $\mathbf{h}.\mathbf{r}_j$, namely, the equivalent behavior of its fractional part in the interval $(0,1)$ whether $\mathbf{h}$ or $\mathbf{r}_j$ is employed as a random variable, it is possible to use either variable to obtain the joint probability distribution. This discussion is not proof that the alternative choices are always possible under the general conditions specified for the $\mathbf{h}$ and $\mathbf{r}_j$, but it appears to explain the good agreement obtained in this type of analysis.

References

BERTAUT, E. F. (1955). *Acta Cryst.* **8**, 537–548.

BERTAUT, E. F. (1956). *Acta Cryst.* **9**, 455–460.

BERTAUT, E. F. (1960*a*). *Acta Cryst.* **13**, 546–552.

BERTAUT, E. F. (1960*b*). *Acta Cryst.* **13**, 643–649.

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.

FISHER, J. E., HANCOCK, J. H. & HAUPTMAN, H. (1970). *The Conditional Probability Distribution of Crystal Structure Invariants*, Naval Research Laboratory Report 7132.

FLIPPEN, J. (1973*a*,*b*). *Acta Cryst.* In the press.

GILARDI, R. D. & KARLE, I. L. (1972). *Acta Cryst.* B**28**, 3420–3425.

HAUPTMAN, H. & KARLE, J. (1953*a*). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* A.C.A. Monograph No. 3, Pittsburgh: Polycrystal Book Service.

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

KARLE, I. L. (1973). To be published.

KARLE, J. (1972). *Acta Cryst.* B**28**, 3362–3369.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.

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

NAYA, S., NITTA, I. & ODA, T. (1964). *Acta Cryst.* **17**, 421–433.

NAYA, S., NITTA, I. & ODA, T. (1965). *Acta Cryst.* **19**, 734–747.

TSOUCARIS, G. (1970). *Acta Cryst.* A**26**, 492–499.

WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 61–67.