GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*, ch. 1. London: Academic Press.
GULIK, A., MONTEILHET, C., DESSEN, P. & FAYAT, G. (1976). *Eur. J. Biochem.* **64**, 295-300.
HENDRIKX, Y. & CHARVOLIN, J. (1981). *J. Phys.* (*Paris*), **42**, 1427-1440.
JACROT, B. (1976). *Rep. Prog. Phys.* **39**, 911-953.
KARTHA, G. (1953). *Acta Cryst.* **6**, 817-820.
KROGH-MOE, J. (1956). *Acta Cryst.* **9**, 951-953.
LANDAU, L. D. & LIFSHITZ, E. M. (1981). *Quantum Mechanics*, 3rd ed., pp. 159-160. Oxford: Pergamon Press.
LUZZATI, V., TARDIEU, A., MATEU, L. & STUHRMANN, H. (1976). *J. Mol. Biol.* **101**, 115-127.
LUZZATI, V., WITZ, J. & NICOLAIEFF, A. (1961). *J. Mol. Biol.* **3**, 367-378, 379-392.
MORAS, D., LORBER, B., ROMBY, P., EBEL, J. P., GIEGE, R., LEWIT-BENTLEY, A. & ROTH, M. (1983). *J. Biomol. Struct. Dyn.* **1**, 209-223.
MORAS, D., PODJARNY, M., THIERRY, J. C., REES, B., GIEGE, R., LEWIT-BENTLEY, A. & ROTH, M. (1986). In preparation.
NORMAN, N. (1957). *Acta Cryst.*, **10**, 370-373.
ROTH, M., LEWIT-BENTLEY, A. & BENTLEY, G. A. (1984). *J. Appl. Cryst.* **17**, 77-84.
ROTHBAUER, R. (1978). *Acta Cryst.* **A34**, 528-533.
STUHRMANN, H. B. (1982). *Small Angle X-ray Scattering*, ch. 6: *Contrast Variation*, edited by O. GLATTER & O. KRATKY. London: Academic Press.
TARDIEU, A., MATEU, L., SARDET, C., WEISS, B., LUZZATI, V., AGGERBECK, L. & SCANU, A. M. (1976). *J. Mol. Biol.* **101**, 129-153.
UNGE, T., LILJAS, L., STRANDBERG, B., VAARA, I., KANNAN, K. K., FRIDBORG, K., NORDMAN, C. E. & LENTZ, P. J. JR (1980). *Nature* (*London*), **285**, 373-377.
WORCESTER, D. L. (1976). *Brookhaven Symp. Neutron Scattering in Biology*, edited by B. SCHOENBORN, pp. III 37-III 57. New York: Brookhaven National Laboratory.
WORCESTER, D. L., GILLIS, J. M. O'BRIEN, E. J. & IBEL, K. (1976). *Brookhaven Symp. Neutron Scattering in Biology*, edited by B. SCHOENBORN, pp. III 101-III 113. New York: Brookhaven National Laboratory.
YSEBAERT, M., VAN EMMELO, I. J. & FRIERS, W. (1980). *J. Mol. Biol.* **143**, 273-287.

Acta Cryst. (1986). A**42**, 240-246

Exact Joint Distribution of $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{h}+\mathbf{k}}$, and the Probability for the Positive Sign of the Triple Product in the Space Group $P\bar{1}$

BY URI SHMUELI

School of Chemistry, Tel Aviv University, Ramat Aviv, 69 978 Tel Aviv, Israel

AND GEORGE H. WEISS

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

(*Received 29 August* 1985; *accepted 23 December* 1985)

Abstract

A recently formulated method of deriving exact Fourier-series representations of joint probability density functions (p.d.f.'s) of several normalized structure factors is applied to the derivation of an exact expression for the conditional probability that the sign of the triple product $E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}$ is positive. The relevant joint and conditional probabilities are derived for the space group $P\bar{1}$. The Fourier coefficients of the p.d.f. are given by rapidly convergent series of Bessel functions, and the convergence properties of the Fourier summations are also found to be favourable. The exact conditional probability is compared with the currently employed approximate one, well known as the hyperbolic tangent formula, for several hypothetical structures. The examples illustrate the effects of the number of atoms in the unit cell, the magnitude of the E values and the atomic composition on the exact and approximate probabilities. It is found, in agreement with previous studies, that the hyperbolic tangent formula may indeed significantly underestimate the probability when the number of equal atoms is small and the E values are only moderately large, and when the structure contains outstandingly heavy atoms. The opposite behaviour, *i.e.* the approximate probability overestimating the exact one, was not observed in the present calculations. For large values of the triple product in equal-atom and heterogeneous models, the agreement between the approximate and exact probabilities is usually good.

Introduction

The well known hyperbolic-tangent formula, from which the probability for the positive sign of a triple-product structure invariant is conventionally estimated, is one of the most important relationships in applications of direct methods to sign determination. The current version of the relationship is based on

0108-7673/86/040240-07$01.50

that derived for unitary structure factors by Cochran & Woolfson (1955), who invoked the central limit theorem (*e.g.* Cramér, 1951) for the purpose of this and related derivations. This form of the conditional probability for the positive sign of the triple product $E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}$, given its magnitude, is therefore a low-order approximation.

Generalizations of the joint probability density function (p.d.f.) of $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{h}+\mathbf{k}}$, as well as those of the conditional probability for the positive sign of the triple product, have been rather extensively investigated (*e.g.* Klug, 1958; Karle, 1972; Karle & Gilardi, 1973; Giacovazzo, 1977). All of these studies lead to improved formulae for the joint and conditional probabilities, but the extent of the improvement is not certain since they are all based on truncated expansions of the Edgeworth or related series. Similar uncertainties were experienced in studies of the simpler univariate generalized p.d.f.'s (*e.g.* Shmueli, 1982), and these were first resolved by the use of random-walk techniques that permit one to obtain exact expressions for probability densities of $|E|$ (Shmueli, Weiss, Kiefer & Wilson, 1984). These techniques were further extended to the multivariate case and a preliminary account of the work on this topic has been published (Weiss, Shmueli, Kiefer & Wilson, 1985).

A rather detailed outline of the general procedures involved in the derivation of exact joint p.d.f.'s of several structure factors, illustrated by a derivation of an exact $\sum_1$ relationship for the space group $P\bar{1}$, is given by Shmueli & Weiss (1985). In view of the importance of the triple-product sign relationship, it seems desirable to apply the method described in the latter reference to the derivation of an exact joint p.d.f. of $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{h}+\mathbf{k}}$, and use it in the construction of an exact expression for the conditional probability for the positive sign of their product. The derivation of such an expression is of interest for its own sake, and may provide reliable answers to several questions. It has been pointed out that the usual hyperbolic tangent formula is an underestimate of an improved expression for this conditional probability (Karle, 1972). This is clearly important, as far as sign-acceptance criteria are concerned, and an exact expression for this probability could usefully show the actual extent of this underestimation. The discrepancy between the exact and approximate conditional probabilities is also bound to depend on the magnitudes of the E values, on the number of atoms in the unit cell and on the atomic composition. A clarification of these matters is of theoretical importance, and may well be of practical significance since it can lead to better-established sign-acceptance criteria, for use with the conventional hyperbolic tangent formula.

The statistical background of the present method is briefly summarized in Appendix *A*, in which we try to place it in the perspective of the conventional formalism of mathematical statistics. The next section deals with the derivation in the space group $P\bar{1}$ of the joint p.d.f. of three normalized structure factors that form a structure invariant. This derivation is followed by that of an exact conditional probability for the positive sign of the corresponding triple product, and we conclude with some numerical examples of the above announced results, a comparison with those given by the conventional hyperbolic tangent formula (Cochran & Woolfson, 1955) and a discussion in view of the questions stated above.

The Σ_2 relationship in $P\bar{1}$

The present derivation is an application of our method for the construction of exact joint p.d.f.'s of several structure factors for centrosymmetric space groups, presented elsewhere (Shmueli & Weiss, 1985). We shall also follow the notation of the latter reference, as well as the set of underlying assumptions there outlined.

We derive an exact expression for the probability that the sign of the triple product $E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}$ is positive, given the magnitudes of the structure factors involved and assuming the space group $P\bar{1}$. As shown elsewhere (Shmueli & Weiss, 1985), we can conveniently start from the characteristic function of the joint p.d.f. of $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{h}+\mathbf{k}}$ which, for the above space group, is given by

$$C(\boldsymbol{\omega}) = \langle \exp[i(\omega_1 E_{\mathbf{h}} + \omega_2 E_{\mathbf{k}} + \omega_3 E_{\mathbf{h}+\mathbf{k}})]\rangle \tag{1}$$

$$= \left\langle \exp\left\{2i \sum_{j=1}^{N/2} n_j[\omega_1 \cos\theta_j + \omega_2 \cos\varphi_j + \omega_3 \cos(\theta_j+\varphi_j)]\right\}\right\rangle \tag{2}$$

$$= \prod_{j=1}^{N/2} C_j(\boldsymbol{\omega}), \tag{3}$$

where N is the number of atoms in the unit cell, n_j is the normalized scattering factor, $\theta_j = 2\pi\mathbf{h}\,.\,\mathbf{r}_j$, $\varphi_j = 2\pi\mathbf{k}\,.\,\mathbf{r}_j$ and

$$C_j(\boldsymbol{\omega}) = (4\pi^2)^{-1} \iint_{-\pi}^{\pi} \exp\{2in_j[\omega_1\cos\theta + \omega_2\cos\varphi + \omega_3\cos(\theta+\varphi)]\}\,d\theta\,d\varphi \tag{4}$$

is the atomic contribution to the characteristic function. Equation (2) is a version of the general form of the characteristic funtion [Shmueli & Weiss, 1985, equation (8)], adapted to the present problem. As in the latter study, we assume that the contributions of different atoms in the asymmetric unit are independent, and the atomic phase factors θ_j and φ_j can be taken as uniformly distributed over the $[0, 2\pi]$ range. The latter assumption is usually valid if all the atoms are located in general positions and effects of rational dependence (Hauptman & Karle, 1959) are negligible. More fundamental aspects of this assumption, pertaining to whether **h** or **r** is taken as the primitive random variable, are discussed in the literature (*e.g.* Hauptman & Karle, 1953; Karle & Gilardi, 1973; Giacovazzo, 1980).

The integral in (4) can be reduced to series form with the aid of the identity

$$\exp(ix\cos\beta)=\sum_{k=-\infty}^{\infty} i^k J_k(x)\exp(ik\beta) \qquad (5)$$

[Gradshteyn & Ryzhik, 1980, entry: 8·511(4)], which leads to a straightforward integration over the angles θ and φ in (4). We obtain

$$C_j(\boldsymbol{\omega})=R_j(\boldsymbol{\omega})+iI_j(\boldsymbol{\omega}), \qquad (6)$$

where

$$R_j(\boldsymbol{\omega})=\prod_{k=1}^{3} J_0(2n_j\omega_k) + 2\sum_{m=1}^{\infty}(-1)^m\prod_{k=1}^{3} J_{2m}(2n_j\omega_k) \qquad (7)$$

and

$$I_j(\boldsymbol{\omega})=2\sum_{m=0}^{\infty}(-1)^{m+1}\prod_{k=1}^{3} J_{2m+1}(2n_j\omega_k). \qquad (8)$$

The characteristic function (1) can now be written as

$$C(\boldsymbol{\omega})=\prod_{j=1}^{N/2}[R_j(\boldsymbol{\omega})+iL_j(\boldsymbol{\omega})]=R(\boldsymbol{\omega})+iI(\boldsymbol{\omega}), \qquad (9)$$

where use is made of (6)-(8) and (3).

As shown elsewhere (Shmueli & Weiss, 1985), the Fourier coefficients of the required p.d.f. can be obtained by replacing the variables ω_r in the characteristic function with quantities related to the appropriate summation indices. In the present application we put

$$\omega_1\to\pi\alpha r,\quad \omega_2\to\pi\alpha s \quad\text{and}\quad \omega_3\to\pi\alpha t, \qquad (10)$$

where r, s, t are the indices of the (triple) Fourier series for the required p.d.f. and

$$1/\alpha=E_{\max}=\sum_{j=1}^{N} n_j \qquad (11)$$

is the maximum magnitude of E, by virtue of which the Fourier expansion of the p.d.f. becomes meaningful (*e.g.* Shmueli, Weiss, Kiefer & Wilson, 1984; Shmueli & Weiss, 1985). Introducing the abbreviation

$$q_{rst}=q(\pi\alpha r, \pi\alpha s, \pi\alpha t), \qquad (12)$$

we can write the joint p.d.f. as

$$p(E_1,E_2,E_3)=(\alpha^3/8)\sum_r\sum_s\sum_t(R_{rst}+iI_{rst}) + \exp[-i(rE_1+sE_2+tE_3), \qquad (13)$$

where $E_1=\pi\alpha E_{\mathbf{h}}$, $E_2=\pi\alpha E_{\mathbf{k}}$, $E_3=\pi\alpha E_{\mathbf{h+k}}$ and, of course, only the real part of (13) is relevant.

The Fourier series in (13) can be reduced to a summation over the non-negative octant of the *rst* index space, when use is made of the symmetry properties of the trigonometric functions, as well as those of R_{rst} and I_{rst}. The latter are found by a direct inspection of (7)-(9), with the aid of some basic properties of Bessel functions $J_n(x)$ of integer order (*e.g.* Abramowitz & Stegun, 1972), to be

(i) R_{rst} is symmetric in all three indices (*i.e.* invariant to their sign changes), and

(ii) I_{rst} changes its sign when one or three indices change sign(s), and is symmetric with respect to the change of signs of two indices.

The simplification of (13) follows in much the same way as the simplification of expressions for the electron density, for given symmetries of the reciprocal lattice (*e.g. International Tables for X-ray Crystallography*, 1952); drawing an analogy between the *rst* and *hkl* Fourier spaces, R_{rst} and I_{rst} can be said to possess symmetries isomorphic to *mmm* and 222 respectively.

Making use of these symmetry considerations, we obtain

$$p(E_1,E_2,E_3)=(\alpha^3/8)(p_{\text{even}}+p_{\text{odd}}), \qquad (14)$$

where

$$\begin{aligned}p_{\text{even}}=1&+2\sum_{r=1}^{\infty}R_{r00}[\cos(rE_1)+\cos(rE_2)+\cos(rE_3)]\\&+4\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}R_{rs0}[\cos(rE_1)\cos(sE_2)\\&+\cos(rE_2)\cos(sE_3)+\cos(rE_3)\cos(sE_1)]\\&+8\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\sum_{t=1}^{\infty}R_{rst}\cos(rE_1)\cos(sE_2)\cos(tE_3)\end{aligned} \qquad (15)$$

and

$$p_{\text{odd}}=-8\sum_{r=1}^{\infty}\sum_{s=1}^{\infty}\sum_{t=1}^{\infty}I_{rst}\sin(rE_1)\sin(sE_2)\sin(tE_3). \qquad (16)$$

Clearly, the p_{even} component is independent of the sign of the triple product, while the sign combinations '−++', '+−+', '++−' and '−−−', which lead to negative triple products, also reverse the sign of p_{odd} in (16). The ratio of the probabilities for negative and positive triple products is therefore given by

$$\frac{p_-(\mathbf{h},\mathbf{k},\mathbf{h+k})}{p_+(\mathbf{h},\mathbf{k},\mathbf{h+k})}=\frac{p_{\text{even}}(E_{\mathbf{h}},E_{\mathbf{k}},E_{\mathbf{h+k}})-p_{\text{odd}}(|E_{\mathbf{h}}|,|E_{\mathbf{k}}|,|E_{\mathbf{h+k}}|)}{p_{\text{even}}(E_{\mathbf{h}},E_{\mathbf{k}},E_{\mathbf{h+k}})+p_{\text{odd}}(|E_{\mathbf{h}}|,|E_{\mathbf{k}}|,|E_{\mathbf{h+k}}|)} \qquad (17)$$

and since

$$p_+=p_+/(p_++p_-)=(1+p_-/p_+)^{-1} \qquad (18)$$

(Bertaut, 1955), the required probability that the triple product $E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}$ is positive is given by

$$p_+(\mathbf{h},\mathbf{k},\mathbf{h}+\mathbf{k}) = \frac{1}{2}\left[1 + \frac{p_{\text{odd}}(|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}+\mathbf{k}}|)}{p_{\text{even}}(|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}+\mathbf{k}}|)}\right], \quad (19)$$

where p_{even} and p_{odd} are defined by (15) and (16) respectively.

It should be pointed out that, since for any non-zero integer n, $J_n(0)=0$ and $J_0(0)=1$, the 'special' coefficients in (15) are given by

$$R_{r00} = \prod_{j=1}^{N/2} J_0(2\pi r\alpha n_j) \quad (20)$$

and

$$R_{rs0} = \prod_{j=1}^{N/2} J_0(2\pi r\alpha n_j)J_0(2\pi s\alpha n_j), \quad (21)$$

since the infinite series of products of Bessel functions in (7) then vanishes identically. This is also the reason for the absence of such special coefficients from p_{odd} in (16). These series of Bessel functions converge after the first few terms, and the convergence behaviour of (15) and (16) is also very good (see next section).

Equations (14) and (19) are exact. It is possible to rederive from the present results the classical asymptotic expression for the probability, *i.e.*

$$p_+(\mathbf{h},\mathbf{k},\mathbf{h}+\mathbf{k}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh[(\sigma_3/\sigma_2^{3/2})|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}|], \quad (22)$$

where $\sigma_m = \sum_{j=1}^{N} f_j^m$ (Hauptman & Karle, 1953; Woolfson, 1954; Cochran & Woolfson, 1955). We do not present here the details of this rederivation, since it follows very closely an analogous treatment of the probability for the positive sign of $E_{2\mathbf{h}}$, given the moduli $|E_{\mathbf{h}}|$ and $|E_{2\mathbf{h}}|$ (see Shmueli & Weiss, 1985, Appendix *A*). As in the latter study, the above rederivation is based on approximating the characteristic function by its lowest-order terms, and recalculating the approximate joint p.d.f. by a Fourier inversion.

Test calculations and discussion

In this section we compare some calculations of well converged results corresponding to (19) with the hyperbolic tangent approximation (22), for several hypothetical $P\bar{1}$ structures defined in terms of their composition. The effects that we would like to illustrate are those due to (i) the number of atoms in the asymmetric unit, (ii) the magnitudes of the E values and (iii) atomic heterogeneity of the structure. Of course, only a very limited number of combinations of the various parameters is possible within the scope of this paper, and the illustrations are by no means exhaustive.

All the numerical computations have been programmed in Fortran, and use has been made of a local library subroutine for the generation of Bessel functions of general order. As can be seen from (7) and (8), such a subroutine is an important prerequisite; its stability is usually achieved by a judicious use of forward and backward recursion techniques. A straightforward summation of the Fourier series, with computed arrays of sines and cosines only, required approximately 0·3 s for each triple product examined. The computation of the Fourier coefficients needs to be done only once for a given structure, and is the least-time-consuming stage also because of the excellent convergence properties of the series of Bessel functions in (7) and (8); for equal-atom structures only three or four terms are needed for a cutoff ratio of 10^{-5}, while more may be required in the presence of heavy atoms. The number of terms in a single Fourier summation required to achieve a comparable accuracy was found to be at least 15, and more than that if the number of atoms in the asymmetric unit is very small. The computing time should be reduced by the use of a fast Fourier algorithm.

The composition of the asymmetric unit was taken as C_nX_m, where X may be carbon or a heavier atom, and all the composition-dependent quantities that are required for the computation of (7), (8) and (22) were expressed in terms of the numbers n and m of the atoms, and the ratio of their atomic numbers, $\rho = Z_X/Z_{\mathrm{C}}$. For example,

$$\begin{aligned}(\sigma_3/\sigma_2^{3/2}) &= 2(nf_{\mathrm{C}}^3 + mf_X^3)/[2(nf_{\mathrm{C}}^2 + mf_X^2)]^{3/2}\\ &= 2(n+m\rho^3)/[2(n+m\rho^2)]^{3/2}\end{aligned}$$

(*cf.* Shmueli & Weiss, 1985).

Figs. 1(*a*) and (*b*) show the probability for a positive sign of the triple product for the compositions C_{32} and $\mathrm{C}_{30}\mathrm{Kr}_2$ respectively, assuming $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 1{\cdot}50$ and letting $|E_{\mathbf{h}+\mathbf{k}}|$ range from 1·0 to 3·0. It is seen that for the equal-atom case (Fig. 1*a*) the two probabilities run quite close to each other, with the discrepancy increasing slightly with increasing value of the triple product. The 'replacement' of two carbons by kryptons ($Z=36$) enhances strongly both exact and approximate probabilities (Fig. 1*b*), but also leads to more serious discrepancies at the low values of the triple product. Taking $p_+ = 0{\cdot}95$ as a tentative threshold of acceptability, it is seen that the exact expression (19) would admit significantly lower $|E|$ values than would the hyperbolic tangent formula (22).

The effect of the magnitude of the E values, for a rather small asymmetric unit, C_{10}, is illustrated in Fig. 2. The discrepancy between the exact and approximate expressions seems to decrease as the value of the triple product increases, and the region of maximum discrepancy moves towards lower values of $|E_{\mathbf{h}+\mathbf{k}}|$ as the values of $|E_{\mathbf{h}}|$ and $|E_{\mathbf{k}}|$ increase. It should be pointed out that the probability computed from (19) is practically unity throughout the range:

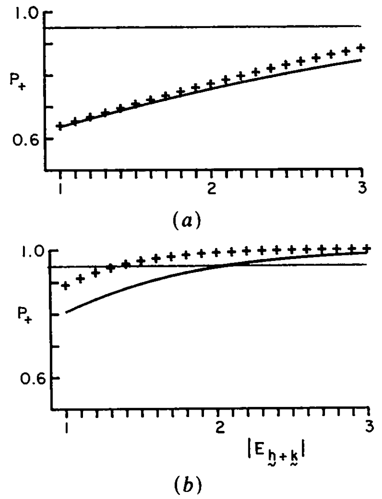

Fig. 1. Enhancement of the probabilities by the presence of heavy atoms. The figure shows the exact (+) and approximate (solid line) probability $p_+(\mathbf{h}, \mathbf{k}, \mathbf{h}+\mathbf{k})$ computed from equations (19) and (22) respectively for two compositions of a 32-atom asymmetric unit of $P\bar{1}$. The $|E|$ values are taken as $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 1{\cdot}50$, and $|E_{\mathbf{h+k}}|$ ranges from 1 to 3. (*a*) C_{32}, (*b*) $C_{30}Kr_2$. A faint line is drawn at $p_+ = 0{\cdot}95$.

$1 < |E_{\mathbf{h+k}}| < 3$, in Fig. 2(*c*), as would be expected on the basis of the related Harker-Kasper inequality (Harker & Kasper, 1948) (note that $U = \alpha E$, and $\alpha = 1/20^{1/2}$ for this composition, where U is the unitary structure factor). Similar qualitative behaviour is observed for other values of N, with increasing discrepancies for smaller N's and decreasing discrepancies for greater number of equal atoms in the unit cell. The overall impression from these tests is that the exact and approximate probabilities tend to agree best in the regions in which they are very useful or nearly useless, when regarded as sign-acceptance criteria, the regions of significant discrepancies being associated with moderately large triple products.

Fig. 3 illustrates the effect of atomic heterogeneity on sign indications for a moderately large asymmetric unit, $C_{95}U_5$. Although the contribution of the uranium atoms ($Z = 92$) is dominant, the discrepancies at lower values of the triple product are significantly smaller than those obtained for an equal-atom structure with five atoms in the asymmetric unit. Still, the favourable effect of the heavy atoms on the sign

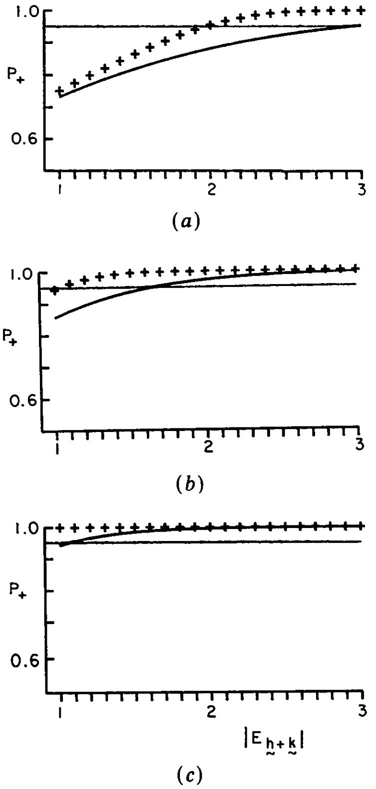

Fig. 2. Effects of paucity of atoms for selected E values. The figure shows the exact (+) and approximate (solid line) probability $p_+(\mathbf{h}, \mathbf{k}, \mathbf{h}+\mathbf{k})$ computed from equations (19) and (22) respectively. The composition of the asymmetric unit of $P\bar{1}$ is C_{10}. the values of $|E_{\mathbf{h}}|$ and $|E_{\mathbf{k}}|$ are taken to be the same and are given by: (*a*) $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 1{\cdot}50$, (*b*) $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 2{\cdot}00$ and (*c*) $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 2{\cdot}50$. A faint line is drawn at $p_+ = 0{\cdot}95$.

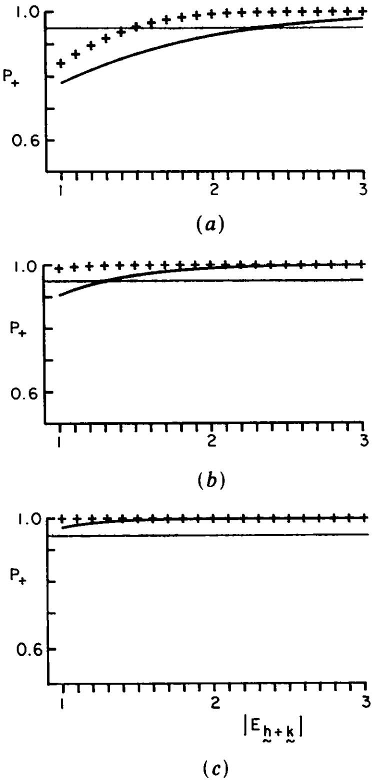

Fig. 3. A small number of very heavy atoms in a moderately large unit. The figure shows the exact (+) and approximate (solid line) probability $p_+(\mathbf{h}, \mathbf{k}, \mathbf{h}+\mathbf{k})$ computed from equations (19) and (22) respectively. The composition of the asymmetric unit of $P\bar{1}$ is $C_{95}U_5$. The values of $|E_{\mathbf{h}}|$ and $|E_{\mathbf{k}}|$ are taken to be the same and are given by: (*a*) $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 1{\cdot}50$, (*b*) $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 2{\cdot}00$ and (*c*) $|E_{\mathbf{h}}| = |E_{\mathbf{k}}| = 2{\cdot}50$. A faint line is drawn at $p_+ = 0{\cdot}95$.

indications is appreciably greater for the probabilities computed from the exact expression (19) than from the approximate one (22).

These preliminary numerical examples indicate that the application of the exact expression (19) to sign determination may promote this process in two respects. Firstly, an implementation of (19) in relevant application software may enhance the reliability of sign determination during the initial stages of this process. Perhaps more important, at least in the near future, are extensive comparisons and analyses of both types of expressions, which may lead to modified and, hopefully, more useful acceptance criteria that might enable one to use the conventional and faster approximate techniques to better advantage.

There seem to be two requirements for the present method to be applicable to real situations. The first of these appears to be the need for accurate E values, since exact probabilities are being used. However, it is too early to make a more definite judgement and only experience may enable one to do so. The other requirement, and probably a more important one, is the very applicability of the usual statistical techniques, which are common to all direct methods, to the structure under consideration. As mentioned above, the exponents of the atomic phase factors are assumed to be uniformly distributed in the $[0, 2\pi]$ range. This assumption, most frequently made in structure-factor statistics, is likely to hold well if the coordinates of any atom are not all expressible in terms of simple fractions; more generally, the atomic coordinates are required to be rationally independent (Hauptman & Karle, 1953, 1959). However, conspicuous rational dependence, especially among the heavy scatterers, is likely to invalidate the predictions of both the asymptotic (22) and the exact (19) expressions, and further studies of such predictable (*e.g.* fixed special positions) and unpredictable effects are certainly of interest.

We also wish to point out what may be considered as an asset of the approximate expressions for $\sum_1$ and $\sum_2$. In none of the test calculations that were performed during the present work did we find an instance in which the approximate conditional probability is greater than the exact one. This is in agreement with the remark made by Karle (1972) in his article on the exponential form of the joint probability distribution. The same is true of our previous study of the $\sum_1$ relationship (Shmueli & Weiss, 1985), for $|E_{\mathbf{h}}|$ values that are of practical importance (*i.e.* exceeding a low-lying threshold of about one, below which the exact probabilities are underestimates of the approximate ones). We have no proof that this is completely general, but at least in the space group $P\bar{1}$ and for a variety of parameter choices the approximate probability indications are usually underestimates of the exact ones, which places the former on the safe side - provided, of course, the $|E|$ values are sufficiently accurate and the assumption of uniformly distributed atomic phase factors is satisfied.

We wish to thank one of the referees for very pertinent philosophical remarks. All the computations related to this work were carried out at the Tel Aviv University Computation Center, on Cyber 170-855 and IBM 4381 computers, using a local library subroutine for the generation of Bessel functions of general order.

APPENDIX *A*
Statistical background

Let $\mathbf{x} = (x_1, \ldots, x_n)$ be a vector of n independent random variables, each obeying the same distribution. For example, $\mathbf{x}$ can be a vector of structure factors, which obey - for a given structure - a certain univariate distribution. The joint probability density function of the components of $\mathbf{x}$, denoted by $p(\mathbf{x}) = p(x_1, \ldots, x_n)$, is associated with a characteristic function which is defined as

$$C(\boldsymbol{\omega}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(\mathbf{x}) \exp(i\boldsymbol{\omega}^T\mathbf{x})\, \mathrm{d}^n\mathbf{x} = \langle \exp(i\boldsymbol{\omega}^T\mathbf{x})\rangle \tag{A1}$$

(*e.g.* Cramér, 1951). This characteristic function (c.f.) is a Fourier transform of the p.d.f. $p(\mathbf{x})$, but it can also be regarded as an expectation value of the exponential $\exp(i\boldsymbol{\omega}^T\mathbf{x})$ in the integrand in (*A*1). The second viewpoint is of a greater practical importance, since the p.d.f. $p(\mathbf{x})$ is usually initially unknown, while $\langle \exp(i\boldsymbol{\omega}^T\mathbf{x})\rangle$ can often be evaluated from known statistical properties of the random vector $\mathbf{x}$. Given the c.f. $C(\boldsymbol{\omega})$, the p.d.f. $p(\mathbf{x})$ can then be written as an inverse Fourier transform of (*A*1):

$$p(\mathbf{x}) = (2\pi)^{-n} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} C(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega}^T\mathbf{x})\mathrm{d}^n\boldsymbol{\omega}. \tag{A2}$$

Equation (*A*2) can be brought to a manageable form, provided the c.f. has been exactly or approximately evaluated and the integral in (*A*2) can be solved. Except for some simple distributions, this integral cannot usually be evaluated in closed form and one often makes use of the moment-generating and cumulant-generating properties of the c.f. $C(\boldsymbol{\omega})$ (*e.g.* Kendall & Stuart, 1969) in constructing approximate expansions for the joint p.d.f., in terms of moments or cumulants, and polynomials that result from the term-by-term integration of (2). These expansions, known as Gram-Charlier or Edgeworth series (*e.g.* Cramér, 1951), are usually truncated after the first few terms, the main reason being the rapidly increasing complexity of their successive terms. Detailed descriptions of the construction of such approximations are given, for example , by Klug (1958) and Giacovazzo (1980).

An exact expression for the joint p.d.f. which, at least for small values of n, can often be accurately evaluated, can be obtained in the important case in which all the components of the random vector $\mathbf{x}$ may be non-zero only in a bounded range, say $-x_M < x_i < +x_M$, $i = 1, \ldots, n$ (Weiss, Shmueli, Kiefer & Wilson, 1985; Shmueli & Weiss, 1985). One can then expand the joint p.d.f. $p(\mathbf{x})$ in a Fourier series

$$p(\mathbf{x}) = (2x_M)^{-n} \sum_{\mathbf{u}} C_{\mathbf{u}} \exp(-\pi i \mathbf{u}^T \mathbf{x}/x_M) \qquad (A3)$$

and write the Fourier coefficients, in the conventional manner, as

$$C_{\mathbf{u}} = \int_{-x_M}^{x_M} \ldots \int_{-x_M}^{x_M} p(\mathbf{x}) \exp(\pi i \mathbf{u}^T \mathbf{x}/x_M)\, \mathrm{d}^n\mathbf{x}. \qquad (A4)$$

Since, however, the random variables, x_i, are confined to the $[-x_M, +x_M]$ range, the probability of finding any of them outside this range is necessarily equal to zero. We can thus replace, with no loss of generality, the limits of integration in ($A4$) by $\pm\infty$ and obtain for the Fourier coefficients an expression analogous to ($A1$). In fact, the Fourier coefficients, $C_{\mathbf{u}}$, are then just the values of the characteristic function at the points: $(\omega_1, \ldots, \omega_n) = (\pi u_1/x_M, \ldots, \pi u_n/x_M)$. The practical significance of ($A3$) is now conditioned by our ability to evaluate the characteristic function, and by the convergence properties of the resulting Fourier summation.

In some applications one has to deal with random vectors in which not all the components are independent. If, for example, $\mathbf{x} = (x_1, x_2, x_3)$ and x_3 depends on x_1 and x_2, we shall still have a triple Fourier series to compute, but the integration leading to the characteristic function (or the Fourier coefficients) involves only the independent variables. For examples of such situations see Shmueli & Weiss (1985) and the derivation in the text.

References

ABRAMOWITZ, M. & STEGUN, I. (1972). *Handbook of Mathematical Functions.* New York: Dover.
BERTAUT, E. F. (1955). *Acta Cryst.* **8**, 823-832.
COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1-12.
CRAMÉR, H. (1951). *Mathematical Methods of Statistics.* Princeton Univ. Press.
GIACOVAZZO, C. (1977). *Acta Cryst.* A**33**, 50-54.
GIACOVAZZO, C. (1980). *Direct Methods in Crystallography.* New York: Academic Press.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Tables of Integrals, Series and Products.* New York: Academic Press.
HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70-75.
HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.
HAUPTMAN, H. & KARLE, J. (1959). *Acta Cryst.* **12**, 846-850.
International Tables for X-ray Crystallography (1952). Vol. I, *Symmetry Groups*, edited by N. F. M. HENRY & K. LONSDALE. Birmingham: Kynoch Press.
KARLE, J. (1972). *Acta Cryst.* B**28**, 3362-3369.
KARLE, J. & GILARDI, R. D. (1973). *Acta Cryst.* A**29**, 401-407.
KENDALL, M. G. & STUART, A. (1969). *The Advanced Theory of Statistics*, Vol. 1, 3rd ed. London: Charles Griffin.
KLUG, A. (1958). *Acta Cryst.* **11**, 515-543.
SHMUELI, U. (1982). *Acta Cryst.* A**38**, 362-371.
SHMUELI, U. & WEISS, G. H. (1985). *Acta Cryst.* A**41**, 401-408.
SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* A**40**, 651-660.
WEISS, G. H., SHMUELI, U., KIEFER, J. E. & WILSON, A. J. C. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 23-42. Guilderland, NY: Adenine Press.
WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 61-64.

Acta Cryst. (1986). A**42**, 246-253

Triplet Phase Invariants from Single Isomorphous Replacement or One-Wavelength Anomalous Dispersion Data, Given Heavy-Atom Information

BY JEROME KARLE

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375-5000, USA

(*Received* 17 *September* 1985; *accepted* 14 *January* 1986)

Abstract

Certain general algebraic formulas for computing triplet phase invariants become accessible when structural information is available concerning the replacement atoms in isomorphous replacement or the predominant type of anomalously scattering atoms in one-wavelength anomalous dispersion experiments. The formulas of interest are presented and subjected to a number of test calculations to obtain insight into their accuracy and to determine the effects of errors in the data. The formulas are simple to calculate and some possible strategies for their use are discussed.

Introduction

On the basis of certain mathematical and physical considerations that pertain to isomorphous replacement or anomalous dispersion experiments, rules

0108-7673/86/040246-08$01.50