Acta Cryst. (1977). A**33**, 873–878

The von Mises Distribution of the Phase of a Structure Invariant

BY J. J. L. HEINERMAN, H. KRABBENDAM AND J. KROON

Laboratorium voor Structuurchemie, Rijksuniversiteit, Padualaan 8, *Utrecht, The Netherlands*

(*Received* 1 *February* 1977; *accepted* 31 *March* 1977)

The main terms in the exponential forms of the phase probability distributions of structure invariants are von Mises distributions $P(\Phi)=[2\pi I_0(k)]^{-1}\exp[k\cos(\Phi-\varphi)]$ (*a*). It is shown that in this formula k and φ are given by $k\exp i\varphi=2\langle F\rangle|F|/\sigma^2$ (*b*), where F, a complex variable, is identified with a structure invariant, $\langle F\rangle$ is its mean value and σ^2 is its variance. From (*a*) and (*b*) are obtained, by simple algebraic calculations of $\langle F\rangle$ and σ^2, the formulae for phase probability distributions of structure invariants, the derivation of which up to now required lengthy calculations *via* joint probability distributions of many structure factors. A new application is the calculation of the phase probability distribution of a quartet employing both the magnitudes of the cross terms and *a priori* structural information.

Introduction

Analogous to Gauss' derivation of the normal distribution von Mises (1918) has derived the normal distribution of a phase,

$$P(\Phi)=\frac{1}{2\pi I_0(k)}\exp[k\cos(\Phi-\varphi)], \qquad (1)$$

which in the literature is referred to as the circular normal distribution or the von Mises distribution (Batschelet, 1965; Mardia, 1972). The mode of (1) is at $\Phi=\varphi$, k is a measure of the sharpness, and I_0 is the modified Bessel function of order zero.

The distribution of the phase of a triple product $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=\mathbf{0}$, given the magnitudes of the normalized structure factors $E_{\mathbf{h}_1}$, $E_{\mathbf{h}_2}$ and $E_{\mathbf{h}_3}$ (Cochran, 1955; Hauptman, 1976; Heinerman, 1977*a*) is a von Mises distribution (see Hendrickson & Lattman, 1970 and Koenig, 1972, 1976) with

$$k\exp i\varphi=2\frac{\sigma_3}{\sigma_2^{3/2}}|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}|. \qquad (2)$$

If *a priori* structural information is used, (2) becomes

$$k\exp i\varphi=2Q_{123}(\exp iq_{123})|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}| \qquad (3)$$

(Main, 1976; Heinerman, 1977*a*).

The distribution of the phase of a quartet $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3+\mathbf{h}_4=\mathbf{0}$, given the magnitudes of $E_{\mathbf{h}_1}$, $E_{\mathbf{h}_2}$, $E_{\mathbf{h}_3}$ and $E_{\mathbf{h}_4}$ (Hauptman, 1976) is a von Mises distribution with

$$k\exp i\varphi=2\frac{\sigma_4}{\sigma_2^2}|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}|. \qquad (4)$$

If, in addition, the magnitudes of the cross terms $E_{\mathbf{h}_1+\mathbf{h}_2}$, $E_{\mathbf{h}_2+\mathbf{h}_3}$ and $E_{\mathbf{h}_3+\mathbf{h}_1}$ are used, the exponential form of the phase distribution of a quartet is again a von Mises distribution (see Giacovazzo, 1976), where now

$$k\exp i\varphi=2\left[\frac{\sigma_4}{\sigma_2^2}+\frac{\sigma_3^2}{\sigma_2^3}(|E_{\mathbf{h}_1+\mathbf{h}_2}|^2+|E_{\mathbf{h}_2+\mathbf{h}_3}|^2+|E_{\mathbf{h}_3+\mathbf{h}_1}|^2-3)\right]|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}| \qquad (5)$$

[from (4.5) of Hauptman (1976) and the method described by Heinerman (1977*b*)].

Expressions (2)–(5) have been derived by lengthy calculations *via* joint probability distributions of from three to seven structure factors. We shall show that there is a general expression for $k\exp i\varphi$ which directly leads to (2)–(5). If the whole structure is known this general expression leads to a delta function for $P(\Phi)$, as it should. This was not the case for (3). As new applications we shall calculate formulae analogous to (4) and (5), now with *a priori* structural information included.

The von Mises distribution

Let us consider a complex variable $F=A+iB$ where A and B are functions of primitive random variables. We assume that both the distribution of the real part A and the distribution of the imaginary part B are Gaussian. In addition we assume that A and B are independent and that their variances are equal. Then it can easily be verified that the joint probability distribution $P(X,Y)$ of A and B is

$$P(X,Y)=\frac{1}{\pi\sigma^2}\exp\left[-\frac{1}{\sigma^2}(X^2+Y^2-2\langle A\rangle X-2\langle B\rangle Y+\langle A\rangle^2+\langle B\rangle^2)\right], \qquad (6)$$

where

$$\sigma^2=\langle|F|^2\rangle-|\langle F\rangle|^2. \qquad (7)$$

Employing the transformation

$$X=R\cos\Phi, \quad Y=R\sin\Phi, \qquad (8)$$

and writing

$$\langle F\rangle = Q\exp iq\,, \tag{9}$$

we obtain

$$P(R,\Phi)=\frac{R}{\pi\sigma^2}\exp\left\{-\frac{1}{\sigma^2}[R^2-2QR\cos(\Phi-q)+Q^2]\right\}. \tag{10}$$

From this distribution it follows that the conditional probability distribution of the phase of F given the magnitude of F is

$$P(\Phi|R)=\frac{1}{L}\exp\left[2\frac{QR}{\sigma^2}\cos(\Phi-q)\right], \tag{11}$$

where L, the normalizing constant, is

$$L=2\pi I_0\left(2\frac{QR}{\sigma^2}\right). \tag{12}$$

Comparing (11) with the von Mises distribution (1), using (9) and replacing R by $|F|$ we obtain

$$k\exp i\varphi=2\frac{\langle F\rangle|F|}{\sigma^2}. \tag{13}$$

We note that $\langle F\rangle$ and σ^2 have to be calculated without fixing the magnitude of F.

Formula (13) gives $k\exp i\varphi$ as a function of the mean value, the variance and the magnitude of F. Identifying F with a structure invariant (in general a structure factor product; see the last section), the magnitude of which is obtained from experiment, it remains to calculate $\langle F\rangle$ and σ^2.

For a triple product a discussion on the choice of the primitive random variables has been given by Heinerman (1977*a*). It was shown that regarding $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$ as the primitive random variables leads to the same result as regarding all the atomic position vectors as the primitive random variables, although the conditions are different. It was also shown that in the approach where the atomic position vectors are regarded as the primitive random variables *a priori* structural information can be used. In the following sections this approach will be the basis of our calculations.

The phase probability of a triple product

We shall first consider the case that there is no *a priori* structural information, except for the contents of the unit cell. The structure factor $F_{\mathbf{h}}$ is defined by

$$F_{\mathbf{h}}=\sum_{j=1}^{N} f_j(\mathbf{h})\exp(2\pi i\mathbf{h}\,.\,\mathbf{r}_j)\,, \tag{14}$$

where N is the number of atoms in the unit cell, $f_j(\mathbf{h})$ is the scattering factor of atom j and $\mathbf{r}_j$ is the position vector of atom j. For simplicity we assume that the space group is $P1$. Next, we identify F in (13) with a triple product $F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=\mathbf{0}$,

$$F\equiv F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}=\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\sum_{j_3=1}^{N} f_{j_1}(\mathbf{h}_1)f_{j_2}(\mathbf{h}_2)f_{j_3}(\mathbf{h}_3)$$
$$\times\exp\{2\pi i[\mathbf{h}_1\,.\,(\mathbf{r}_{j_1}-\mathbf{r}_{j_3})+\mathbf{h}_2\,.\,(\mathbf{r}_{j_2}-\mathbf{r}_{j_3})]\}\,. \tag{15}$$

Simple algebraic calculations (see Appendix I) give

$$\langle F\rangle=\sum_{j=1}^{N} f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3) \tag{16}$$

and

$$\langle|F|^2\rangle=[\sum_{j=1}^{N} f_j(\mathbf{h}_1)^2][\sum_{j=1}^{N} f_j(\mathbf{h}_2)^2][\sum_{j=1}^{N} f_j(\mathbf{h}_3)^2]$$
$$+[\sum_{j=1}^{N} f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3)]^2$$
$$-\sum_{j=1}^{N} f_j(\mathbf{h}_1)^2f_j(\mathbf{h}_2)^2f_j(\mathbf{h}_3)^2\,, \tag{17}$$

which lead to the variance

$$\sigma^2=[\sum_{j=1}^{N} f_j(\mathbf{h}_1)^2][\sum_{j=1}^{N} f_j(\mathbf{h}_2)^2][\sum_{j=1}^{N} f_j(\mathbf{h}_3)^2]$$
$$-\sum_{j=1}^{N} f_j(\mathbf{h}_1)^2f_j(\mathbf{h}_2)^2f_j(\mathbf{h}_3)^2\,. \tag{18}$$

From (13), (16) and (18) we obtain the result

$$k\exp i\varphi=2\frac{[\sum_{j=1}^{N} f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3)]|F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}|}{[\sum_{j=1}^{N} f_j(\mathbf{h}_1)^2][\sum_{j=1}^{N} f_j(\mathbf{h}_2)^2][\sum_{j=1}^{N} f_j(\mathbf{h}_3)^2]-\sum_{j=1}^{N} f_j(\mathbf{h}_1)^2f_j(\mathbf{h}_2)^2f_j(\mathbf{h}_3)^2}. \tag{19}$$

If we use the definitions for $E_{\mathbf{h}}$ and σ_n,

$$E_{\mathbf{h}}=\frac{F_{\mathbf{h}}}{\langle|F_{\mathbf{h}}|^2\rangle^{1/2}}\,, \tag{20}$$

where

$$\langle|F_{\mathbf{h}}|^2\rangle=\sum_{j=1}^{N} f_j(\mathbf{h})^2\,, \tag{21}$$

and

$$\sigma_n=\sum_{j=1}^{N} Z_j^n\,, \tag{22}$$

where Z_j is the atomic number of atom j, formula (19) leads to (2) provided we neglect the last summation in the denominator of (19) [this is justified if the normalized scattering factors

$$f_j(\mathbf{h})/(\sum_{j=1}^{N} f_j(\mathbf{h})^2)^{1/2}$$

are small] and if we assume that all the atoms have the same unitary scattering factor $\hat{f}(\mathbf{h})$,

$$f_j(\mathbf{h}) = Z_j \hat{f}(\mathbf{h}) . \tag{23}$$

The conditions for (19) are $\mathbf{h}_i \neq \mathbf{0}$ $i=1,2,3$ and $\mathbf{h}_i \neq \mathbf{h}_j$ $i,j=1,2,3$ $i \neq j$ (see Appendix I). Heinerman (1977*a*) also arrived at the condition $2\mathbf{h}_i \neq \mathbf{h}_j$. The two cases $2\mathbf{h}_i \neq \mathbf{h}_j$ and $2\mathbf{h}_i = \mathbf{h}_j$ lead to different joint probability distributions for the phases of $E_{\mathbf{h}_1}$, $E_{\mathbf{h}_2}$ and $E_{\mathbf{h}_3}$. However, both distributions give the same probability distribution for the phase of $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}$ (see Appendix II; there we also give the probability distribution for the phase of $E^2_{\mathbf{h}_1}E_{\mathbf{h}_2}$, $2\mathbf{h}_1 + \mathbf{h}_2 = \mathbf{0}$).

We now consider the case of *a priori* structural information (see Heinerman, 1977*a*), and write $F_{\mathbf{h}}$ as

$$F_{\mathbf{h}} = \sum_{j=1}^{p} g_j(\mathbf{h}) \exp(2\pi i \mathbf{h} . \mathbf{r}_j) , \tag{24}$$

where for $1 \leq j \leq p_1$ the $\mathbf{r}_j$'s are the atomic position vectors and the $g_j(\mathbf{h})$ are the scattering factors $f_j(\mathbf{h})$, and for $p_1 + 1 \leq j \leq p$ the $\mathbf{r}_j$'s are the group position vectors and the $g_j(\mathbf{h})$ are the group scattering factors. The independent primitive random variables are (*a*) $1 \leq j \leq p_1$, the atomic position vectors, (*b*) $p_1 + 1 \leq j \leq p_2$, the position vectors and orientational parameters of the groups, (*c*) $p_2 + 1 \leq j \leq p$, the position vectors of the groups with known orientation. Then, identifying F with $F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}$, $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = \mathbf{0}$, as before, we obtain for F expression (15) with N replaced by p and the $f_j(\mathbf{h})$ replaced by $g_j(\mathbf{h})$. For $\langle F \rangle$ and $\langle |F|^2 \rangle$ we find

$$\langle F \rangle = \sum_{j=1}^{p} \langle g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3) \rangle \tag{25}$$

and

$$\langle |F|^2 \rangle = \sum_{j_1=1}^{p} \sum_{j_2=1}^{p} \sum_{j_3=1}^{p} \langle |g_{j_1}(\mathbf{h}_1) g_{j_2}(\mathbf{h}_2) g_{j_3}(\mathbf{h}_3)|^2 \rangle + \langle |\sum_{j=1}^{p} g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3)|^2 \rangle - \sum_{j=1}^{p} \langle |g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3)|^2 \rangle , \tag{26}$$

where in the right-hand sides of (25) and (26) the $g_j(\mathbf{h})$ with $p_1 + 1 \leq j \leq p_2$ depend on the orientational parameters with respect to which the averaging still has to be carried out. (Group) scattering factors and products of (group) scattering factors do not depend on the (group) position vectors. Therefore averaging these with respect to the primitive random variables means averaging with respect to the orientational parameters of the groups j, $p_1 + 1 \leq j \leq p_2$. From (25) and (26) it follows that

$$\sigma^2 = \sum_{j_1=1}^{p} \sum_{j_2=1}^{p} \sum_{j_3=1}^{p} \langle |g_{j_1}(\mathbf{h}_1) g_{j_2}(\mathbf{h}_2) g_{j_3}(\mathbf{h}_3)|^2 \rangle - \sum_{j=1}^{p} |\langle g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3) \rangle|^2 , \tag{27}$$

which with (13) and (25) leads to

$$k \exp i\varphi = 2 \frac{[\sum_{j=1}^{p} \langle g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3) \rangle] |F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3}|}{\sum_{j_1=1}^{p} \sum_{j_2=1}^{p} \sum_{j_3=1}^{p} \langle |g_{j_1}(\mathbf{h}_1) g_{j_2}(\mathbf{h}_2) g_{j_3}(\mathbf{h}_3)|^2 \rangle - \sum_{j=1}^{p} |\langle g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3) \rangle|^2} . \tag{28}$$

If there is no structural information (28) reduces to (19).

If there is only a small amount of structural information [*i.e.* the magnitudes of the normalized (group) scattering factors

$$g_j(\mathbf{h}) / (\sum_{j=1}^{p} \langle |g_j(\mathbf{h})|^2 \rangle)^{1/2}$$

are small] the variance (27) can be approximated by

$$\sigma^2 \simeq [\sum_{j=1}^{p} \langle |g_j(\mathbf{h}_1)|^2 \rangle] [\sum_{j=1}^{p} \langle |g_j(\mathbf{h}_2)|^2 \rangle] [\sum_{j=1}^{p} \langle |g_j(\mathbf{h}_3)|^2 \rangle] . \tag{29}$$

If the denominator of (28) is replaced by (29) and (20) employed, where now

$$\langle |F_{\mathbf{h}}|^2 \rangle = \sum_{j=1}^{p} \langle |g_j(\mathbf{h})|^2 \rangle , \tag{30}$$

(28) leads to (3).

For the case that the entire structure is known φ is equal to the true phase of the triple product and k is infinite, which means that then (28) leads to a delta function centred on the true phase of the triple product; this is in strong contrast to (3) in this limiting case (see Heinerman, 1977*a*).

The phase probability of a quartet

Analogous to the derivation of (28) we obtain for $F \equiv F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}F_{\mathbf{h}_4}$, $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = \mathbf{0}$,

$$k \exp i\varphi = 2 \frac{[\sum_{j=1}^{p} \langle g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3) g_j(\mathbf{h}_4) \rangle] |F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3} F_{\mathbf{h}_4}|}{\sum_{j_1=1}^{p} \cdots \sum_{j_4=1}^{p} \langle |g_{j_1}(\mathbf{h}_1) g_{j_2}(\mathbf{h}_2) g_{j_3}(\mathbf{h}_3) g_{j_4}(\mathbf{h}_4)|^2 \rangle - \sum_{j=1}^{p} |\langle g_j(\mathbf{h}_1) g_j(\mathbf{h}_2) g_j(\mathbf{h}_3) g_j(\mathbf{h}_4) \rangle|^2} . \tag{31}$$

If no structural information is available, $p=N$ and $g_j(\mathbf{h})=f_j(\mathbf{h})$, $j=1,\ldots,N$; then (4) is obtained by using (20)–(23) and neglecting the last sum in the denominator of (31).

where $K_{\mathbf{h}_1\mathbf{h}_2}$, a complex quantity, is obtained from a least-squares fitting procedure [see Appendix III; the normalized (group) scattering factors are assumed to be small],

$$K_{\mathbf{h}_1\mathbf{h}_2}=\frac{[\sum_{j=1}^{p}\langle g_j(\mathbf{h}_1)g_j(\mathbf{h}_2)g_j(\overline{\mathbf{h}_1+\mathbf{h}_2})\rangle][\sum_{j=1}^{p}\langle g_j(\mathbf{h}_3)g_j(\mathbf{h}_4)g_j(\overline{\mathbf{h}_3+\mathbf{h}_4})\rangle]}{[\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_1+\mathbf{h}_2)|^2\rangle]^2}. \tag{34}$$

The cross terms $F_{\mathbf{h}_1+\mathbf{h}_2}$, $F_{\mathbf{h}_2+\mathbf{h}_3}$ and $F_{\mathbf{h}_3+\mathbf{h}_1}$

The expression for a quartet can be written as

$$\begin{aligned}F=F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}F_{\mathbf{h}_4}&=\sum_{j=1}^{p}g_j(\mathbf{h}_1)g_j(\mathbf{h}_2)g_j(\mathbf{h}_3)g_j(\mathbf{h}_4)\\&+\sum_{\substack{j_1=1\\j_1\neq j_2}}^{p}\sum_{j_2=1}^{p}g_{j_1}(\mathbf{h}_1)g_{j_1}(\mathbf{h}_2)g_{j_2}(\mathbf{h}_3)g_{j_2}(\mathbf{h}_4)\\&\times\exp[2\pi i(\mathbf{h}_1+\mathbf{h}_2).(\mathbf{r}_{j_1}-\mathbf{r}_{j_2})]\\&+\sum_{\substack{j_1=1\\j_1\neq j_2}}^{p}\sum_{j_2=1}^{p}g_{j_2}(\mathbf{h}_1)g_{j_1}(\mathbf{h}_2)g_{j_1}(\mathbf{h}_3)g_{j_2}(\mathbf{h}_4)\\&\times\exp[2\pi i(\mathbf{h}_2+\mathbf{h}_3).(\mathbf{r}_{j_1}-\mathbf{r}_{j_2})]\\&+\sum_{\substack{j_1=1\\j_1\neq j_2}}^{p}\sum_{j_2=1}^{p}g_{j_1}(\mathbf{h}_1)g_{j_2}(\mathbf{h}_2)g_{j_1}(\mathbf{h}_3)g_{j_2}(\mathbf{h}_4)\\&\times\exp[2\pi i(\mathbf{h}_3+\mathbf{h}_1).(\mathbf{r}_{j_1}-\mathbf{r}_{j_2})]\\&+\sum_{\substack{j_1=1\\\text{not } j_1=j_2\wedge j_3=j_4\\\text{not } j_1=j_4\wedge j_2=j_3\\\text{not } j_1=j_3\wedge j_2=j_4}}^{p}\sum_{j_2=1}^{p}\sum_{j_3=1}^{p}\sum_{j_4=1}^{p}g_{j_1}(\mathbf{h}_1)g_{j_2}(\mathbf{h}_2)g_{j_3}(\mathbf{h}_3)g_{j_4}(\mathbf{h}_4)\\&\times\exp\{2\pi i[\mathbf{h}_1.(\mathbf{r}_{j_1}-\mathbf{r}_{j_4})+\mathbf{h}_2.(\mathbf{r}_{j_2}-\mathbf{r}_{j_4})\\&+\mathbf{h}_3.(\mathbf{r}_{j_3}-\mathbf{r}_{j_4})]\}.\end{aligned} \tag{32}$$

For equal atoms and no *a priori* structural information the double summations in (32) are directly related to $|F_{\mathbf{h}_1+\mathbf{h}_2}|^2$, $|F_{\mathbf{h}_2+\mathbf{h}_3}|^2$ and $|F_{\mathbf{h}_3+\mathbf{h}_1}|^2$. For unequal atoms and/or *a priori* structural information we therefore make the following estimation

$$\begin{aligned}&\sum_{\substack{j_1=1\\j_1\neq j_2}}^{p}\sum_{j_2=1}^{p}g_{j_1}(\mathbf{h}_1)g_{j_1}(\mathbf{h}_2)g_{j_2}(\mathbf{h}_3)g_{j_2}(\mathbf{h}_4)\\&\quad\times\exp[2\pi i(\mathbf{h}_1+\mathbf{h}_2).(\mathbf{r}_{j_1}-\mathbf{r}_{j_2})]\\&\quad\simeq K_{\mathbf{h}_1\mathbf{h}_2}\sum_{\substack{j_1=1\\j_1\neq j_2}}^{p}\sum_{j_2=1}^{p}g_{j_1}(\mathbf{h}_1+\mathbf{h}_2)g_{j_2}(\mathbf{h}_3+\mathbf{h}_4)\\&\quad\times\exp[2\pi i(\mathbf{h}_1+\mathbf{h}_2).(\mathbf{r}_{j_1}-\mathbf{r}_{j_2})]\\&\quad=K_{\mathbf{h}_1\mathbf{h}_2}(|F_{\mathbf{h}_1+\mathbf{h}_2}|^2-\sum_{j=1}^{p}|g_j(\mathbf{h}_1+\mathbf{h}_2)|^2),\end{aligned} \tag{33}$$

The results for the two remaining double summations are obtained by permutation of $\mathbf{h}_1$, $\mathbf{h}_2$ and $\mathbf{h}_3$. Then

$$\begin{aligned}\langle F\rangle\simeq&\sum_{j=1}^{p}\langle g_j(\mathbf{h}_1)g_j(\mathbf{h}_2)g_j(\mathbf{h}_3)g_j(\mathbf{h}_4)\rangle\\&+K_{\mathbf{h}_1\mathbf{h}_2}[|F_{\mathbf{h}_1+\mathbf{h}_2}|^2-\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_1+\mathbf{h}_2)|^2\rangle]\\&+K_{\mathbf{h}_2\mathbf{h}_3}[|F_{\mathbf{h}_2+\mathbf{h}_3}|^2-\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_2+\mathbf{h}_3)|^2\rangle]\\&+K_{\mathbf{h}_3\mathbf{h}_1}[|F_{\mathbf{h}_3+\mathbf{h}_1}|^2-\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_3+\mathbf{h}_1)|^2\rangle].\end{aligned} \tag{35}$$

With the same assumption which led to (34) we obtain for the variance

$$\begin{aligned}\sigma^2\simeq&[\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_1)|^2\rangle][\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_2)|^2\rangle]\\&\times[\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_3)|^2\rangle][\sum_{j=1}^{p}\langle|g_j(\mathbf{h}_4)|^2\rangle].\end{aligned} \tag{36}$$

The final result expressed in normalized (group) scattering factors,

$$t_j(\mathbf{h})=\frac{g_j(\mathbf{h})}{\langle|F_{\mathbf{h}}|^2\rangle^{1/2}}, \tag{37}$$

and in normalized structure factors (20), where $\langle|F_{\mathbf{h}}|^2\rangle$ is given by (30), is

$$\begin{aligned}k\exp i\varphi\simeq 2\{&\sum_{j=1}^{p}\langle t_j(\mathbf{h}_1)t_j(\mathbf{h}_2)t_j(\mathbf{h}_3)t_j(\mathbf{h}_4)\rangle\\&+[\sum_{j=1}^{p}\langle t_j(\mathbf{h}_1)t_j(\mathbf{h}_2)t_j(\overline{\mathbf{h}_1+\mathbf{h}_2})\rangle]\\&\times[\sum_{j=1}^{p}\langle t_j(\mathbf{h}_3)t_j(\mathbf{h}_4)t_j(\overline{\mathbf{h}_3+\mathbf{h}_4})\rangle](|E_{\mathbf{h}_1+\mathbf{h}_2}|^2-1)\\&+[\sum_{j=1}^{p}\langle t_j(\mathbf{h}_2)t_j(\mathbf{h}_3)t_j(\overline{\mathbf{h}_2+\mathbf{h}_3})\rangle]\\&\times[\sum_{j=1}^{p}\langle t_j(\mathbf{h}_1)t_j(\mathbf{h}_4)t_j(\overline{\mathbf{h}_1+\mathbf{h}_4})\rangle](|E_{\mathbf{h}_2+\mathbf{h}_3}|^2-1)\\&+[\sum_{j=1}^{p}\langle t_j(\mathbf{h}_3)t_j(\mathbf{h}_1)t_j(\overline{\mathbf{h}_3+\mathbf{h}_1})\rangle]\\&\times[\sum_{j=1}^{p}\langle t_j(\mathbf{h}_2)t_j(\mathbf{h}_4)t_j(\overline{\mathbf{h}_2+\mathbf{h}_4})\rangle](|E_{\mathbf{h}_3+\mathbf{h}_1}|^2-1)\}\\&\times|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_3}E_{\mathbf{h}_4}|.\end{aligned} \tag{38}$$

If there is no *a priori* structural information, then with the use of (22) and (23) formula (5) is obtained. It can easily be seen that if, for example, $|F_{\mathbf{h}_1+\mathbf{h}_2}|$ is unknown, the term in (38) with $|E_{\mathbf{h}_1+\mathbf{h}_2}|^2-1$ vanishes, or in other words $|E_{\mathbf{h}_1+\mathbf{h}_2}|^2$ is then replaced by its mean value, which is equal to one. Analogous remarks apply to $|F_{\mathbf{h}_2+\mathbf{h}_3}|$ and $|F_{\mathbf{h}_3+\mathbf{h}_1}|$ (see also Heinerman, 1977*b*).

The fourfold summation in (32) contains double summations that are related to the squares of $|F_{\mathbf{h}_1}|$, $|F_{\mathbf{h}_2}|$, $|F_{\mathbf{h}_3}|$ and $|F_{\mathbf{h}_4}|$. Fixing these magnitudes in the calculations of the mean and variance of $F\equiv F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}F_{\mathbf{h}_4}$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3+\mathbf{h}_4=\mathbf{0}$, implies that $|F|$ is fixed, which is not allowed in the basic formula (13). The same holds for the magnitudes $|F_{\mathbf{h}_1}|$, $|F_{\mathbf{h}_2}|$ and $|F_{\mathbf{h}_3}|$ in the calculations of the mean and variance of $F\equiv F_{\mathbf{h}_1}F_{\mathbf{h}_2}F_{\mathbf{h}_3}$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=\mathbf{0}$ (see also Kroon & Krabbendam, 1970).

The calculations for higher-order structure invariants can be performed in the same way as described for a quartet. We calculated the von Mises distribution for a quintet employing the magnitudes of the ten cross terms and assuming point atoms. This distribution appears to be identical to the exponential form of formula (3.3) of Hauptman & Fortier (1977).

Discussion

The orientational averages $\langle g_j(\mathbf{h}_1)g_j(\mathbf{h}_2)g_j(\mathbf{h}_3)\rangle$, $p_1+1\le j\le p_2$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=\mathbf{0}$, are calculated with the $B(z,t)$ formula (Hauptman, 1965) which gives an expression for the average of $\exp\{2\pi i[\mathbf{h}_1.(\mathbf{r}_1-\mathbf{r}_3)+\mathbf{h}_2.(\mathbf{r}_2-\mathbf{r}_3)]\}$ over all orientations of the triangle formed by the atoms 1, 2 and 3. A formula for the orientational average of $\exp\{2\pi i[\mathbf{h}_1.(\mathbf{r}_1-\mathbf{r}_4)+\mathbf{h}_2.(\mathbf{r}_2-\mathbf{r}_4)+\mathbf{h}_3.(\mathbf{r}_3-\mathbf{r}_4)]\}$, needed for the calculation of $\langle g_j(\mathbf{h}_1)g_j(\mathbf{h}_2)g_j(\mathbf{h}_3)g_j(\mathbf{h}_4)\rangle$, $p_1+1\le j\le p_2$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3+\mathbf{h}_4=\mathbf{0}$, is not yet available. We suggest a numerical (computer) calculation. This may have the advantage that partial information about the orientational parameters, if available, can be used in the averaging process.

So far $\langle F\rangle$ and σ^2 have been calculated for structures in space group $P1$. It should be stressed that formula (13) is more generally valid: both $\langle F\rangle$ and σ^2 can be computed for any space group. Structure (sem)invariants of which the phases are restricted to 0 and π (or α and $\alpha+\pi$), however, are excluded because of the requirement that the real and imaginary parts of F have to be independent. For such structure (sem)invariants the phase probability can be obtained from a one-dimensional Gauss distribution. Then for centrosymmetric structures

$$P_+=0{\cdot}5+0{\cdot}5\tanh\frac{|F|\langle F\rangle}{\sigma^2} \tag{39}$$

[see, for example, Klug (1958), Kitaigorodskii (1961), Giacovazzo (1975) and for the use of *a priori* structural information Kroon & Krabbendam (1970)].

In general F can be identified with any structure factor product or with a single structure factor. For the latter, including *a priori* structural information, (13) leads to the formula of Sim (1959), while (39) gives Woolfson's (1956) formula.

We are indebted to Dr H. Hauptman for supplying preprints of his work.

APPENDIX I

The calculation of (16) and (17)

Formula (15) can be written as

$$\begin{aligned}F=&\sum_{j=1}^{N}f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3)\\&+\sum_{\substack{j_1=1\\ \text{not } j_1=j_2=j_3}}^{N}\sum_{j_2=1}^{N}\sum_{j_3=1}^{N}f_{j_1}(\mathbf{h}_1)f_{j_2}(\mathbf{h}_2)f_{j_3}(\mathbf{h}_3)\\&\times\exp\{2\pi i[\mathbf{h}_1.(\mathbf{r}_{j_1}-\mathbf{r}_{j_3})+\mathbf{h}_2.(\mathbf{r}_{j_2}-\mathbf{r}_{j_3})]\}.\end{aligned} \tag{I.1}$$

Averaging with respect to the atomic position vectors gives zero for all terms in the last summation if $\mathbf{h}_i\neq\mathbf{0}$, $i=1,2,3$,

$$\langle F\rangle=\sum_{j=1}^{N}f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3). \tag{I.2}$$

The expression for $|F|^2$ is

$$\begin{aligned}|F|^2=&\sum_{j_1=1}^{N}\cdots\sum_{i_3=1}^{N}f_{j_1}(\mathbf{h}_1)f_{j_2}(\mathbf{h}_2)f_{j_3}(\mathbf{h}_3)f_{i_1}(\mathbf{h}_1)f_{i_2}(\mathbf{h}_2)f_{i_3}(\mathbf{h}_3)\\&\times\exp(2\pi i\{\mathbf{h}_1.[(\mathbf{r}_{j_1}-\mathbf{r}_{i_1})-(\mathbf{r}_{j_3}-\mathbf{r}_{i_3})]\\&+\mathbf{h}_2.[(\mathbf{r}_{j_2}-\mathbf{r}_{i_2})-(\mathbf{r}_{j_3}-\mathbf{r}_{i_3})]\})\\=&\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\sum_{j_3=1}^{N}f_{j_1}(\mathbf{h}_1)^2f_{j_2}(\mathbf{h}_2)^2f_{j_3}(\mathbf{h}_3)^2\\&+\sum_{j=1}^{N}\sum_{\substack{i=1\\ j\neq i}}^{N}f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3)f_i(\mathbf{h}_1)f_i(\mathbf{h}_2)f_i(\mathbf{h}_3)\\&+\sum_{\substack{j_1=1\\ \text{not } j_1=i_1\wedge j_2=i_2\wedge j_3=i_3\\ \text{not } j_1=j_2=j_3\wedge i_1=i_2=i_3}}^{N}\cdots\sum_{i_3=1}^{N}f_{j_1}(\mathbf{h}_1)\ldots f_{i_3}(\mathbf{h}_3)\\&\times\exp(2\pi i\{\mathbf{h}_1.[(\mathbf{r}_{j_1}-\mathbf{r}_{i_1})-(\mathbf{r}_{j_3}-\mathbf{r}_{i_3})]\\&+\mathbf{h}_2.[(\mathbf{r}_{j_2}-\mathbf{r}_{i_2})-(\mathbf{r}_{j_3}-\mathbf{r}_{i_3})]\}).\end{aligned} \tag{I.3}$$

Again, averaging with respect to the atomic position vectors gives zero for all terms in the last summation if $\mathbf{h}_i\neq\mathbf{0}$, $i=1,2,3$, and $\mathbf{h}_i\neq\mathbf{h}_j$, $i,j=1,2,3$ $i\neq j$,

$$\begin{aligned}\langle|F|^2\rangle=&\Big(\sum_{j=1}^{N}f_j(\mathbf{h}_1)^2\Big)\Big(\sum_{j=1}^{N}f_j(\mathbf{h}_2)^2\Big)\Big(\sum_{j=1}^{N}f_j(\mathbf{h}_3)^2\Big)\\&+\Big(\sum_{j=1}^{N}f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3)\Big)^2-\sum_{j=1}^{N}f_j(\mathbf{h}_1)^2f_j(\mathbf{h}_2)^2f_j(\mathbf{h}_3)^2.\end{aligned} \tag{I.4}$$

APPENDIX II

The condition $2\mathbf{h}_i \neq \mathbf{h}_j$

From (21) of Heinerman (1977*a*), where $2\mathbf{h}_i \neq \mathbf{h}_j$, it follows that the joint probability distribution $P(\Phi_1,\Phi_2,\Phi_3|R_1,R_2,R_3)$ of the phases of $E_{\mathbf{h}_1}$, $E_{\mathbf{h}_2}$ and $E_{\mathbf{h}_3}$, given $|E_{\mathbf{h}_1}|$, $|E_{\mathbf{h}_2}|$ and $|E_{\mathbf{h}_3}|$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=\mathbf{0}$, is equal to

$$P(\Phi_1,\Phi_2,\Phi_3|R_1,R_2,R_3) \simeq C \times \exp\left[2\frac{\sigma_3}{\sigma_2^{3/2}} R_1R_2R_3\cos(\Phi_1+\Phi_2+\Phi_3)\right]. \quad \text{(II.1)}$$

For $2\mathbf{h}_1=\mathbf{h}_2$ it can be shown that the result is

$$P(\Phi_1,\Phi_2,\Phi_3|R_1,R_2,R_3) \simeq C \times \exp\left[2\frac{\sigma_3}{\sigma_2^{3/2}} R_1R_2R_3\cos(\Phi_1+\Phi_2+\Phi_3) + \frac{\sigma_3}{\sigma_2^{3/2}} R_1^2R_2\cos(2\Phi_1-\Phi_2)\right]. \quad \text{(II.2)}$$

Integrating (II.1) and (II.2) with respect to Φ_1, Φ_2 and Φ_3, such that $\Phi_1+\Phi_2+\Phi_3=\Phi$, in both cases leads to

$$P(\Phi|R_1,R_2,R_3) \simeq C' \exp\left(2\frac{\sigma_3}{\sigma_2^{3/2}} R_1R_2R_3\cos\Phi\right). \quad \text{(II.3)}$$

It follows that the condition $2\mathbf{h}_i \neq \mathbf{h}_j$ is relaxed in the process of integration with respect to the phases.

From (II.2) we can calculate the probability distribution of the phase of $E_{\mathbf{h}_1}^2 E_{\mathbf{h}_2}$, $2\mathbf{h}_1+\mathbf{h}_2=\mathbf{0}$, by replacing $\mathbf{h}_2$ and Φ_2 by $-\mathbf{h}_2$ and $-\Phi_2$ respectively and next integrating with respect to Φ_1, Φ_2 and Φ_3, such that $2\Phi_1+\Phi_2=\Phi$. This gives

$$P(\Phi|R_1,R_2) \simeq C'' \exp\left(\frac{\sigma_3}{\sigma_2^{3/2}} R_1^2R_2\cos\Phi\right). \quad \text{(II.4)}$$

The same result is obtained from (13) by calculating $\langle F\rangle$ and σ^2 for $F \equiv F_{\mathbf{h}_1}^2 F_{\mathbf{h}_2}$, $2\mathbf{h}_1+\mathbf{h}_2=\mathbf{0}$.

APPENDIX III

$K_{\mathbf{h}_1\mathbf{h}_2}$ in (33)

The 'least-squares' value for the complex quantity $K_{\mathbf{h}_1\mathbf{h}_2}$ in (33) is obtained by minimizing the expression

$$\Big\langle \Big| \sum_{\substack{j_1=1 \\ j_1\neq j_2}}^{p} \sum_{j_2=1}^{p} [g_{j_1}(\mathbf{h}_1)g_{j_1}(\mathbf{h}_2)g_{j_2}(\mathbf{h}_3)g_{j_2}(\mathbf{h}_4) - K_{\mathbf{h}_1\mathbf{h}_2}g_{j_1}(\mathbf{h}_1+\mathbf{h}_2)g_{j_2}(\mathbf{h}_3+\mathbf{h}_4)] \times \exp\left[2\pi i(\mathbf{h}_1+\mathbf{h}_2).(\mathbf{r}_{j_1}-\mathbf{r}_{j_2})\right]\Big|^2\Big\rangle .$$

This leads to

$$K_{\mathbf{h}_1\mathbf{h}_2} = \frac{\sum\limits_{\substack{j_1=1 \\ j_1\neq j_2}}^{p} \sum\limits_{j_2=1}^{p} \langle g_{j_1}(\mathbf{h}_1)g_{j_1}(\mathbf{h}_2)g_{j_1}(\overline{\mathbf{h}_1+\mathbf{h}_2})\rangle \langle g_{j_2}(\mathbf{h}_3)g_{j_2}(\mathbf{h}_4)g_{j_2}(\overline{\mathbf{h}_3+\mathbf{h}_4})\rangle}{\sum\limits_{\substack{j_1=1 \\ j_1\neq j_2}}^{p} \sum\limits_{j_2=1}^{p} \langle |g_{j_1}(\mathbf{h}_1+\mathbf{h}_2)|^2\rangle \langle |g_{j_2}(\mathbf{h}_1+\mathbf{h}_2)|^2\rangle}. \quad \text{(III.1)}$$

If the magnitudes of the normalized (group) scattering factors

$$g_j(\mathbf{h})/\Big[\sum_{j=1}^{p} \langle |g_j(\mathbf{h})|^2\rangle\Big]^{1/2}$$

are small we may include the terms with $j_1=j_2$.

References

BATSCHELET, E. (1965). *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Washington: The American Institute of Biological Sciences.

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.

HAUPTMAN, H. (1965). *Z. Kristallogr.* **121**, 1–8.

HAUPTMAN, H. (1976). *Acta Cryst.* A**32**, 877–882.

HAUPTMAN, H. & FORTIER, S. (1977). *Acta Cryst.* A**33**, 575–580.

HEINERMAN, J. J. L. (1977*a*). *Acta Cryst.* A**33**, 100–106.

HEINERMAN, J. J. L. (1977*b*). In *Proceedings of the Intercongress Symposium on Direct Methods in Crystallography*, Buffalo. To be published.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

KITAIGORODSKII, A. I. (1961). *The Theory of Crystal Structure Analysis*. New York: Consultants Bureau.

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

KOENIG, D. F. (1972). *Acta Cryst.* A**28**, S5.

KOENIG, D. F. (1976). *Chem. Script.* **9**, 14–17.

KROON, J. & KRABBENDAM, H. (1970). *Acta Cryst.* B**26**, 312–314.

MAIN, P. (1976). *Crystallographic Computing Techniques*. Edited by F. R. AHMED, pp. 97–105. Copenhagen: Munksgaard.

MARDIA, K. V. (1972). *Statistics of Directional Data*. London: Academic Press.

MISES, R. VON (1918). *Phys. Z.* **19**, 490–500.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

WOOLFSON, M. M. (1956). *Acta Cryst.* **9**, 804–810.