STEURER, W. (1991). *J. Phys. Condens. Matter*, **3**, 3397-3410.
STEURER, W. & MAYER, J. (1989). *Acta Cryst.* B**45**, 355-359.
YAMAMOTO, A. & ISHIHARA, K. N. (1988). *Acta Cryst.* A**44**, 707-714.

Acta Cryst. (1992). A**48**, 901-906

The Joint Probability Distribution of Any Set of Phases Given Any Set of Diffraction Magnitudes. I. Theoretical Considerations

BY C. GIACOVAZZO

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche CNR, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 *Bari, Italy*

M. C. BURLA

Dipartimento di Scienze della Terra, Università, 06100 *Perugia, Italy*

AND G. CASCARANO

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche CNR, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 *Bari, Italy*

(*Received* 13 *December* 1992; *accepted* 6 *May* 1992)

Abstract

Conditional joint probability distribution functions $P(\varphi_1, \ldots, \varphi_n | R_1, \ldots, R_n, \ldots, R_p)$ of any set of n phases given any set of p diffraction moduli are calculated. The distributions include terms up to order $1/N$ and involve both triplet and quartet contributions. Two types of formulae are derived, which may be considered as developments of two mathematical approaches described by Hauptman [*Acta Cryst.* (1975). A**31**, 680-687] and by Giacovazzo [*Acta Cryst.* (1975). A**31**, 252-259; *Acta Cryst.* (1976). A**32**, 91-99] for the estimation of the quartet invariants.

1. Introduction

The discovery (Hauptman & Karle, 1953) of the properties of the structure invariants and seminvariants has played a crucial role in the solution of the phase problem. Their estimation was the mainspring for the development of the joint probability distribution methods (Hauptman & Karle, 1953; Klug 1958). Such methods rely on the idea that certain combinations of phases (*i.e.* the structure invariants and seminvariants) can be estimated when the related structure factors have their observed values.

More recently, this point of view has been generalized by the neighbourhood principle (Hauptman, 1975, 1978) and by the representation method (Giacovazzo, 1977*a*, 1980*a*). Such contributions extended the range of application of direct methods; indeed, single n-phase structure invariants could be estimated *via* the overall prior information provided by p moduli of structure factors, where p may be much larger than n. Asymptotically, p may coincide with the number of measured diffraction magnitudes. The standard technique is as follows.

(i) The joint probability distribution function

$$P(\varphi_1, \ldots, \varphi_n, \ldots, \varphi_p, R_1, \ldots, R_p) \quad (1)$$

is first calculated.

(ii) The marginal distribution

$$P(\varphi_1, \ldots, \varphi_n, R_1, \ldots, R_n, \ldots, R_p) \quad (2)$$

with $n < p$ is derived. Accordingly,

$$P(\varphi_1, \ldots, \varphi_n, R_1, \ldots, R_n, \ldots, R_p) = \int \ldots \int P(\varphi_1, \ldots, \varphi_n, \ldots, \varphi_p, R_1, \ldots, R_n, \ldots, R_p)\, d\varphi_{n+1} \ldots d\varphi_p.$$

(iii) The conditional distribution

$$P(\varphi_1, \ldots, \varphi_n | R_1, \ldots, R_p) \quad (3)$$

is derived, where $\varphi_1, \ldots, \varphi_n$ are the phases that compose the n-phase structure invariant

$$\Phi = \varphi_1 + \varphi_2 + \ldots + \varphi_n$$

that one wishes to estimate.

(iv) The conditional distribution

$$P(\Phi | R_1, \ldots, R_p) \quad (4)$$

is obtained, which provides the desired estimate of Φ.

In this paper we will focus our attention on distributions (3) characterized by large values of n. The aim is not that of deriving estimates of single n-phase

0108-7673/92/060901-06$06.00

structure invariants or seminvariants [as from (4)], but that of obtaining joint probability distributions of any set of n phases given any set of p moduli. In particular, n may be the dimension of any subset of the measured reflections characterized by large values of R and/or by any other useful condition.

The proposed distributions will enclose terms up to order N^{-1}. As shown in the following paper (Burla, Cascarano & Giacovazzo, 1992), the properties of such distributions are of great theoretical interest and can provide useful suggestions for practical applications.

2. The distribution $P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)$ in $P1$ and $P\bar{1}$

Deriving a general mathematical expression for the joint probability distribution (3) that includes terms up to order N^{-1} and is valid in all the space groups is not a trivial task. We will first focus our attention on the space groups $P1$ and $P\bar{1}$; the results will then be extended to space groups of higher symmetry. We will follow the procedure used by Giacovazzo (1980*b*) to derive, from a seven-variate distribution, the conditional probability of a quartet phase given seven moduli and a variable number (up to seven) of phases. Giacovazzo's procedure enables the estimation of a quartet phase when, as well as moduli, phases of some cross terms are *a priori* known. Here we are not interested in triplet or quartet phases; we aim to describe a joint probability distribution of the individual phases. In particular:

(i) The distributions $P(\varphi_1,\ldots,\varphi_7,R_1,\ldots,R_7)$ obtained by Hauptman and by Giacovazzo for the estimation of the quartet invariants will be recalled. Hauptman's (1975) results were obtained *via* the exponential form of the characteristic function while Giacovazzo's (1975, 1976) formulae were derived by means of the Gram-Charlier expansion of the characteristic function. The two distributions do not coincide and will be examined separately; they lead to different expressions for (3).

(ii) Marginal and conditional distributions of $P(\varphi_1,\ldots,\varphi_7,R_1,\ldots,R_7)$ will be recalled to discover how the coefficients of the various terms in the distribution depend on the prior phase information.

(iii) Generalized coefficients will then be derived and used for constructing a general mathematical expression for the distribution (3). Expressions for $P1$ and $P\bar{1}$ will be separately described.

(*a*) *Space group P*1. *Extension of the Hauptman formulation*

For brevity we will use the notation

$$E_1=E_{\mathbf{h}},\quad E_2=E_{\mathbf{k}},\quad E_3=E_{\mathbf{l}},\quad E_4=E_{\mathbf{h+k+l}},$$
$$E_5=E_{\mathbf{h+k}},\quad E_6=E_{\mathbf{h+l}},\quad E_7=E_{\mathbf{k+l}}.$$

Then, according to Hauptman (1975), we have

$$\begin{aligned}
&P(\varphi_1,\ldots,\varphi_7,R_1,\ldots,R_7)\\
&=\prod_{i=1}^{7}[(R_i/\pi)\exp(-R_i^2)]\\
&\times\exp\{(2/N^{1/2})[R_1R_2R_5\cos(\varphi_1+\varphi_2-\varphi_5)\\
&+R_1R_3R_6\cos(\varphi_1+\varphi_3-\varphi_6)\\
&+R_1R_4R_7\cos(\varphi_1-\varphi_4+\varphi_7)\\
&+R_2R_3R_7\cos(\varphi_2+\varphi_3-\varphi_7)\\
&+R_2R_4R_6\cos(\varphi_2-\varphi_4+\varphi_6)\\
&+R_3R_4R_5\cos(\varphi_3-\varphi_4+\varphi_5)]\\
&-(2/N)[R_1R_2R_6R_7\cos(\varphi_1-\varphi_2-\varphi_6+\varphi_7)\\
&+R_1R_3R_5R_7\cos(\varphi_1-\varphi_3-\varphi_5+\varphi_7)\\
&+R_1R_4R_5R_6\cos(\varphi_1+\varphi_4-\varphi_5-\varphi_6)\\
&+R_2R_3R_5R_6\cos(\varphi_2-\varphi_3-\varphi_5+\varphi_6)\\
&+R_2R_4R_5R_7\cos(\varphi_2+\varphi_4-\varphi_5-\varphi_7)\\
&+R_3R_4R_6R_7\cos(\varphi_3+\varphi_4-\varphi_6-\varphi_7)\\
&+2R_1R_2R_3R_4\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)]\}.
\end{aligned}\tag{5}$$

The following list shows basis and cross terms for each quartet involved in distribution (5):

$$\begin{array}{ll}
\varphi_1-\varphi_2-\varphi_6+\varphi_7 & E_{\mathbf{h-k}},E_3,E_4;\\
\varphi_1-\varphi_3-\varphi_5+\varphi_7 & E_{\mathbf{h-l}},E_2,E_4;\\
\varphi_1+\varphi_4-\varphi_5-\varphi_6 & E_{\mathbf{2h-k+l}},E_2,E_3;\\
\varphi_2-\varphi_3-\varphi_5+\varphi_6 & E_{\mathbf{k-l}},E_1,E_4;\\
\varphi_2+\varphi_4-\varphi_5-\varphi_7 & E_{\mathbf{h+2k+l}},E_1,E_3;\\
\varphi_3+\varphi_4-\varphi_6-\varphi_7 & E_{\mathbf{h+k+2l}},E_1,E_2;\\
\varphi_1+\varphi_2+\varphi_3-\varphi_4 & E_5,E_6,E_7.
\end{array}\tag{6}$$

It is easily seen from (6) that the first six quartets are characterized by two *a priori* known cross magnitudes: in (5), the numerical coefficient $-2/N$ is associated with them. The last quartet in (6) is characterized by three *a priori* known cross magnitudes and has coefficient $-4/N$. Additional calculations, which for brevity are not shown here [refer to Giacovazzo (1977*b*) and Heinerman (1977)] indicate that: (i) in the marginal distribution $P(\varphi_1,\ldots,\varphi_5,R_1,\ldots,R_5)$, where only one cross magnitude of the quartet $(\varphi_1+\varphi_2+\varphi_3-\varphi_4)$ is involved, the coefficient of $\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)$ in the exponential term is zero; (ii) in the marginal distribution $P(\varphi_1,\ldots,\varphi_4,R_1,\ldots,R_4)$, where no cross magnitudes of $(\varphi_1+\varphi_2+\varphi_3-\varphi_4)$ are involved, the coefficient of $\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)$ is $2/N$.

The following rule may thus be obtained: in distributions of type (2), quartet terms with 3, 2, 1, 0 cross magnitudes *a priori* known in modulus and phase have coefficients $-4/N$, $-2/N$, 0, $2/N$, respectively. Generally, the contribution to (2) due to the general quartet $(\varphi_i+\varphi_j+\varphi_l+\varphi_m)$ may be expressed as (indices 5, 6, 7 of the following expression refer to the cross terms of the quartet)

$$\exp\left[(2R_iR_jR_lR_m/N)(w-w_5-w_6-w_7)\times\cos(\varphi_i+\varphi_j+\varphi_l+\varphi_m)\right],$$

where $w=1$, $w_\nu=0$ always except when the cross magnitude R is *a priori* known. In this case, $w_\nu=1$.

To see how a distribution of type (2) can be modelled when a modulus is known but its phase is unknown we consider the marginal distribution

$$\begin{aligned}P(\varphi_1,\ldots,\varphi_6,R_1,\ldots,R_7)&\\ \simeq\prod_{i=1}^{7}&[(R_i/\pi)\exp(-R_i^2)]\\ &\times\exp\{(2/N^{1/2})[R_1R_2R_5\cos(\varphi_1+\varphi_2-\varphi_5)\\ &+R_1R_3R_6\cos(\varphi_1+\varphi_3-\varphi_6)\\ &+R_2R_4R_6\cos(\varphi_2-\varphi_4+\varphi_6)\\ &+R_3R_4R_5\cos(\varphi_3-\varphi_4+\varphi_5)]\\ &-(2/N)[R_1R_4R_5R_6\cos(\varphi_1+\varphi_4-\varphi_5-\varphi_6)\\ &+R_2R_3R_5R_6\cos(\varphi_2-\varphi_3-\varphi_5+\varphi_6)\\ &+2R_1R_2R_3R_4\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)]\}\\ &\times I_0(R_7Z_7)\end{aligned}$$

where

$$Z_7=(2/N^{1/2})[R_2^2R_3^2+R_1^2R_4^2+2R_1R_2R_3R_4\times\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)]^{1/2}.$$

R_7 is a cross term of only one quartet in (5), say $(\varphi_1+\varphi_2+\varphi_3-\varphi_4)$. The lack of information on φ_7 cancels from (5) the cosine terms in which φ_7 is involved as a basis term, generates the new term $I_0(R_7Z_7)$ but does not modify the coefficient of $\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)$ in the exponential term. In conclusion, the overall contribution to (2) arising from a quartet relationship is

$$\begin{aligned}&\exp[(2R_iR_jR_lR_m/N)(w-w_5-w_6-w_7)\\ &\quad\times\cos(\varphi_i+\varphi_j+\varphi_l+\varphi_m)]\\ &\quad\times I_0(w_5'R_5Z_5)I_0(w_6'R_6Z_6)I_0(w_7'R_7Z_7),\end{aligned}$$

where

$$\begin{aligned}Z_5&=(2/N^{1/2})[R_1^2R_2^2+R_3^2R_4^2+2R_1R_2R_3R_4\\ &\quad\times\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)]^{1/2},\\ Z_6&=(2/N^{1/2})[R_1^2R_3^2+R_2^2R_4^2+2R_1R_2R_3R_4\\ &\quad\times\cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)]^{1/2}\end{aligned}$$

w_ν' is always zero, except when R_ν is *a priori* known and φ_ν is unknown; in this case $w_\nu'=1$.

We can now write the general expression of a distribution of type (2),

$$\begin{aligned}P(\varphi_1,\ldots,\varphi_n,R_1,\ldots,R_p)&\\ \simeq\prod_{i=1}^{p}[(R_i/\pi\exp(-R_i^2)]\exp&\Bigg[\sum_{\text{triplets}}T_{ijl}\cos t_{ijl}\\ &+\sum_{\text{quartets}}B_{ijlm}(w-w_5-w_6-w_7)\cos q_{ijlm}\Bigg]\\ &\times\prod_\nu I_0(w_\nu'R_\nu Z_\nu),\end{aligned}\qquad(7)$$

where

$$\begin{aligned}T_{ijl}&=2R_iR_jR_l/N^{1/2},\\ t_{ijl}&=\varphi_i+\varphi_j+\varphi_l,\\ B_{ijlm}&=2R_iR_jR_lR_m/N,\\ q_{ijlm}&=\varphi_i+\varphi_j+\varphi_l+\varphi_m.\end{aligned}$$

The product of I_0 terms in (7) includes all the *a priori* known magnitudes with unknown phase value that are cross terms of some quartet in (7). The general conditional distribution $P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)$ may thus be written as

$$\begin{aligned}P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)&\\ \simeq(1/L)\exp&\Bigg[\sum_{\text{triplets}}T_{ijl}\cos t_{ijl}\\ &+\sum_{\text{quartets}}B_{ijlm}(w-w_5-w_6-w_7)\cos q_{ijlm}\Bigg]\\ &\times\prod_\nu I_0(w_\nu'R_\nu Z_\nu),\end{aligned}\qquad(8)$$

where L is a scale factor whose value does not depend on the phase values.

(*b*) *Space group* $P\bar{1}$. *Extension of the Green & Hauptman formulation*

According to Green & Hauptman (1976),

$$\begin{aligned}P(E_1,\ldots,E_7)\simeq(2\pi)^{-7/2}\exp\{&-\tfrac{1}{2}(E_1^2+\ldots+E_7^2)\\ &+(1/N^{1/2})(E_1E_2E_5+E_1E_3E_6\\ &+E_1E_4E_7+E_2E_3E_7\\ &+E_2E_4E_6+E_3E_4E_5)\\ &-(1/N)(E_1E_2E_6E_7+E_1E_3E_5E_7\\ &+E_1E_4E_5E_6+E_2E_3E_5E_6\\ &+E_2E_4E_5E_7+E_3E_4E_6E_7\\ &+2E_1E_2E_3E_4)\}.\end{aligned}\qquad(9)$$

Data in (6) and considerations similar to those made for the space group $P1$ suggest the following rule: in distributions of type (2), quartet terms with 3, 2, 1, 0 cross magnitudes *a priori* known in modulus and phase have coefficients equal to $-2/N$, $-1/N$, 0, $1/N$, respectively. Generally, the contribution of the quartet $E_iE_jE_lE_m$ to (2) is

$$\exp[(E_iE_jE_lE_m/N)(w-w_5-w_6-w_7)], \quad (10)$$

where $w=1$, $w_\nu=0$ always except when the cross magnitude R_ν is *a priori* known; in this case, $w_\nu=1$.

To model a distribution of type (3) we first write (9) as $P(s_1,\ldots,s_7,R_1,\ldots,R_7)$, where s_i is the sign of E_i, and we then calculate $P(s_1,\ldots,s_7|R_1,\ldots,R_7)$,

$$\begin{aligned}P(s_1,\ldots,s_7,R_1,\ldots,R_7) &= (1/L)\exp[(1/N^{1/2})(s_1s_2s_5R_1R_2R_5+\ldots\\&\quad+s_3s_4s_5R_3R_4R_5)\\&\quad-(1/N)(s_1s_2s_6s_7R_1R_2R_6R_7+\ldots\\&\quad+2s_1s_2s_3s_4R_1R_2R_3R_4)], \quad (11)\end{aligned}$$

where L is a scaling factor whose value does not depend on the sign values. We now calculate the marginal distribution

$$\begin{aligned}P(s_1,\ldots,s_6|R_1,\ldots,R_7) &= \sum_{s_7=\pm1} P(s_1,\ldots,s_7|R_1,\ldots,R_7)\\&\simeq(1/L)\exp\{(1/N^{1/2})(E_1E_2E_5+E_1E_3E_6\\&\quad+E_2E_4E_6+E_3E_4E_5)\\&\quad-(1/2N)(E_1E_4E_5E_6+E_2E_3E_5E_6\\&\quad+2E_1E_2E_3E_4)\}\cosh(R_5Z'_5),\end{aligned}$$

where $Z'_5=(E_1E_2+E_3E_4)/N^{1/2}$.

It is clear that the lack of information on s_7 cancels from (11) the invariants in which E_7 is not involved as a basis term, generates the new term $\cosh(R_5Z'_5)$ but does not modify the coefficient of $E_1E_2E_3E_4$ in the exponential term. In conclusion, we can write

$$\begin{aligned}P(s_1,\ldots,s_n|R_1,\ldots,R_p) &\simeq(1/L)\exp\Big[(1/N^{1/2})\sum_{\text{triplets}}E_iE_jE_l\\&\quad+\sum_{\text{quartets}}(E_iE_jE_lE_m/N)(w-w_5-w_6-w_7)\Big]\\&\quad\times\prod_\nu\cosh(w'_\nu R_\nu Z'_\nu), \quad (12)\end{aligned}$$

where

$$Z'_5=(E_iE_j+E_lE_m)/N^{1/2}$$
$$Z'_6=(E_iE_l+E_jE_m)/N^{1/2},$$
$$Z'_7=(E_jE_l+E_iE_m)/N^{1/2}.$$

The product in (12) includes all the *a priori* known magnitudes with unknown phase that are cross terms of some quartet in (12). We have $w'_\nu=0$ except when R_ν is *a priori* known and φ_ν is unknown; in this case $w'_\nu=1$.

(c) Distributions in pure exponential form

In practical applications, distributions (8) and (12) may present some difficulties. In fact, when the numbers of triplets and quartets considered in the formulae is high the exponent of the exponential function may become a very large positive or negative number according to circumstances. Consequently, the multiplication of such an exponential function by the product of a large number of modified Bessel or hyperbolic cosine functions may be difficult to perform. A more useful technique is the approximation of (8) and (12) in pure exponential form. If x is sufficiently small then

$$I_0(x)\simeq1+x^2/4\simeq\exp(x^2/4). \quad (13)$$

Hence (8) can be written in a pure exponential form,

$$P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)\simeq(1/L)\exp\left(\sum_{\text{triplets}}T_{ijl}\cos t_{ijl}+\sum_{\text{quartets}}Q_{ijlm}\cos q_{ijlm}\right) \quad (14)$$

where

$$\begin{aligned}Q_{ijlm}&=B_{ijlm}[w+(w'_5R_5^2-w_5)\\&\quad+(w'_6R_6^2-w_6)+(w'_7R_7^2-w_7)]. \quad (15)\end{aligned}$$

Terms not dependent on the phase values are not emphasized; they merely affect the value of the scaling factor L.

Instead of using (14), a further approximation of (8) (valid for any practical values of the Rs) may be obtained by following a procedure suggested by Giacovazzo, Camalli & Spagna (1989). Accordingly, each I_0 function in (8) is replaced as in the following:

$$I_0(w'_5R_5Z_5)\simeq\exp[w'_5\alpha_5\cos(\varphi_i+\varphi_j+\varphi_l+\varphi_m)],$$
$$I_0(w'_6R_6Z_6)\simeq\exp[w'_6\alpha_6\cos(\varphi_i+\varphi_j+\varphi_l+\varphi_m)],$$
$$I_0(w'_7R_7Z_7)\simeq\exp[w'_7\alpha_7\cos(\varphi_i+\varphi_j+\varphi_l+\varphi_m)],$$

where α_5, α_6, α_7 satisfy the equations

$$D_1(\alpha_5)=D_1(2R_iR_jR_5/N^{1/2})D_1(2R_lR_mR_5/N^{1/2}),$$
$$D_1(\alpha_6)=D_1(2R_iR_lR_6/N^{1/2})D_1(2R_jR_mR_6/N^{1/2}),$$
$$D_1(\alpha_7)=D_1(2R_jR_lR_7/N^{1/2})D_1(2R_iR_mR_7/N^{1/2})$$

and $D_1=I_1/I_0$ is the ratio of the two modified Bessel functions of order 1 and 0, respectively. Again, terms not depending on the phase values are not emphasized; they affect the value of the scale factor L. Equation (8) may thus be written in a pure exponential form:

$$P(\varphi_1,\ldots,\varphi_n|R_1,\ldots,R_p)\simeq(1/L)\exp\left(\sum_{\text{triplets}}T_{ijl}\cos t_{ijl}+\sum_{\text{quartets}}Q'_{ijlm}\cos q_{ijlm}\right), \quad (16)$$

where

$$Q'_{ijlm} = wB_{ijlm} + (w'_5\alpha_5 - B_{ijlm}w_5) + (w'_6\alpha_6 - B_{ijlm}w_6) + (w'_7\alpha_7 - B_{ijlm}w_7). \quad (17)$$

(d) Extension of the Giacovazzo formulation

According to Giacovazzo (1976),

$$\begin{aligned}
P(\varphi_1, \ldots, \varphi_7 | R_1, \ldots, R_7) \\
\simeq (1/L)\{1 + (2/N^{1/2})[R_1R_2R_5 \cos(\varphi_1+\varphi_2-\varphi_5) \\
+ R_1R_3R_6 \cos(\varphi_1+\varphi_3-\varphi_6) \\
+ R_1R_4R_7 \cos(\varphi_1-\varphi_4+\varphi_7) \\
+ R_2R_3R_7 \cos(\varphi_2+\varphi_3-\varphi_7) \\
+ R_2R_4R_6 \cos(\varphi_2-\varphi_4+\varphi_6) \\
+ R_3R_4R_5 \cos(\varphi_3-\varphi_4+\varphi_5)] \\
+ (1/N)[R_1^2R_2^2R_5^2 \cos 2(\varphi_1+\varphi_2-\varphi_5) \\
+ R_1^2R_3^2R_6^2 \cos 2(\varphi_1+\varphi_3-\varphi_6) \\
+ R_1^2R_4^2R_7^2 \cos 2(\varphi_1-\varphi_4+\varphi_7) \\
+ R_2^2R_3^2R_7^2 \cos 2(\varphi_2+\varphi_3-\varphi_7) \\
+ R_2^2R_4^2R_6^2 \cos 2(\varphi_2-\varphi_4+\varphi_6) \\
+ R_3^2R_4^2R_5^2 \cos 2(\varphi_3-\varphi_4+\varphi_5)] \\
+ (2/N)[R_1R_2R_6R_7(1+\varepsilon_3+\varepsilon_4) \\
\times \cos(\varphi_1-\varphi_2-\varphi_6+\varphi_7) \\
+ R_1R_3R_5R_7(1+\varepsilon_2+\varepsilon_4) \cos(\varphi_1-\varphi_3-\varphi_5+\varphi_7) \\
+ R_1R_4R_5R_6(1+\varepsilon_2+\varepsilon_3) \cos(\varphi_1+\varphi_4-\varphi_5-\varphi_6) \\
+ R_2R_3R_5R_6(1+\varepsilon_1+\varepsilon_4) \cos(\varphi_2-\varphi_3-\varphi_5+\varphi_6) \\
+ R_2R_4R_5R_7(1+\varepsilon_1+\varepsilon_3) \cos(\varphi_2+\varphi_4-\varphi_5-\varphi_7) \\
+ R_3R_4R_6R_7(1+\varepsilon_1+\varepsilon_2) \cos(\varphi_3+\varphi_4-\varphi_6-\varphi_7) \\
+ R_1R_2R_3R_4(1+\varepsilon_5+\varepsilon_6+\varepsilon_7) \\
\times \cos(\varphi_1+\varphi_2+\varphi_3-\varphi_4)]\}, \quad (18)
\end{aligned}$$

where $\varepsilon_i = R_i^2 - 1$. The contribution of a generic quartet to (18) is therefore

$$B_{ijlm}(w + w_5\varepsilon_5 + w_6\varepsilon_6 + w_7\varepsilon_7),$$

where $w = 1$ and w_5, w_6, w_7 are 0 or 1 according to whether the cross magnitudes R_5, R_6, R_7 are known or unknown, respectively. Simple calculations show in addition that the lack of information on φ_7 does not modify the coefficients of the various quartet contributions but only cancels from (18) the triplet and quartet terms in which φ_7 is involved as a basis term. In conclusion, a general expression for a probability distribution of type (2) is

$$\begin{aligned}
P(\varphi_1, \ldots, \varphi_n | R_1, \ldots, R_p) \\
\simeq (1/L)\Big\{1 + \sum_{\text{triplets}} [T_{ijl} \cos t_{ijl} \\
+ (R_i^2R_j^2R_l^2/N) \cos 2t_{ijl}] \\
+ \sum_{\text{quartets}} Q''_{ijlm} \cos q_{ijlm}\Big\}, \quad (19)
\end{aligned}$$

where

$$Q''_{ijlm} = B_{ijlm}(w + w_5\varepsilon_5 + w_6\varepsilon_6 + w_7\varepsilon_7). \quad (20)$$

Equation (19) may be approximated by the exponential form

$$P(\varphi_1, \ldots, \varphi_n | R_1, \ldots, R_p) \simeq (1/L) \exp\Big(\sum_{\text{triplets}} T_{ijl} \cos t_{ijl} + \sum_{\text{quartets}} Q''_{ijlm} \cos q_{ijlm}\Big). \quad (21)$$

The corresponding distribution in a centrosymmetric space group is (see Giacovazzo, 1975)

$$\begin{aligned}
P(s_1, \ldots, s_n | R_1, \ldots, R_p) \\
\simeq (1/L) \exp\Big[\sum_{\text{triplets}} (E_iE_jE_l/N^{1/2}) \\
+ \sum_{\text{quartets}} (E_iE_jE_lE_m/N) \\
\times (w + w_5\varepsilon_5 + w_6\varepsilon_6 + w_7\varepsilon_7)\Big]. \quad (22)
\end{aligned}$$

It must be emphasized that, according to Cochran (1955),

$$D_1(T_{ijl}) = \langle\cos(t_{ijl})\rangle \quad (23)$$

and, according to Giacovazzo (1976),

$$D_1(Q''_{ijlm}) = \langle\cos(q_{ijlm})\rangle, \quad (24)$$

where $D_1(x) = I_1(x)/I_0(x)$ is the ratio of the two modified Bessel functions of order 1 and 0, respectively. This observation provides a statistical meaning for the coefficients in (21). Practical use of Q'' revealed that for small structures Q'' can overestimate the cosine values of the positive estimated quartets. A more carefully formulated probabilistic theory (Giacovazzo, 1980*b*) suggests that Q''_{ijlm} as defined by (20) should be replaced by

$$Q''_{ijlm} = B_{ijlm}(w + w_5\varepsilon_5 + w_6\varepsilon_6 + w_7\varepsilon_7)/(1 + Z_{ijlm}), \quad (25)$$

where

$$Z_{ijlm} = [(\varepsilon_1\varepsilon_2 + \varepsilon_3\varepsilon_4)w_5\varepsilon_5 + (\varepsilon_1\varepsilon_3 + \varepsilon_2\varepsilon_4)w_6\varepsilon_6 + (\varepsilon_1\varepsilon_4 + \varepsilon_2\varepsilon_3)w_7\varepsilon_7]/2N.$$

From now on we will always assume that Q''_{ijlm} is defined by (25).

3. Concluding remarks

The probabilistic approaches for the estimation of quartet invariants in $P1$ and $P\bar{1}$ by Hauptman and by Giacovazzo have been further developed to derive the joint probability distribution function (including terms up to order N^{-1}) of n phases given $p \geq n$ moduli. The formulae (14) and (21) are obtained.

In (14) the weights w and w' suggest that a cross term of a quartet with large known modulus but unknown phase provides a positive contribution to Q [see (15)], while a negative contribution is provided by a cross term when both its modulus and its phase are known. According to (21), a cross term with known large modulus will provide a positive contribution to Q'' no matter whether the corresponding phase is known or not. Such behaviour should have striking consequences in practical applications. Indeed, as long as the largest $|E|$ values are phased during the phasing process, (14) and (21) will use such information in different ways. In particular, the positivity of the quartet term is expected to decrease in (14) and increase in (21).

In conclusion, while the approaches of Hauptman and Giacovazzo produce nearly equivalent quartet estimates (Giacovazzo, Camalli & Spagna, 1989) when only moduli are *a priori* known, the two formalisms lead to quite different estimates when applied to a situation in which a large number of phases are also known. This unexpected result will prove of large practical interest, as shown in the paper by Burla, Cascarano & Giacovazzo (1992), and suggests that the probabilistic quartet theory, as formulated so far, is not completely satisfactory. Indeed, the different mathematical approximations involved in the approaches of Hauptman and Giacovazzo are far from being insignificant if they cause such striking differences.

References

Burla, M. C., Cascarano, G. & Giacovazzo, C. (1992). *Acta Cryst.* A**48**, 906-912.
Cochran, W. (1955). *Acta Cryst.* **8**, 473-478.
Giacovazzo, C. (1975). *Acta Cryst.* A**31**, 252-259.
Giacovazzo, C. (1976). *Acta Cryst.* A**32**, 91-99.
Giacovazzo, C. (1977*a*). *Acta Cryst.* A**33**, 933-944.
Giacovazzo, C. (1977*b*). *Acta Cryst.* A**33**, 50-54.
Giacovazzo, C. (1980*a*). *Acta Cryst.* A**36**, 362-372.
Giacovazzo, C. (1980*b*). *Direct Methods in Crystallography.* London: Academic Press.
Giacovazzo, C., Camalli, M. & Spagna, R. (1989). *Acta Cryst.* A**45**, 141-143.
Green, E. A. & Hauptman, H. (1976). *Acta Cryst.* A**32**, 43-45.
Hauptman, H. (1975). *Acta Cryst.* A**31**, 680-687.
Hauptman, H. (1978). *Acta Cryst.* A**34**, 525-528.
Hauptman, H. & Karle, J. (1953). *The Solution of the Phase Problem* I. *The Centrosymmetric Crystal*, ACA Monograph No. 3. New York: Polycrystal Book Service.
Heinerman, J. J. L. (1977). Thesis, Univ. of Utrecht, The Netherlands.
Klug, A. (1958). *Acta Cryst.* **11**, 515-543.

Acta Cryst. (1992). A**48**, 906-912

The Joint Probability Distribution of Any Set of Phases Given Any Set of Diffraction Magnitudes. II. Practical Applications

By M. C. Burla

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

and G. Cascarano and C. Giacovazzo

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche CNR, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 Bari, Italy

(*Received* 13 *December* 1992; *accepted* 6 *May* 1992)

Abstract

In the first paper in this series [Giacovazzo, Burla & Cascarano (1992). *Acta Cryst.* A**48**, 901-906], the conditional joint probability distribution function of n phases given $p \geq n$ moduli was derived. The properties of the concluding formulae are checked here. It is found that the distribution is not maximized by the correct phases, mostly because of bias in the formulae. If the triplets are estimated *via* the $P10$ formula [Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori (1984). *Acta Cryst.* A**40**, 278-283] instead of being estimated by the Cochran relationship [Cochran (1952). *Acta Cryst.* **5**, 65-67], the situation is remarkably improved but further improvements are needed. A practical procedure is also described that successfully uses phase relationships to solve difficult structures.

0108-7673/92/060906-07$06.00