BUNGE, H. J. (1974). *Krist. Tech.* **9**, 4, 413-424.
ESLING, C., BUNGE, H. J., PHILIPPE, M. J. & MULLER, J. (1987). In *Theoretical Methods of Texture Analysis*, edited by H. J. BUNGE, pp. 105-126. Berlin: Informationsgesellschaft Verlag.
HASHIN, Z. & SHTRIKMAN, S. (1962). *J. Mech. Phys. Solids*, **10**, 343-352.
HIRTH, J. P. & LOTHE, J. (1968). *Theory of Dislocations.* New York: McGraw-Hill.
MATTHIES, S. & VINEL, G. W. (1983). *Phys. Status Solidi B*, **112**, K115-K120.
REUSS, A. (1929). *Z. Angew. Math. Mech.* **9**, 49-58.
VOIGT, W. (1928). *Lehrbuch der Kristallphysik.* Leipzig: Teubner.

Acta Cryst. (1989). A**45**, 361-367

Exact Conditional Distribution of a Three-Phase Invariant in the Space Group *P*1. I. Derivation and Simplification of the Fourier Series

BY URI SHMUELI AND SAVELY RABINOVICH

School of Chemistry, The Sackler Faculty of Sciences, Tel Aviv University, 69 978 *Tel Aviv, Israel*

AND GEORGE H. WEISS

Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892, *USA*

(*Received* 7 *June* 1988; *accepted* 3 *January* 1989)

Abstract

An exact expression is derived for the conditional probability density function of a three-phase invariant and the general result is applied to the space group *P*1. The expression for the conditional density is given in terms of a sixfold Fourier series. A straightforward numerical evaluation of this series, without further analysis, is extremely time consuming. Much of the present paper is therefore devoted to an exposition of symmetries hidden in the various summations. The computational effort required for the evaluation of the above expression is thereby reduced to manageable proportions in a number of interesting cases. Results of numerical computations of the exact conditional density are given in the second paper in this series. It is also shown that the exact expression for the conditional probability reduces to that given by Cochran [*Acta Cryst.* (1955), **8**, 473-478].

Introduction

Current approaches to phase determination by direct methods are based on the use of linear combinations of phases invariant under a shift of the origin of the unit cell. Such combinations are generally termed structure invariants (Hauptman & Karle, 1953). Of these, one of the most extensively used is the so-called three-phase structure invariant associated with the product of the three normalized structure factors $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{-\mathbf{h}-\mathbf{k}}$, *i.e.*

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}. \qquad (1)$$

In order to use the three-phase invariant for structure determination one needs to know the conditional probability density function (c.p.d.f.) of Φ, where the conditioning is on the values $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$ and $|E_{-\mathbf{h}-\mathbf{k}}|$. An approximate form of this c.p.d.f., based on the central limit theorem, was first derived by Cochran (1955). Corrections to this result were subsequently calculated by a number of investigators in terms of Gram-Charlier, Edgeworth, or exponentiated series [*e.g.* Naya, Nitta & Oda (1965); Hauptman (1971); Karle (1972); Karle & Gilardi (1973); Giacovazzo (1974); Peschar & Schenk (1986); a rich source of references is the book by Giacovazzo (1980)]. As in the above studies we shall assume that the primitive random variables of the problem are the atomic coordinates, which vary independently and uniformly over the (0, 1) range. However, none of the approximate c.p.d.f.'s have been compared with exact results to assess their accuracy. Such an assessment is the motivation for the present study.

We have recently developed methods, not involving the approximations based on the central limit theorem, for calculating exact representations of p.d.f.'s useful in crystallographic applications (*e.g.* Shmueli, Weiss, Kiefer & Wilson, 1984; Shmueli & Weiss, 1985, 1986). Some of the problems considered include that of finding exact representations of the p.d.f.'s for $|E|$ in various space groups as well as for joint p.d.f.'s for E from which exact $\sum_1$ and $\sum_2$ relationships can be recovered. In the present two papers we present some analogous results for the p.d.f. of the three-phase invariant Φ relevant to the solution of the phase problem in non-centrosymmetric structures. As in the above studies we shall assume that the primitive random variables of the problem are the atomic coordinates which, in the space group *P*1, vary independently and uniformly over the (0, 1) range.

0108-7673/89/060361-07$03.00

A p.d.f. for the n-phase invariant has been derived in closed form by Castleden (1987). However, no numerical computations accompanied his derivation and, indeed, even in the simple case of the three-phase invariant, straightforward evaluation of the Fourier-series representation of the c.p.d.f. presents formidable numerical problems. In this first paper of this series we outline a derivation of an exact Fourier representation of the c.p.d.f. of Φ given values of $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$ and $|E_{-\mathbf{h}-\mathbf{k}}|$, for space group $P1$. We also show that Cochran's approximation to the p.d.f. follows directly from the form of the characteristic function corresponding to the c.p.d.f. of Φ. We then briefly discuss some of the symmetrization techniques required for reducing the numerical evaluation of the exact c.p.d.f. to manageable form, and present some detailed illustrations thereof in Appendix *A*. In the accompanying paper we present some results of specific numerical calculations, showing that approximations based on the central limit theorem underestimate the probabilities of phase determination. These conclusions are consistent with those found earlier for methods based on $\sum_1$ and $\sum_2$ relationships for $P\bar{1}$ (Shmueli & Weiss, 1985, 1986).

Derivation

The first step of our derivation is the construction of a joint probability density function (j.p.d.f.) of the real and imaginary parts of the three structure factors related to Φ. Let these structure factors be written as $E_{\mathbf{h}} = A_1 + iB_1$, $E_{\mathbf{k}} = A_2 + iB_2$ and $E_{-\mathbf{h}-\mathbf{k}} = A_3 + iB_3$. Observe that all the A's and B's can only differ from zero in the range $(-1/\alpha, 1/\alpha)$, where $1/\alpha$ is the maximum value of $|E|$. We may therefore express the required j.p.d.f. as the sixfold Fourier series

$$p(\mathbf{E}) = K \sum_{\mathbf{u}} C_{\mathbf{u}} \exp\left[-\pi i\alpha \sum_{k=1}^{3} (u_{2k-1}A_k + u_{2k}B_k)\right], \quad (2)$$

where $\mathbf{E}^T = (A_1, B_1, A_2, B_2, A_3, B_3)$, K is a normalization constant, $\mathbf{u}^T = (u_1, u_2, u_3, u_4, u_5, u_6)$ is the vector of summation indices and $C_{\mathbf{u}}$ is a Fourier coefficient (see *Example* below). The following abbreviations will be employed:

$$E_1 \equiv |E(\mathbf{h})|, \quad E_2 \equiv |E(\mathbf{k})|, \quad E_3 \equiv |E(-\mathbf{h}-\mathbf{k})|,$$

$$\varphi_1 \equiv \varphi_{\mathbf{h}}, \quad \varphi_2 \equiv \varphi_{\mathbf{k}} \quad \text{and} \quad \varphi_3 \equiv \varphi_{-\mathbf{h}-\mathbf{k}}.$$

We can therefore write $A_i = E_i \cos\varphi_i$ and $B_i = E_i \sin\varphi_i$, which allows us to transform (2) to the form

$$p(\mathbf{E}) = K \sum_{\mathbf{u}} C_{\mathbf{u}} \exp\left[-\pi i\alpha \sum_{k=1}^{3} E_k(u_{2k-1}\cos\varphi_k + u_{2k}\sin\varphi_k)\right] \quad (3)$$

$$= K \sum_{\mathbf{u}} C_{\mathbf{u}} \exp\left[-i \sum_{k=1}^{3} \Omega_k \cos(\varphi_k - \Delta_k)\right], \quad (4)$$

where

$$\Omega_k = \pi\alpha E_k (u_{2k-1}^2 + u_{2k}^2)^{1/2} \quad (5)$$

and

$$\Delta_k = \tan^{-1}(u_{2k}/u_{2k-1}). \quad (6)$$

The j.p.d.f. (4) depends on the individual moduli and phases of the three structure factors, and its dependence on the three-phase invariant can be established by replacing, for example, φ_3 by $\Phi - \varphi_1 - \varphi_2$. If we now integrate out φ_1 and φ_2, the required c.p.d.f. is obtained. The required angular integrals can be evaluated numerically, but it is simpler to represent the exponential in terms of Bessel functions, by making use of the relation

$$\exp(ix\cos\beta) = \sum_{k=-\infty}^{\infty} i^k J_k(x) \exp(ix\beta) \quad (7)$$

(*e.g.* Gradshteyn & Ryzhik, 1980). The required integral can therefore be written as

$$\int_0^{2\pi} d\varphi_1 \int_0^{2\pi} d\varphi_2 \exp\left[-i \sum_{k=1}^{3} \Omega_k \cos(\varphi_k - \Delta_k)\right]$$
$$= \sum_p \sum_q \sum_r i^{p+q+r} J_p(-\Omega_1) J_q(-\Omega_2) J_r(-\Omega_3)$$
$$\times \exp[i(r\Phi - p\Delta_1 - q\Delta_2 - r\Delta_3)]$$
$$\times \int_0^{2\pi} \int_0^{2\pi} \exp\{i[(p-r)\varphi_1 + (q-r)\varphi_2]\}\, d\varphi_1\, d\varphi_2. \quad (8)$$

The angular integral in the last line vanishes unless both $p = r$ and $q = r$. Hence, if we observe that $J_n(-x) = (-1)^n J_n(x)$ and omit a factor of $4\pi^2$, the right-hand side of (8) simplifies to

$$Z_{\mathbf{u}}(\Phi) = \sum_{p=-\infty}^{\infty} i^p \left[\prod_{k=1}^{3} J_p(\Omega_k)\right] \exp[ip(\Phi - \Delta)], \quad (9)$$

where $\Delta = \Delta_1 + \Delta_2 + \Delta_3$. The general form of the c.p.d.f. of the three-phase invariant is therefore given by

$$p(\Phi|E_1, E_2, E_3) = K' \sum_{\mathbf{u}} C_{\mathbf{u}} Z_{\mathbf{u}}(\Phi), \quad (10)$$

where K' is a normalization constant, $Z_{\mathbf{u}}(\Phi)$ is defined by (5), (6) and (9), and the Fourier coefficient $C_{\mathbf{u}}$ is to be evaluated for the (crystallographic and/or non-crystallographic) symmetry under consideration. Straightforward integration of (10) over Φ shows that the (real) normalization constant is given by

$$K' = \left\{2\pi \mathcal{R}e\left[\sum_{\mathbf{u}} C_{\mathbf{u}} \mathbf{G}_0\right]\right\}^{-1}, \quad (11)$$

where $\mathbf{G}_0$ is given by (*A*.4).

Example: space group P1

The c.p.d.f. (10) depends on the space-group symmetry *via* the Fourier coefficient $C_{\mathbf{u}}$ and we proceed

to show this dependence for the simplest non-centrosymmetric space group $P1$. Since the Fourier coefficient is just the value of the characteristic function at a point depending in a certain way on the summation indices (*e.g.* Shmueli & Weiss, 1985), the general expression for this coefficient is given by

$$C_{\mathbf{u}} = \left\langle \exp\left[i \sum_{k=1}^{3} (\omega_{2k-1} A_k + \omega_{2k} B_k) \right] \right\rangle, \quad (12)$$

where $\omega_m = \pi \alpha u_m$. The expressions for A_k and B_k, for the space group $P1$, are given by

$$A_k = \sum_{j=1}^{N} n_j \cos \Theta_{jk} \quad \text{and} \quad B_k = \sum_{j=1}^{N} n_j \sin \Theta_{jk}, \quad (13)$$

where

$$n_j = f_j \Big/ \left(\sum_{k=1}^{N} f_k^2 \right)^{1/2} \simeq Z_j \Big/ \left(\sum_{k=1}^{N} Z_k^2 \right)^{1/2} \quad (14)$$

is the normalized scattering factor of the jth atom, f_j being its conventional scattering factor, and $\Theta_{jk} = 2\pi \mathbf{h}_k . \mathbf{r}_j$, here $\mathbf{h}_1 \equiv \mathbf{h}$, $\mathbf{h}_2 \equiv \mathbf{k}$ and $\mathbf{h}_3 \equiv -\mathbf{h} - \mathbf{k}$. The atomic number approximation in (14) is exact for the equal-atom case, and can still be conveniently employed for moderate atomic heterogeneities. If we make use of (13) and of the assumption that the atomic contributions are independent, (12) can be rearranged to read

$$C_{\mathbf{u}} = \prod_{j=1}^{N} \left\langle \exp\left[i \sum_{k=1}^{3} \Lambda_{jk} \cos (\Theta_{jk} - \Delta_k) \right] \right\rangle, \quad (15)$$

where

$$\Lambda_{jk} = \pi \alpha n_j (u_{2k-1}^2 + u_{2k}^2)^{1/2} \quad (16)$$

and Δ_k is given by (6). The atomic average in (15) can be evaluated analogously to the integration of the phase factor outlined earlier. The exponential is a product of three terms of the form $\exp(ix \cos \beta)$, and each of these can be replaced by a series of Bessel functions (7). It is further observed that $\Theta_{j3} = -\Theta_{j1} - \Theta_{j2}$. Some straightforward algebra then leads to

$$C_{\mathbf{u}}(\Delta) = \prod_{j=1}^{N} C_{\mathbf{u},j}(\Delta), \quad (17)$$

where

$$C_{\mathbf{u},j}(\Delta) = \sum_{q=-\infty}^{\infty} i^{3q} \left[\prod_{k=1}^{3} J_q(\Lambda_{jk}) \right] \exp(-iq\Delta), \quad (18)$$

where $\Delta = \Delta_1 + \Delta_2 + \Delta_3$ and $\Delta_k = \tan^{-1}(u_{2k}/u_{2k-1})$.

We point out that our expression for the characteristic function is analogous to that obtained by Peschar & Schenk (1986), except that the latter authors confine their derivation to the equal-atom case and the present equation (18) is applicable to any atomic composition. Another difference between the treatment of Peschar & Schenk and ours is that the latter authors derive approximations to the c.p.d.f. of Φ by evaluating a sixfold Fourier integral transform of a suitably expanded and truncated characteristic function, while our starting point is a sixfold Fourier series. The coefficients of this series are given in terms of the exact rather than approximate characteristic function and the series is then evaluated numerically (Shmueli, Rabinovich & Weiss, 1989). Some practical aspects of bringing this rather formidable summation to manageable form are considered in the next section.

Reduction of the exact result to Cochran's (1955) approximation

Approximate results for the various p.d.f's in structure-factor statistics, in particular those based on the central limit theorem, can be obtained by retaining only the terms of lowest order in the expansion for the appropriate characteristic function and by performing the required Fourier inversion (*e.g.* Shmueli & Weiss, 1985). It should therefore be possible to reduce our exact expression for the c.p.d.f. of the three-phase invariant to the well known Cochran's (1955) approximation to this c.p.d.f. Let us consider the Fourier inversion integral

$$p(\mathbf{E}) = [1/(2\pi)^6] \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} C(\boldsymbol{\omega}) \exp(-i\boldsymbol{\omega} . \mathbf{E}) \, \mathrm{d}^6\boldsymbol{\omega}, \quad (19)$$

and approximate to $C(\boldsymbol{\omega})$ by keeping only the lowest-order terms of its real and imaginary parts. In the present case these can be found from (18) or, more conveniently, from the terms $\mathbf{K}_{0,j}$ and $\mathbf{K}_{1,j}$ in (*A*.8). The lowest-order terms in the expression for the atomic contribution to the characteristic function are

$$C_j(\boldsymbol{\omega}) \simeq \prod_{k=1}^{3} J_0(n_j \Gamma_k) - 2i \left[\prod_{k=1}^{3} J_1(n_j \Gamma_k) \right] \cos \Delta, \quad (20)$$

where $\Gamma_k = (\omega_{2k-1}^2 + \omega_{2k}^2)^{1/2}$. When the approximations $J_0(x) \simeq 1 - x^2/4$ and $J_1(x) \simeq x/2$ are made, one finds from (20) the further approximation

$$C_j(\boldsymbol{\omega}) \simeq 1 - (n_j^2/4)(\Gamma_1^2 + \Gamma_2^2 + \Gamma_3^2) - i(n_j^3/4)\Gamma_1\Gamma_2\Gamma_3 \cos \Delta. \quad (21)$$

Since

$$\cos \Delta = \cos(\Delta_1 + \Delta_2 + \Delta_3) = \mathbf{V}(\boldsymbol{\omega})/\Gamma_1\Gamma_2\Gamma_3, \quad (22)$$

where

$$\mathbf{V}(\boldsymbol{\omega}) = \omega_1\omega_3\omega_5 - \omega_2\omega_4\omega_5 - \omega_2\omega_3\omega_6 - \omega_1\omega_4\omega_6, \quad (23)$$

the atomic contribution to the characteristic function can be approximated by

$$C_j \simeq 1 - (\boldsymbol{\omega} . \boldsymbol{\omega}/4) n_j^2 - i[\mathbf{V}(\boldsymbol{\omega})/4] n_j^3. \quad (24)$$

If we make use of the approximation $1 - x \simeq \exp(-x)$,

we find

$$C(\omega) \simeq \exp\left[-\omega \cdot \omega/4 - i[V(\omega)/4] \sum_{j=1}^{N} n_j^3\right], \quad (25)$$

which agrees with the expression given by Giacovazzo (1980). One can then follow Giacovazzo's (1980) derivation to find that to the order given

$$p(\Phi|E_1E_2E_3) \simeq [2\pi I_0(\kappa)]^{-1} \exp(\kappa \cos \Phi), \quad (26)$$

where

$$\kappa = 2\left(\sum_{j=1}^{N} n_j^3\right) E_1E_2E_3$$
$$= 2\sigma_3\sigma_2^{-3/2}E_1E_2E_3, \quad (27)$$

with

$$\sigma_n = \sum_{j=1}^{N} f_j^n.$$

Equation (26) reduces to the Cochran (1955) result for the equal-atom case, *i.e.* for $n_1 = \ldots = n_N = N^{-1/2}$. Higher-order corrections may be obtained by considering further terms in (*A*.6) and (*A*.7).

Simplification of the expressions and some programming considerations

Numerical computation of the expression for the c.p.d.f. (10) presents considerable practical difficulties because of prohibitive computing times. However, the computing time can be reduced by several orders of magnitude by making use of the symmetries inherent in the summations. In what follows we briefly survey several such possible simplifications.

(1) It is convenient to decompose (10) into the different kinds of summations involving five, four, three, two, one and no zero indices respectively. This leads to more efficient and more readily debugged programs.

(2) The arguments of the Bessel functions appearing in the Fourier coefficient C_u and in the phase factor Z_u of our c.p.d.f. each depend on the sum of the squares of two summation indices of the sixfold Fourier series. This indicates that the summation can be restricted to non-negative indices only, provided the phase factors [*e.g.* (6)] are appropriately modified.

(3) The Fourier summation indices appear in pairs [(1, 2), (3, 4) and (5, 6)] both in the arguments of the Bessel functions and in the phase factors [*e.g.* (6)]. If, for example, u_1 and u_2 both run over all the positive integers, we may try to reduce the corresponding pair of summations to the range $u_1 > u_2$. This is straightforward in the case of the Bessel functions, but requires some bookkeeping in the case of the phases. The resulting relations allow one to restrict the general range of summation to $u_1 \geq u_2$, $u_3 \geq u_4$ and $u_5 \geq u_6$.

(4) The 'intra-pair' symmetry outlined in point 3 usually leads to several kinds of phase shifts, as shown in Appendix *A*. However, an examination of the trigonometric functions of these seemingly different phases often reduces the number of different terms to be summed.

(5) As pointed out above, the sixfold summation essentially reduces to summations over three pairs of indices. These pairs can often be interchanged, subject to a proper symmetrization of the Bessel function products that depend on (different) magnitudes of normalized structure factors, while leaving the value of the sixfold summation unchanged. This leads to another significant reduction of the computational effort.

(6) Major simplifications, confined to the programming of the computations, are brought about by avoiding repetitive calculations. For example, the required Bessel functions need only be computed once outside the nested summation loop. Such a simplification may call for rather large arrays, possibly requiring a computer with virtual memory. Casting the expressions into vectorizable form may result in some savings in time but this appears to be much less important than the reduction of the computing time resulting from the symmetries outlined above.

The above outlined simplifications appear to be a necessary condition for carrying out numerical calculations of the exact c.p.d.f. of the three-phase invariant. They thus deserve some detailed illustrations, which have been postponed to Appendix *A*.

Concluding remarks

We have derived an exact representation of the c.p.d.f. of a three-phase invariant in terms of a Fourier series for the space group $P1$. However, the derivation can be generalized to derive representations of c.p.d.f's of higher invariants for higher symmetries. While the resulting formulation is fairly concise, its numerical implementation poses serious difficulties, and becomes possible only when the symmetry inherent in the summation is properly exploited. In fact, much of the present work has been concerned with simplification of the expressions to be evaluated and the computations, the results of which are presented and discussed in the second paper of this series (Shmueli, Rabinovich & Weiss, 1989), have thereby been reduced to manageable proportions.

Analogous Fourier representations of c.p.d.f.'s of quartet, quintet *etc.* invariants can be readily constructed but we do not deal with them here since the relevant computational difficulties still appear to be too serious.

This research was supported in part by grant No. 84-00076 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. All the

computations related to this paper were carried out at the Tel Aviv University Computation Center on a Cyber 180-990 computer.

APPENDIX A

We present here an example of simplifying the summations involved in the computation of the c.p.d.f. of the three-phase invariant and try to illustrate most of the programming considerations outlined in the paper and implemented in actual computations. We first require a decomposition of the complex quantities appearing in our expressions into their real and imaginary parts.

The phase factor (9) of the c.p.d.f. depends on the three-phase invariant Φ and can be decomposed into its real and imaginary parts as

$$Z_{\mathbf{u}}(\Phi) = X_{\mathbf{u}}(\Phi) + iY_{\mathbf{u}}(\Phi), \tag{A.1}$$

where

$$X_{\mathbf{u}}(\Phi) = \mathbf{G}_0 + 2\sum_{p=1}^{\infty} (-1)^p \mathbf{G}_{2p}(\Phi) \cos[2p(\Phi - \Delta)], \tag{A.2}$$

and

$$Y_{\mathbf{u}}(\Phi) = -2\sum_{p=1}^{\infty} (-1)^p \mathbf{G}_{2p-1}(\Phi) \cos[(2p-1)(\Phi - \Delta)], \tag{A.3}$$

with

$$\mathbf{G}_m = \prod_{k=1}^{3} J_m(\Omega_k), \tag{A.4}$$

where Ω_k is given by equation (5).

Similarly, the jth component of the characteristic function, given by (18), can be decomposed into its real and imaginary parts as

$$C_{\mathbf{u},j} = R_{\mathbf{u},j} + iI_{\mathbf{u},j}, \tag{A.5}$$

where

$$R_{\mathbf{u},j} = \mathbf{K}_{0,j} + 2\sum_{q=1}^{\infty} (-1)^q \mathbf{K}_{2q,j} \cos(2q\Delta) \tag{A.6}$$

and

$$I_{\mathbf{u},j} = 2\sum_{q=1}^{\infty} (-1)^q \mathbf{K}_{2q-1,j} \cos[(2q-1)\Delta] \tag{A.7}$$

with

$$\mathbf{K}_{m,j} = \prod_{k=1}^{3} J_m[\pi\alpha n_j (u_{2k-1}^2 + u_{2k}^2)^{1/2}]. \tag{A.8}$$

Summation over positive indices only

Suppose we restrict our attention to non-zero summation indices. There are altogether $2^6 = 64$ sign combinations for any given sextet of the summation indices. Hence, for any given six magnitudes of the indices u_k in (10) there are 64 terms in that summation that depend on the given set of index magnitudes but with all their possible combinations of signs. The dependence of the c.p.d.f. on the summation indices is restricted to (i) the arguments of the Bessel functions, which are of the form $\kappa(u_{2m}^2 + u_{2m-1}^2)^{1/2}$ and depend on magnitudes only, and (ii) the phase factors $\Delta = \Delta_1 + \Delta_2 + \Delta_3$, where Δ_k is given in (6). Clearly, any Δ_k changes its sign when only *one* of the u's changes sign, and does not change its sign when either *both* or *none* of the u's change sign. There are therefore eight sign combinations of the u's leaving Δ unchanged, and eight possible forms of the phase Δ,

$$\delta_i = \pm|\Delta_1| \pm |\Delta_2| \pm |\Delta_3|, \tag{A.9}$$

where the values 0, 1, 2, 3, 4, 5, 6 and 7 of the index i in (A.9) correspond to the sign combinations: +++, −++, +−+, ++−, +−−, −+−, −−+ and −−−, respectively. The simplification that immediately suggests itself is to group the products $C_{\mathbf{u}}Z_{\mathbf{u}}$, internally, for the above eight values of the phase, and carry out the sixfold summation on positive values of the indices only. The c.p.d.f. (10) thus becomes

$$p(\Phi|E_1, E_2, E_3) \propto \sum_{\mathbf{u}>0} \sum_{s=0}^{7} C_{\mathbf{u}}(\delta_s) Z_{\mathbf{u}}(\Phi, \delta_s). \tag{A.10}$$

Further simplifications require a consideration of the dependence of $C_{\mathbf{u}}$ and $Z_{\mathbf{u}}$ on the phase Δ, and the symmetry of the phases as in (A.9). The eight forms of the phase in (A.9) split in two groups of four, which are related by a change of the sign of Δ. It is seen from (A.6) and (A.7) that $C_{\mathbf{u}}$ is independent of the sign of Δ, but (A.2) and (A.3) show that $Z_{\mathbf{u}}$ changes its value when Δ changes sign. Equation (A.10) can therefore be rewritten as

$$p(\Phi|E_1, E_2, E_3) \propto \sum_{\mathbf{u}>0} \sum_{s=0}^{3} C_{\mathbf{u}}(\delta_s) \times [Z_{\mathbf{u}}(\Phi, \delta_s) + Z_{\mathbf{u}}(\Phi, -\delta_s)]. \tag{A.11}$$

The sum in the square brackets in (A.11) reduces to

$$Z_{\mathbf{u}}(\Phi, \delta) + Z_{\mathbf{u}}(\Phi, -\delta) \equiv 2W_{\mathbf{u}}(\Phi, \delta) = 2[U_{\mathbf{u}}(\Phi, \delta) + iV_{\mathbf{u}}(\Phi, \delta)], \tag{A.12}$$

where

$$U_{\mathbf{u}}(\Phi, \delta) = \mathbf{G}_0 + 2\sum_{p=1}^{\infty} (-1)^p \mathbf{G}_{2p} \cos(2p\Phi) \cos(2p\delta) \tag{A.13}$$

and

$$V_{\mathbf{u}}(\Phi, \delta) = -2\sum_{p=1}^{\infty} (-1)^p \mathbf{G}_{2p-1} \times \cos[(2p-1)\Phi] \cos[(2p-1)\delta] \tag{A.14}$$

and (*A*.11) further reduces to

$$p(\Phi|E_1, E_2, E_3) \propto \sum_{\mathbf{u}>0} \sum_{s=0}^{3} C_{\mathbf{u}}(\delta_s) W_{\mathbf{u}}(\Phi, \delta_s). \quad (A.15)$$

Another consequence of these considerations is the (expected) symmetry of the c.p.d.f. with respect to a change of the sign of Φ. Simplifications similar to those of the last paragraph follow readily for the summations having one or more indices equal to zero. Note that we omitted here explicit reference to the normalization constant, which can be computed from (11) at any stage of the symmetrization/simplification.

Intra-pair symmetry and other simplifications

As pointed out above, the characteristic function $C_{\mathbf{u}}$ and the modified and integrated phase factor $W_{\mathbf{u}}(\Phi)$ of the Fourier c.p.d.f. depend on the summation indices *via* the arguments of the Bessel functions, and the phases Δ_k defined in (6). For each pair of indices (u_{2k-1}, u_{2k}) we have either $u_{2k-1} = u_{2k}$, with $\Delta_k = \pi/4$, or $u_{2k-1} \neq u_{2k}$. If we consider all the positive indices, as outlined in the previous section, we encounter, for any pair of unequal indices (u_{2k-1}, u_{2k}), the pair of indices (u_{2k}, u_{2k-1}). The phases of these two pairs are obviously related by

$$\begin{aligned}\Delta'_k &= \tan^{-1}(u_{2k-1}/u_{2k}) \\ &= \cot^{-1}(u_{2k}/u_{2k-1}) = (\pi/2) - \Delta_k. \end{aligned} \quad (A.16)$$

We can therefore confine the summations on u_1 and u_2 to $u_1 > u_2$, and treat in a similar manner the other pairs of summations on (u_3, u_4) and (u_5, u_6).

An example of this intra-pair symmetry is shown below by considering together the following configurations of the summation indices: (1) $u_1 \neq u_2$, $u_3 = u_4$ and $u_5 = u_6$, (ii) $u_1 = u_2$, $u_3 \neq u_4$ and $u_5 = u_6$, and (iii) $u_1 = u_2$, $u_3 = u_4$ and $u_5 \neq u_6$, where all the indices are positive. The appropriate version of the summation in (*A*.15) is

$$\sum_{u_1 \neq u_2=1}^{\infty} \sum_{u_3,u_5=1}^{\infty} \sum_{s=0}^{3} C_{u_1u_2u_3u_5}(\delta_s) W'_{u_1u_2u_3u_5}(\Phi, \delta_s), \quad (A.17)$$

where $C_{u_1u_2u_3u_5}$ [an abbreviation for $C_{u_1u_2u_3u_3u_5u_5}$] is defined as in (17), and $W'_{u_1u_2u_3u_5}(\Phi)$ is defined as in (*A*.2)-(*A*.4) except that the quantity $\mathbf{G}_m$ that appears in those equations is to be replaced by

$$\mathbf{G}^3_m = B^m_{112}B^m_{233}B^m_{355} + B^m_{212}B^m_{333}B^m_{155} + B^m_{312}B^m_{133}B^m_{255}, \quad (A.18)$$

where

$$B^m_{skl} = \begin{cases} J_m[\pi\alpha E_s(u_k^2 + u_l^2)^{1/2}] & \text{if } k \neq l \\ J_m(\pi a E_s u_k \sqrt{2}) & \text{if } k = l. \end{cases} \quad (A.19)$$

Equations (*A*.18) and (*A*.19) account for the three index configurations, while only the first one appears in the summation. The fact that there are two pairs of equal indices introduces into the phase a constant term.

We now have to consider the four phases, δ_s, for the cases (i) $u_1 > u_2$, and (ii) $u_1 < u_2$, and will take (i) as the reference case. We thereby obtain eight phases, only four of which are different (up to a change of the sign): (i) $\Delta_1 - \pi$, (ii) $\Delta_1 - \pi/2$, (iii) Δ_1 and (iv) $\Delta_1 + \pi/2$, with multiplicities of 1, 3, 3 and 1, respectively. If we group internally the expressions that depend on these phases, the summation in (*A*.17) reduces to

$$\sum_{u_1>u_2=1}^{\infty} \sum_{u_3,u_5=1}^{\infty} \sum_{s=0}^{3} \binom{3}{s} [C_{u_1u_2u_3u_5}(\Delta_1 - \pi + \pi s/2) \times W'_{u_1u_2u_3u_5}(\Phi, \Delta_1 - \pi + \pi s/2)]. \quad (A.20)$$

The number of different types of products $C_{\mathbf{u}}W'_{\mathbf{u}}(\Phi)$ in (*A*.20) can be further reduced from four to two by evaluating the functions $\cos(2n\delta_s)$ and $\cos[(2n-1)\delta_s]$ for the different values of the phases. One sees that if the phase Δ_1 in $C_{\mathbf{u}}(\Delta_1)W'_{\mathbf{u}}(\Phi, \Delta_1)$ is replaced by $\Delta_1 - \pi$, the product transforms to its complex conjugate. If Δ_1 in $C_{\mathbf{u}}(\Delta_1)W'_{\mathbf{u}}(\Phi, \Delta_1)$ is replaced by $\Delta_1 + \pi/2$ another type of product is obtained, which we denote by $\tilde{C}_{\mathbf{u}}(\Delta_1)\tilde{W}'_{\mathbf{u}}(\Phi, \Delta_1)$ where, in the space group $P1$,

$$\tilde{C}_{\mathbf{u}}(\Delta) = \prod_{j=1}^{N} \tilde{C}_{\mathbf{u},j}, \quad (A.21)$$

where

$$\mathscr{R}e[\tilde{C}_{\mathbf{u},j}(\Delta)] = \mathbf{K}_{0,j} + 2\sum_{q=1}^{\infty} \mathbf{K}_{2q,j} \cos(2q\Delta), \quad (A.22)$$

$$\mathscr{I}m[\tilde{C}_{\mathbf{u},j}(\Delta)] = 2\sum_{q=1}^{\infty} \mathbf{K}_{2q-1,j} \cos[(2q-1)\Delta], \quad (A.23)$$

$$\mathscr{R}e[\tilde{W}'_{\mathbf{u}}(\Phi, \Delta)] = \mathbf{G}^3_0 + 2\sum_{p=1}^{\infty} \mathbf{G}^3_{2p} \cos(2p\Phi) \cos(2p\Delta), \quad (A.24)$$

and

$$\mathscr{I}m[\tilde{W}'_{\mathbf{u}}(\Phi,\Delta)] = -2\sum_{p=1}^{\infty} \mathbf{G}^3_{2p-1} \cos[(2p-1)\Phi] \times \cos[(2p-1)\Delta], \quad (A.25)$$

where $\mathbf{K}_{m,j}$ is defined by (*A*.48) and $\mathbf{G}^3_m$ is given by (*A*.18). When, finally, the phase Δ_1 in $C_{\mathbf{u}}(\Delta_1)W'_{\mathbf{u}}(\Phi, \Delta_1)$ is replaced by $\Delta_1 - \pi/2$, the product transforms to the complex conjugate of $\tilde{C}_{\mathbf{u}}(\Delta_1)\tilde{W}'_{\mathbf{u}}(\Phi, \Delta_1)$. The (relevant) real part of the inner summation in (*A*.20) thus becomes

$$4\mathscr{R}e\{[C_{\mathbf{u}}(\Delta_1)W'_{\mathbf{u}}(\Phi, \Delta_1)] + [\tilde{C}_{\mathbf{u}}(\Delta_1)\tilde{W}'_{\mathbf{u}}(\Phi, \Delta_1)]\}. \quad (A.26)$$

There are still eight products in the summation but only two of them differ.

Inter-pair symmetry

Consider the summation over six non-zero indices, where none of (u_1, u_2), (u_3, u_4) and (u_5, u_6) is a pair of equal numbers. This case is the most time-consuming part of the calculation. As explained above, the summation can be restricted to positive indices only, where $u_1 > u_2$, $u_3 > u_4$ and $u_5 > u_6$. Suppose the sixfold summation converges when the upper limit of a single summation is some number, say M, and each of the indices u_1, u_3 and u_5 ranges from 1 to M. It is easily verified that when the three $|E|$ values are equal, the above index configuration leads to some redundant calculations the result of which is invariant under a permutation of the pairs of indices. Accordingly, in order to simplify the calculation it is convenient to define new indices as

$$\nu_k = \tfrac{1}{2}(u_{2k-1} - 1)u_{2k-1} + u_{2k}, \quad k = 1, 2, 3. \quad (A.27)$$

It is then sufficient to let u_1 range from 1 to M, u_3 from 1 to u_1 and u_5 from 1 to u_3. We further introduce multiplicity factors depending on whether ν_1, ν_2, ν_3 are all different, $\nu_1 = \nu_2 \neq \nu_3$, $\nu_1 \neq \nu_2 = \nu_3$ or $\nu_1 = \nu_2 = \nu_3$. The E-dependent Bessel functions are stored outside the summation loop in arrays of the form

$$D_j(p, \nu_k) = J_p[\pi\alpha E_j(u_{2k-1}^2 + u_{2k}^2)^{1/2}] \quad (A.28)$$

and the triple products such as those in (*A*.4) can be computed as

$$T_{123} = D_1(p, \nu_1)D_2(p, \nu_2)D_3(p, \nu_3), \quad (A.29)$$

where the subscripts on T pertain to the subscripts on the $|E|$ values. In the general case, *i.e.* for unequal $|E|$'s, we can still retain the restricted ranges of the summations by symmetrizing T_{123}. This is simply achieved by computing the expression

$$\tfrac{1}{6}(T_{123} + T_{312} + T_{231} + T_{132} + T_{321} + T_{213}). \quad (A.30)$$

Since only some of the terms require such a symmetrization, the computing effort is thus again significantly reduced.

References

CASTLEDEN, I. R. (1987). *Acta Cryst.* A**43**, 384-393.
COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.
GIACOVAZZO, C. (1974). *Acta Cryst.* A**30**, 626-630.
GIACOVAZZO, C. (1980). *Direct Methods in Crystallography.* London: Academic Press.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Table of Integrals, Series, and Products.* New York: Academic Press.
HAUPTMAN, H. (1971). *Z. Kristallogr.* **134**, 28-43.
HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* ACA Monogr. No. 3. Pittsburgh: Polycrystal Book Service.
KARLE, J. (1972). *Acta Cryst.* B**28**, 3362-3369.
KARLE, J. & GILARDI, R. D. (1973). *Acta Cryst.* A**29**, 401-407.
NAYA, S., NITTA, I. & ODA, T. (1965). *Acta Cryst.* **19**, 734-747.
PESCHAR, R. & SCHENK, H. (1986). *Acta Cryst.* A**42**, 309-317.
SHMUELI, U., RABINOVICH, S. & WEISS, G. H. (1989). *Acta Cryst.* A**45**, 367-371.
SHMUELI, U. & WEISS, G. H. (1985). *Acta Cryst.* A**41**, 401-408.
SHMUELI, U. & WEISS, G. H. (1986). *Acta Cryst.* A**42**, 240-246.
SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* A**40**, 651-660.

Acta Cryst. (1989). A**45**, 367-371

Exact Conditional Distribution of a Three-Phase Invariant in the Space Group *P*1. II. Calculations and Comparison with the Cochran Approximation

BY URI SHMUELI AND SAVELY RABINOVICH

School of Chemistry, The Sackler Faculty of Sciences, Tel Aviv University, 69 978 Tel Aviv, Israel

AND GEORGE H. WEISS

Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20892, *USA*

(*Received* 7 *June* 1988; *accepted* 3 *January* 1989)

Abstract

The conditional probability density function of a three-phase invariant is computed from exact expressions derived and discussed in the first paper of this series [Shmueli, Rabinovich & Weiss (1989). *Acta Cryst.* A**45**, 361-367] and comparisons are presented of these computations with the approximate conditional density due to Cochran [*Acta Cryst.* (1955), **8**, 473-478]. Conditional variances computed from the exact and approximate expressions are also compared. The computations are carried out for the space group *P*1. This is the first numerical comparison of conditional phase-invariant statistics evaluated from exact and approximate expressions. The discrepancy between these two kinds of statistics appears to be negligible if the E values involved are small and the number of atoms in the cell is moderately

0108-7673/89/060367-05$03.00