Acta Cryst. (1975). A**31**, 680

A New Method in the Probabilistic Theory of the Structure Invariants

BY HERBERT HAUPTMAN

Medical Foundation of Buffalo, 73 *High Street, Buffalo, N.Y.* 14203, *U.S.A.*

(*Received* 23 *December* 1974; *accepted* 17 *March* 1975)

It is assumed that a crystal structure in $P1$ is fixed and that the seven non-negative numbers $R_1, R_2, R_3, R_4, R_{12}, R_{23}, R_{31}$ are also specified. The random variables (vectors) $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$ are assumed to be uniformly and independently distributed in the regions of reciprocal space defined by

$$|E_{\mathbf{h}}|=R_1\,,\quad |E_{\mathbf{k}}|=R_2\,,\quad |E_{\mathbf{l}}|=R_3\,,\quad |E_{\mathbf{m}}|=R_4\,, \tag{1}$$

$$|E_{\mathbf{h}+\mathbf{k}}|=R_{12}\,,\quad |E_{\mathbf{k}+\mathbf{l}}|=R_{23}\,,\quad |E_{\mathbf{l}+\mathbf{h}}|=R_{31}\,, \tag{2}$$

and

$$\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}=0\,. \tag{3}$$

Then the structure invariant, $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, as a function of the primitive random variables $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$, is itself a random variable, and its conditional probability distribution, given (1) and (2), is derived and compared with the conditional distribution when only (1) is given. The distribution leads to improved estimates for $\cos\varphi$ in terms of the seven magnitudes (1) and (2). The results secured here suggest a generalization which is described in terms of the 'neighborhood concept'; and a 'principle of nested neighborhoods' is formulated. This terminology permits, in turn, an analogy with interpolation formulas.

1. Introduction

The theory of the structure invariants $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$ initiated recently (Hauptman, 1974*a*,*b*) was based on a joint probability distribution of three structure factors. The theory led to an estimate for the cosine invariant $\cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}})$ dependent on the seven magnitudes

$$|E_{\mathbf{h}}|,|E_{\mathbf{k}}|,|E_{\mathbf{l}}|,|E_{\mathbf{m}}|,|E_{\mathbf{h}+\mathbf{k}}|,|E_{\mathbf{k}+\mathbf{l}}|,|E_{\mathbf{l}+\mathbf{h}}|\,. \tag{1.1}$$

Since this theory was based on a distribution of only three structure factors, it was not possible to take into account the concerted effect of all seven magnitudes (1.1) on the value of the cosine invariant. For this reason the resulting estimate for the cosine was biased, especially when the cosine was negative, and this bias had been tentatively attributed to Patterson overlap. In the light of the new theory described here, which does take into account the mutual correlations among the seven magnitudes (1.1), Patterson overlap is ruled out as the primary cause of the observed bias and the true source is correctly identified and eliminated almost completely. Thus the present theory, which in effect consists of a method for deriving the conditional probability distribution of the structure invariant $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, given the seven magnitudes (1.1), is not only more satisfactory than the earlier one, but yields improved estimates for the cosine invariants as well.

Although only the four-phase structure invariant is treated in detail here, the method used is clearly sufficiently general to be applicable to structure invariants and seminvariants in general, and some preliminary work with the three-phase invariants and selected seminvariants in $P\bar{1}$ has already been carried out.

The joint probability distribution of seven structure factors derived in the preceding paper [Hauptman, 1975, equation (2.5)], is the source of a great variety of conditional distributions. Only two of these are derived here.

2. The joint conditional probability distribution of the four phases $\varphi_{\mathbf{h}}, \varphi_{\mathbf{k}}, \varphi_{\mathbf{l}}, \varphi_{\mathbf{m}}$, given the seven magnitudes $|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}|, |E_{\mathbf{m}}|, |E_{\mathbf{h}+\mathbf{k}}|, |E_{\mathbf{k}+\mathbf{l}}|, |E_{\mathbf{l}+\mathbf{h}}|$

Suppose that a crystal structure consisting of N identical atoms in the space group $P1$ is specified and that the seven non-negative numbers $R_1, R_2, R_3, R_4, R_{12}, R_{23}, R_{31}$ are also fixed. As in the previous paper (Hauptman, 1975) define the fourfold Cartesian product $\mathbf{S}\times\mathbf{S}\times\mathbf{S}\times\mathbf{S}$ of reciprocal space $\mathbf{S}$ to be the collection of all ordered quadruples $(\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m})$ where $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$ are reciprocal vectors. Suppose next that the ordered quadruple of reciprocal vectors $(\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m})$ is a random variable which is uniformly distributed over the subset of $\mathbf{S}\times\mathbf{S}\times\mathbf{S}\times\mathbf{S}$ for which

$$|E_{\mathbf{h}}|=R_1,\ |E_{\mathbf{k}}|=R_2,\ |E_{\mathbf{l}}|=R_3,\ |E_{\mathbf{m}}|=R_4, \tag{2.1}$$

$$|E_{\mathbf{h}+\mathbf{k}}|=R_{12},\ |E_{\mathbf{k}+\mathbf{l}}|=R_{23},\ |E_{\mathbf{l}+\mathbf{h}}|=R_{31}, \tag{2.2}$$

and

$$\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}=0. \tag{2.3}$$

In view of (2.3), the random variables $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$, the components of the ordered quadruple $(\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m})$, are not independently distributed in reciprocal space. Note also that in order to insure that the domain of the random variable $(\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m})$ be non-vacuous, it is necessary to interpret the exact equality $|E_\mathbf{h}|=R_1$ of (2.1), for example, as an inequality, $R_1 \le |E_\mathbf{h}| \le R_1 + \mathrm{d}R_1$, where $\mathrm{d}R_1$ is a 'small' positive quantity; and similarly with the remainder of (2.1) and (2.2). Then $\varphi_\mathbf{h},\varphi_\mathbf{k},\varphi_\mathbf{l},\varphi_\mathbf{m}$, as functions of the primitive random variables $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$, are themselves random variables. Denote by $P(\Phi_1,\Phi_2,\Phi_3,\Phi_4|R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31})$ the joint conditional probability distribution of the four phases $\varphi_\mathbf{h},\varphi_\mathbf{k},\varphi_\mathbf{l},\varphi_\mathbf{m}$, given (2.1), (2.2) and (2.3). Then $P(\Phi_1,\Phi_2,\Phi_3,\Phi_4|R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31})$ is found from (2.5) of Hauptman (1975) by fixing $R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31}$, integrating with respect to $\Phi_{12},\Phi_{23},\Phi_{31}$ from 0 to 2π, and multiplying the result by a suitable normalizing constant. In view of the previous paper, none of these operations presents any difficulty. Thus, in order to carry out the Φ_{12} integration, one collects the terms involving Φ_{12} in the exponent of (2.5) of the earlier paper to obtain

$$\frac{2R_{12}}{N^{1/2}}\,Y_{12}\cos(\Phi_{12}+\eta_{12}), \qquad (2.4)$$

where Y_{12} and η_{12} are independent of Φ_{12}, so that the Φ_{12} integration is immediate. The Φ_{23} and Φ_{31} integrations are done in the same way and one finally obtains, correct up to and including terms of order $1/N$ [since $O(1/N)$ of the previous paper consists of all terms of order $1/N$ or higher in which the terms of order $1/N$ are independent of the Φ's],

$$P(\Phi_1,\Phi_2,\Phi_3,\Phi_4|R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31})$$
$$\approx\frac{1}{K}\exp\{-2B\cos(\Phi_1+\Phi_2+\Phi_3+\Phi_4)\}$$
$$\times I_0\left(\frac{2R_{12}Y_{12}}{N^{1/2}}\right)I_0\left(\frac{2R_{23}Y_{23}}{N^{1/2}}\right)I_0\left(\frac{2R_{31}Y_{31}}{N^{1/2}}\right), \qquad (2.5)$$

where

$$B=\frac{2}{N}\,R_1R_2R_3R_4, \qquad (2.6)$$

$$Y_{12}=[R_1^2R_2^2+R_3^2R_4^2+2R_1R_2R_3R_4\cos(\Phi_1+\Phi_2+\Phi_3+\Phi_4)]^{1/2}, \qquad (2.7)$$

$$Y_{23}=[R_2^2R_3^2+R_1^2R_4^2+2R_1R_2R_3R_4\cos(\Phi_1+\Phi_2+\Phi_3+\Phi_4)]^{1/2}, \qquad (2.8)$$

$$Y_{31}=[R_3^2R_1^2+R_2^2R_4^2+2R_1R_2R_3R_4\cos(\Phi_1+\Phi_2+\Phi_3+\Phi_4)]^{1/2}, \qquad (2.9)$$

K is a suitable normalizing constant, independent of $\Phi_1,\Phi_2,\Phi_3,\Phi_4$, and I_0 is the modified Bessel function. Although K is readily found by integrating (2.5) with respect to $\Phi_1,\Phi_2,\Phi_3,\Phi_4$ from 0 to 2π and setting the result equal to unity, the value of this normalizing factor is not needed for the present purpose and is therefore not derived explicitly.

It is clear from (2.5)–(2.9) that the distribution (2.5) is a function of the sum $\Phi=\Phi_1+\Phi_2+\Phi_3+\Phi_4$. Hence (2.5) leads directly to the conditional distribution, given (2.1) and (2.2), of the sum $\varphi=\varphi_\mathbf{h}+\varphi_\mathbf{k}+\varphi_\mathbf{l}+\varphi_\mathbf{m}$, as is shown in the next section.

3. The conditional probability distribution of the structure invariant $\varphi=\varphi_\mathbf{h}+\varphi_\mathbf{k}+\varphi_\mathbf{l}+\varphi_\mathbf{m}$, given the seven magnitudes, $|E_\mathbf{h}|, |E_\mathbf{k}|, |E_\mathbf{l}|, |E_\mathbf{m}|, |E_{\mathbf{h}+\mathbf{k}}|, |E_{\mathbf{k}+\mathbf{l}}|, |E_{\mathbf{l}+\mathbf{h}}|$

Under the same hypotheses as in § 2, the structure invariant

$$\varphi=\varphi_\mathbf{h}+\varphi_\mathbf{k}+\varphi_\mathbf{l}+\varphi_\mathbf{m} \qquad (3.1)$$

is a random variable whose conditional probability distribution, given (2.1) and (2.2), $P(\Phi|R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31})$, is readily found from (2.5)–(2.9). Thus, correct up to and including terms of order $1/N$, the major result of this paper is given by

$$P(\Phi|R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31})$$
$$\approx\frac{1}{L}\exp\{-2B\cos\Phi\}$$
$$\times I_0\left(\frac{2R_{12}Z_{12}}{N^{1/2}}\right)I_0\left(\frac{2R_{23}Z_{23}}{N^{1/2}}\right)I_0\left(\frac{2R_{31}Z_{31}}{N^{1/2}}\right), \qquad (3.2)$$

where B is defined by (2.6),

$$Z_{12}=[R_1^2R_2^2+R_3^2R_4^2+2R_1R_2R_3R_4\cos\Phi]^{1/2}, \qquad (3.3)$$
$$Z_{23}=[R_2^2R_3^2+R_1^2R_4^2+2R_1R_2R_3R_4\cos\Phi]^{1/2}, \qquad (3.4)$$
$$Z_{31}=[R_3^2R_1^2+R_2^2R_4^2+2R_1R_2R_3R_4\cos\Phi]^{1/2}, \qquad (3.5)$$

and the normalizing constant L is readily found to be

$$L=2\pi\sum_{\substack{\mu,\nu,\varrho\\-\infty}}^{\infty}(-1)^{\mu+\nu+\varrho}I_{\mu\nu\varrho}I_{\mu+\nu+\varrho}(2B), \qquad (3.6)$$

where $I_{\mu\nu\varrho}$ is defined by

$$I_{\mu\nu\varrho}=I_\mu\left(\frac{2R_{12}R_1R_2}{N^{1/2}}\right)I_\mu\left(\frac{2R_{12}R_3R_4}{N^{1/2}}\right)I_\nu\left(\frac{2R_{23}R_2R_3}{N^{1/2}}\right)$$
$$\times I_\nu\left(\frac{2R_{23}R_1R_4}{N^{1/2}}\right)I_\varrho\left(\frac{2R_{31}R_3R_1}{N^{1/2}}\right)I_\varrho\left(\frac{2R_{31}R_2R_4}{N^{1/2}}\right) \qquad (3.7)$$

and I is the modified Bessel function. In the same way the conditional expected value of $\cos\varphi$, given (2.1) and (2.2), is found from (3.2).

$$\epsilon\{\cos\varphi|R_1,R_2,R_3,R_4,R_{12},R_{23},R_{31}\}$$
$$=\frac{\sum\limits_{\substack{\mu,\nu,\varrho\\-\infty}}^{\infty}(-1)^{\mu+\nu+\varrho+1}I_{\mu\nu\varrho}I_{\mu+\nu+\varrho+1}(2B)}{\sum\limits_{\substack{\mu,\nu,\varrho\\-\infty}}^{\infty}(-1)^{\mu+\nu+\varrho}I_{\mu\nu\varrho}I_{\mu+\nu+\varrho}(2B)}. \qquad (3.8)$$

Although (3.6)–(3.8) are explicit expressions for the normalizing parameter L and the conditional expected value of the cosine in terms of the seven magnitudes

(1.1), their values are more readily obtained in any given case by first calculating the distribution (3.2) itself and then computing numerically the value of L (if desired) and the expectation value (3.8). The average value of $|\varphi|$ ($\langle|\varphi|\rangle$) and the most probable value of $|\varphi|$ ($|\varphi|_{\text{mode}}$) are also readily obtainable from the distribution, and all these calculations are sufficiently rapid that it is altogether feasible to do several thousands of them in any given case.

3.1. *The special case* $|E_{\mathbf{h}+\mathbf{k}}|\approx|E_{\mathbf{k}+\mathbf{l}}|\approx|E_{\mathbf{l}+\mathbf{h}}|\approx 0$

In the case that

$$|E_{\mathbf{h}+\mathbf{k}}|=R_{12}\approx 0\,, \tag{3.9}$$

$$|E_{\mathbf{k}+\mathbf{l}}|=R_{23}\approx 0\,, \tag{3.10}$$

$$|E_{\mathbf{l}+\mathbf{h}}|=R_{31}\approx 0\,, \tag{3.11}$$

(3.2) reduces to [since $I_0(0)=1$]

$$P(\Phi\,|\,R_1,R_2,R_3,R_4,R_{12}\approx R_{23}\approx R_{31}\approx 0) \approx \frac{1}{2\pi I_0(2B)}\exp\{-2B\cos\Phi\}\,, \tag{3.12}$$

and in this special case it is readily verified [directly or from (3.8)] that the conditional expected value of $\cos\varphi$, given R_1,R_2,R_3,R_4, and $R_{12}\approx R_{23}\approx R_{31}\approx 0$, is given by

$$\varepsilon\{\cos\varphi\,|\,R_1,R_2,R_3,R_4,R_{12}\approx R_{23}\approx R_{31}\approx 0\} = -\frac{I_1(2B)}{I_0(2B)}\,. \tag{3.13}$$

It need hardly be emphasized that (3.13) is always negative and tends to -1 with increasing B.

Equation (3.13) is to be compared with (3.2) and (4.28) (Hauptman, 1974*a*) of the earlier theory. The presence of $2B$ in the argument of the Bessel functions of (3.13), rather than the B which occurs in the older theory, almost completely eliminates the bias in the older estimates of these cosines which had been previously observed, *i.e.* the estimates were not sufficiently negative. (In order to eliminate the bias completely it is necessary to take into account terms of order $1/N^{3/2}$, as was done in the older theory, but is not done here.) This comparison already clearly illustrates the improvement which results from the ability to take into account the simultaneous interactions of all seven magnitudes on which the estimate for the cosine depends. The inability to do this in the earlier theory, which employed the distribution of only three structure factors instead of the seven used here, biased the estimates of the negative cosines.

4. The conditional probability distribution of the structure invariant $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, given the four magnitudes, $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$, $|E_{\mathbf{l}}|$, $|E_{\mathbf{m}}|$

If, instead of being given the seven magnitudes (2.1) and (2.2), one is given only the four magnitudes (2.1), then the conditional probability distribution of φ, given (2.1), is found from (3.1) of Hauptman (1975) to be, correct up to and including terms of order $1/N$,

$$P(\Phi\,|\,R_1,R_2,R_3,R_4)\approx\frac{1}{2\pi I_0(B)}\exp\{B\cos\Phi\}. \tag{4.1}$$

In this case one finds the conditional expected value of $\cos\varphi$ to be

$$\varepsilon\{\cos\varphi\,|\,R_1,R_2,R_3,R_4\}=\frac{I_1(B)}{I_0(B)} \tag{4.2}$$

which should be compared with (3.13). This comparison clearly illustrates the dramatic changes which may take place when all seven magnitudes (2.2) and (2.1) are assumed to be given instead of merely the four magnitudes (2.1). In sharp contrast to (3.13), (4.2) is always positive and tends to unity with increasing B. Comparison of (4·1) with (3.2) is also illuminating and is done in some detail in the next section.

5. The applications

If only the four magnitudes (2.1) are known, the conditional distribution of the cosine invariant $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$ is given by (4.1). If all seven magnitudes (2.1) and (2.2) are known, the conditional distribution of φ is given by (3.2). It is remarkable that knowledge of only the three additional magnitudes

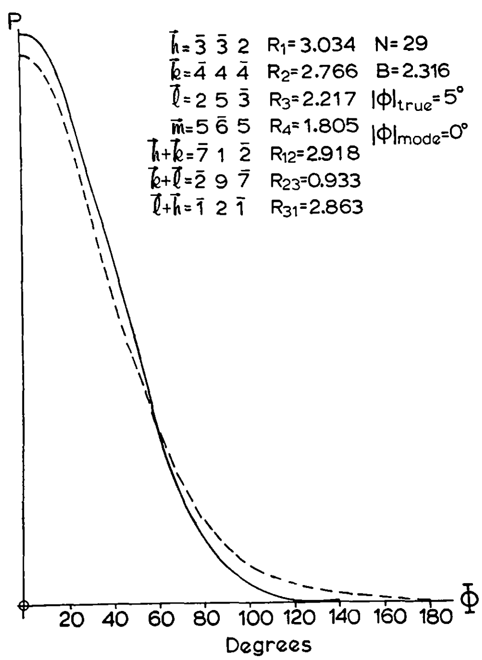

Fig. 1. The distributions (3·2) (——) and (4·1) (– – –) for the values (2·1) and (2·2) shown. The mode of (3·2) is 0, of (4·1) always 0.

(2.2) can effect enormous change in the shape of the distribution [as the comparison of (4.2) with (3.13) already suggests]. Thus (4.1) always has a unique maximum at $\varphi=0$. Hence if one is given only the four

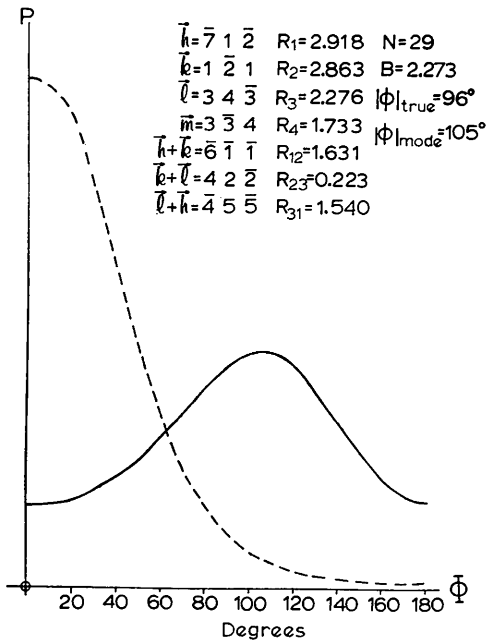

Fig. 2. The distributions (3·2) (——) and (4·1) (---) for the values (2·1) and (2·2) shown. The mode of (3·2) is 105°, of (4·1) always 0.

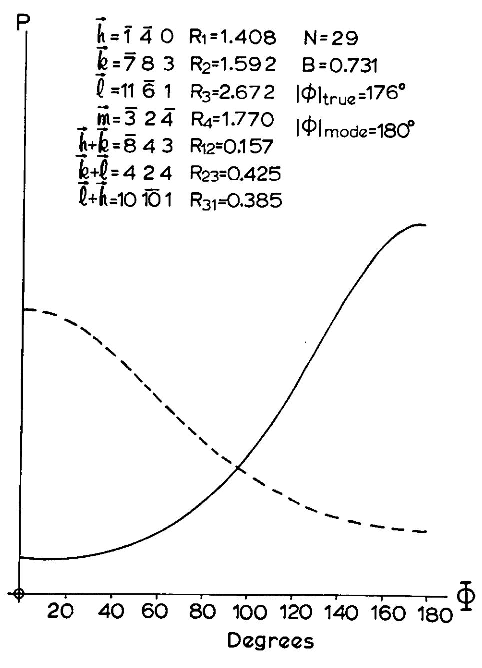

Fig. 3. The distributions (3·2) (——) and (4·1) (---) for the values (2·1) and (2·2) shown. The mode of (3·2) is 180°, of (4·1) always 0.

magnitudes (2.1), then the most probable value of φ is zero and the larger the value of B the more likely it is that $\varphi\approx 0$. The maximum of (3.2), on the other hand, may lie anywhere between 0 and 180° (or, owing to symmetry, between $-180°$ and 0). Thus if one is given all seven magnitudes (2.1) and (2.2), the most probable value of $|\varphi|$ may be 0 or 180° or any value between these limits. In particular, if R_{12}, R_{23}, R_{31} are all very large then it is likely that $|\varphi|$ is near zero, while if R_{12}, R_{23}, R_{31} are all very small then the most probable value of $|\varphi|$ is near 180°. It is remarkable in fact, as reference to (3.2) and (3.12) shows, that, if (3.9) – (3.11) hold, then $|\varphi|$ is almost certainly close to 180° (unless B is very small) and the larger the value of B the closer to 180° $|\varphi|$ is likely to be. These properties of the distributions are illustrated by Figs. 1–3 showing the conditional distributions (3.2) (solid line) and (4.1) (dashed line) for several representative sets of values (2.1) and (2.2) for a hypothetical 29-atom structure in $P1$. [Actually only the distributions of $|\varphi|$ are plotted but, owing to symmetry, these are the same as (3.2) and (4.1) except for a doubling of the scaling parameter. In effect then, Figs. 1–3 are the restrictions of (3.2) and (4.1) to the interval $0\leq\varphi\leq 180°$.] It is noteworthy that in all cases (some 160 distributions have been calculated) the conditional distribution (3.2) has been found to be unimodal, *i.e.* has precisely one maximum, in the interval 0–180°. It is conjectured that this distribution is always unimodal.

Next, an idealized structure consisting of $N=29$ identical point atoms in the space group $P1$ was constructed (the same structure as that used for Figs. 1–3 and in Hauptman, 1974*a*, *b*) and cosine invariants calculated as shown in Tables 1–3.* In all cases three estimates were obtained by using the distribution (3.2). The first, columns 9, were simply the average values of $\cos\varphi$ obtained numerically from (3.2). [As already explained this is more efficient than using the explicit expression (3.8) which requires the calculation of two triple sums.] The second estimates, columns 11, are the values of $\cos(\langle|\varphi|\rangle)$ where $\langle|\varphi|\rangle$ is the expected value of $|\varphi|$, again obtained numerically from the distribution (3.2). The final estimates, columns 13, are the values of $\cos(|\varphi|_m)$ where $|\varphi|_m$ is the most probable value of $|\varphi|$ (the mode), obtained by inspection from the distribution (3.2) by simply reading off the position of the (unique) maximum in the interval 0 to 180°. The columns 14 contain the true values of the cosines, $\cos\varphi_t$, and columns 15–17 the discrepancies between the true cosine values and the three estimates respectively. Thus each entry in the columns 15 is the difference between the corresponding entries in the

* Only fragments of Tables 1–3 are included here. The complete Tables 1–3 have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 31039 (4 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 13 White Friars, Chester CH1 1NZ, England.

Table 1. 31 *invariants* φ *estimated* [*from* (3.2)] *to be in the neighborhood of* 180° *with B in the interval* $0{\cdot}760 < B < 0{\cdot}900$

Column number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Serial number	**h** $\lvert E\rvert$	**k** $\lvert E\rvert$	**l** $\lvert E\rvert$	**m** $\lvert E\rvert$	**h+k** $\lvert E\rvert$	**k+l** $\lvert E\rvert$	**l+h** $\lvert E\rvert$	B	$\langle\cos\varphi\rangle$ from (3·2)	$\langle\lvert\varphi\rvert\rangle$ from (3·2) (°)	$\cos\langle\lvert\varphi\rvert\rangle$	$\lvert\varphi\rvert_m$ from (3·2) (°)	$\cos\lvert\varphi\rvert_m$	$\cos\varphi_t$	$\Delta_1=$ (14)–(9)	$\Delta_2=$ (14)–(11)	$\Delta_3=$ (14)–(13)
501	5 1 $\bar{6}$	0 $\bar{1}$ 1	1 $\bar{3}$ 1	$\bar{6}$ 3 4	5 0 $\bar{3}$	1 $\bar{4}$ 2	6 $\bar{2}$ $\bar{3}$										
	2·158	1·803	1·864	1·794	0·327	0·474	0·470	0·897	−0·574	133	−0·682	180	−1·000	−0·939	−0·365	−0·257	0·061
502	0 5 0	0 $\bar{2}$ 1	1 $\bar{2}$ 1	$\bar{1}$ $\bar{1}$ $\bar{2}$	0 3 1	1 $\bar{4}$ 2	1 3 1										
	1·258	1·644	2·863	2·171	0·421	0·474	0·427	0·887	−0·564	132	−0·669	180	−1·000	−0·757	−0·193	−0·088	0·243
503	$\bar{4}$ $\bar{2}$ 5	$\bar{4}$ 5 $\bar{4}$	9 $\bar{3}$ $\bar{1}$	$\bar{1}$ 2 0	$\bar{8}$ 3 1	5 0 $\bar{5}$	5 $\bar{7}$ 4										
	1·487	1·771	2·202	2·217	0·190	0·327	0·231	0·887	−0·643	139	−0·755	180	−1·000	−0·591	0·052	0·164	0·409
504	5 $\bar{3}$ 7	3 3 $\bar{5}$	$\bar{2}$ 4 $\bar{1}$	$\bar{6}$ $\bar{2}$ $\bar{1}$	8 $\bar{2}$ 2	1 7 $\bar{6}$	3 $\bar{1}$ 6										
	1·749	2·001	1·815	2·020	0·254	0·463	0·398	0·885	−0·594	135	−0·707	180	−1·000	−0·646	−0·051	0·060	0·354
505	$\bar{2}$ 4 $\bar{2}$	2 2 1	2 5 $\bar{2}$	$\bar{2}$ $\bar{1}\bar{1}$ 3	0 6 $\bar{1}$	4 7 $\bar{1}$	0 9 $\bar{4}$										
	2·166	1·602	1·726	2·139	0·259	0·277	0·403	0·883	−0·621	137	−0·731	180	−1·000	−0·875	−0·254	−0·144	0·125

Table 2. 32 *invariants* φ *estimated* [*from* (3.2)] *to be in the neighborhood of* 180° *with B in the interval* $0{\cdot}900 < B < 2{\cdot}160$

Column number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Serial number	**h** $\lvert E\rvert$	**k** $\lvert E\rvert$	**l** $\lvert E\rvert$	**m** $\lvert E\rvert$	**h+k** $\lvert E\rvert$	**k+l** $\lvert E\rvert$	**l+h** $\lvert E\rvert$	B	$\langle\cos\varphi\rangle$ from (3·2)	$\langle\lvert\varphi\rvert\rangle$ from (3·2) (°)	$\cos\langle\lvert\varphi\rvert\rangle$	$\lvert\varphi\rvert_m$ from (3·2) (°)	$\cos\lvert\varphi\rvert_m$	$\cos\varphi_t$	$\Delta_1=$ (14)–(9)	$\Delta_2=$ (14)–(11)	$\Delta_3=$ (14)–(13)
101	$\bar{1}$ $\bar{5}$ $\bar{2}$	4 4 0	4 $\bar{3}$ 5	$\bar{7}$ 4 $\bar{3}$	3 $\bar{1}$ $\bar{2}$	8 1 5	3 $\bar{8}$ 3										
	3·087	2·242	2·136	2·014	0·674	0·922	0·282	2·053	−0·726	145	−0·819	180	−1·000	−0·507	0·219	0·312	0·493
103	$\bar{1}$ 2 $\bar{1}$	11 $\bar{6}$ 1	$\bar{3}$ $\bar{4}$ 3	$\bar{7}$ 8 $\bar{3}$	10 $\bar{4}$ 0	8 $\bar{10}$ 4	$\bar{4}$ $\bar{2}$ 2										
	2·862	2·672	2·275	1·700	1·106	0·404	0·222	2·040	−0·758	147	−0·839	180	−1·000	−0·950	−0·192	−0·111	0·050
104	$\bar{8}$ 1 3	1 8 1	1 $\bar{3}$ 1	6 $\bar{6}$ $\bar{5}$	$\bar{7}$ 9 4	2 5 2	$\bar{7}$ $\bar{2}$ 4										
	2·314	1·973	1·864	1·576	0·661	0·253	0·111	0·925	−0·602	135	−0·707	180	−1·000	−0·999	−0·397	−0·292	0·001
105	$\bar{10}$ 4 $\bar{1}$	7 $\bar{4}$ 3	7 $\bar{8}$ $\bar{3}$	$\bar{4}$ 8 1	$\bar{3}$ 0 2	14 $\bar{12}$ 0	$\bar{3}$ $\bar{4}$ $\bar{4}$										
	2·454	2·014	1·591	1·870	0·398	0·612	0·177	1·015	−0·623	137	−0·731	180	−1·000	−0·982	−0·359	−0·251	0·018

Table 3. 30 *cosine invariants* cos φ *estimated* [*from* (3.2)] *to be in the middle range* (−0·1 *to* +0·8, *column* 9) *with* B *in the interval* $2{\cdot}29 < B < 2{\cdot}50$

Column number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Serial number	**h** $\|E\|$	**k** $\|E\|$	**l** $\|E\|$	**m** $\|E\|$	**h+k** $\|E\|$	**k+l** $\|E\|$	**l+h** $\|E\|$	B	$\langle\cos\varphi\rangle$ from (3·2)	$\langle\|\varphi\|\rangle$ from (3·2) (°)	$\cos\langle\|\varphi\|\rangle$	$\|\varphi\|_m$ from (3·2) (°)	$\cos\|\varphi\|_m$	$\cos\varphi_t$	$\Delta_1=$ (14)–(9)	$\Delta_2=$ (14)–(11)	$\Delta_3=$ (14)–(13)
601	$\bar{1}\ \bar{5}\ \bar{2}$	$\bar{3}\ 4\ 4$	$\bar{4}\ 8\ \bar{2}$	$8\ \bar{7}\ 0$	$\bar{4}\ \bar{1}\ 2$	$\bar{7}\,12\,2$	$\bar{5}\ 3\ \bar{4}$										
	3·088	2·489	2·183	2·153	1·872	0·652	1·869	2·491	0·438	58	0·530	60	0·500	1·000	0·562	0·470	0·500
602	$\bar{1}\ \bar{5}\ \bar{2}$	$\bar{7}\ 1\ \bar{2}$	$7\ \bar{4}\ 3$	$1\ 8\ 1$	$\bar{8}\ \bar{4}\ \bar{4}$	$0\ \bar{3}\ 1$	$6\ \bar{9}\ 1$										
	3·088	2·918	2·015	1·973	2·135	0·907	2·141	2·471	0·638	43	0·731	0	1·000	0·982	0·344	0·251	−0·018
603	$\bar{1}\ \bar{5}\ \bar{2}$	$3\ 3\ \bar{2}$	$\bar{8}\ 1\ 3$	$6\ 1\ 1$	$2\ \bar{2}\ \bar{4}$	$\bar{5}\ 4\ 1$	$\bar{9}\ \bar{4}\ 1$										
	3·088	3·034	2·314	1·631	2·143	0·998	2·409	2·439	0·690	38	0·788	0	1·000	0·851	0·160	0·063	−0·149
604	$\bar{3}\ \bar{3}\ 2$	$\bar{1}\ 2\ \bar{1}$	$\bar{7}\ 7\ \bar{3}$	$11\,\bar{6}\,2$	$\bar{4}\ \bar{1}\ 1$	$\bar{8}\ 9\ \bar{4}$	$\overline{10}\,4\,\bar{1}$										
	3·034	2·863	2·023	2·012	1·732	0·997	2·454	2·438	0·656	41	0·755	0	1·000	0·949	0·293	0·194	−0·051
605	$0\ \bar{7}\ \bar{2}$	$\bar{7}\ 1\ \bar{2}$	$\bar{1}\ 2\ \bar{1}$	$8\ 4\ 5$	$\bar{7}\ \bar{6}\ \bar{4}$	$\bar{8}\ 3\ \bar{3}$	$\bar{1}\ 5\ \bar{3}$										
	3·054	2·918	2·863	1·385	1·325	0·867	2·073	2·437	0·223	73	0·292	80	0·174	0·317	0·094	0·025	0·143

columns 14 and 9, *etc.* Table 4 is a summary of Tables 1–3.

It is clear from the tables that the present theory leads to estimates for the cosine invariants which are substantially better than those derived from the earlier theory. The estimates for those cosines calculated to be most positive had already been reliable even with the older theory (Hauptman, 1974*a*, Table 2) and this calculation is not repeated here. Those cosines calculated to be most negative reliably identified the negative cosines, even in the older theory, but quantitative agreement had not been quite attained earlier, and an inexplicable bias had also been observed. As shown by the present Tables 1, 2, and 4 the present estimates for the negative cosines are noticeably improved and the bias substantially reduced. In fact the bias is virtually eliminated if the estimate $\cos(\langle|\varphi|\rangle)$ is used even when B values are as small as 0·8. (Naturally the observed bias in the estimates $\cos(|\varphi|_m)$ of Tables 1 and 2 is a necessary consequence of the requirement that $|\varphi|_m = 180°$ for the 63 invariants in these tables.)

Those cosines estimated to be in the middle range had been, under the old theory, in poor agreement with the true values. As shown by Tables 3 and 4 the present theory yields substantially better agreement, especially if the estimates $\cos(\langle|\varphi|\rangle)$ or $\cos(|\varphi|_m)$ are used. However these cosines are still determined least reliably because of the relatively large value of the variance (compare Fig. 2 with Figs. 1 and 3).

In summary then, it seems clear that the present theory yields estimates for the cosines which are already sufficiently reliable to justify further tests and even efforts to make applications to structure determination. Discrepancies which still persist probably arise from a number of causes: (1) the probabilistic nature of the estimates; (2) the omission of terms of order higher than $1/N$; (3) artifacts of the structure, *e.g.* Patterson overlap, pseudo-symmetry, rational dependence among atomic coordinates, *etc.*

Finally, it should be pointed out that initial applications to the solution of unknown structures of the four-phase invariants have already been made (*e.g.* DeTitta, Edmonds, Langs & Hauptman, 1974; Einspahr, Gartland, Freeman & Schenk, 1974). It is anticipated that with the improvements described in this paper more effective use of these invariants will be made in the near future.

6. The neighborhood concept and the principle of nested neighborhoods

It has been known for many years that provided no homometric solutions (other than enantiomorphs) exist the structure factor magnitudes $|E|$ uniquely determine, in general, the values of the cosine invariants, and therefore, except for a twofold ambiguity in sign, the values of the structure invariants. In fact earlier formulas, *e.g.* $B_{3,0}$ (Karle & Hauptman, 1958), the modified triple product (Hauptman, 1964; Hauptman, Fischer, Hancock & Norton, 1969), and MDKS (Hauptman, 1972, p. 192) are dependent on all available structure-factor magnitudes or at least large numbers of them. The methods and results of this paper and the preceding one make plausible a refinement of this perception. It is now suggested that the value of a cosine invariant is primarily determined by the values of one or more small sets of appropriately selected structure-factor magnitudes, and is relatively insensitive to the values of the great bulk of the remaining magnitudes. Thus a rough estimate, in the probabilistic sense, of the magnitude of the structure invariant

$$\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}} \tag{6.1}$$

depends in the first instance on the four magnitudes

$$|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}|, |E_{\mathbf{m}}|\,, \tag{6.2}$$

and the larger the value of B (2.6) the more reliable is the estimate, zero in this case (if one uses the mode: near zero if the mean is used). If one adjoins to the set (6.2) the three magnitudes

$$|E_{\mathbf{h}+\mathbf{k}}|, |E_{\mathbf{k}+\mathbf{l}}|, |E_{\mathbf{l}+\mathbf{h}}| \tag{6.3}$$

then one obtains a different estimate, in general, for $|\varphi|$ [dependent on the seven magnitudes (6.2) and (6.3)], and this estimate may lie anywhere between 0 and π. The reliability of the latter estimate depends on both B and the estimate, the larger the value of B and the closer to 0 or to π the value of the estimate, the greater the reliability of the estimate. Estimates in the middle range (near $\pi/2$) or for small values of B are least reliable. Since the first estimate, dependent on only the four magnitudes (6.2), uses less information than the second estimate, dependent on the seven magnitudes (6.2) and (6.3), it is natural to expect that the latter will, at least in favorable cases, be more reliable than the former, and this in fact turns out to be the case. In any event a measure of the reliability is given by the variance of the conditional probability distribution of φ, (4.1) or (3.2) respectively. It is natural to conjecture that there exists an additional small set of magnitudes which, when added to the sets (6.2) and (6.3), will, in combination with (6.2) and (6.3), yield a still better estimate for $|\varphi|$; *etc.* One is led in this way to the concept of nested neighborhoods of a structure invariant depicted schematically in Fig. 4 for the invariant (6.1). The first neighborhood consists of the four vectors $\mathbf{h}, \mathbf{k}, \mathbf{l}, \mathbf{m}$ [or the corresponding magnitudes (6.2)]. The second neighborhood is the set-theoretic union of the first neighborhood and the three additional vectors $\mathbf{h}+\mathbf{k}, \mathbf{k}+\mathbf{l}, \mathbf{l}+\mathbf{h}$ [or the associated magnitudes (6.3)] shown in the second shell of Fig. 4. The identity of the third shell of vectors (or magnitudes) indicated by the ?'s of Fig. 4, to be added to the second neighborhood in order to obtain the third, is an open question. One proceeds in this way to obtain a sequence of nested neighborhoods of the invariant φ, each a subset of the succeeding one and having the property defined by the principle of nested neighborhoods: There exists an estimate for the cosine invariant $\cos \varphi$ (or $|\varphi|$) dependent on the magnitudes $|E|$ constituting any neighborhood of φ, and the more magnitudes in the neighborhood the better the estimate, in the probabilistic sense.

It should be emphasized that the principle of nested neighborhoods is not only strongly suggested by the available evidence, but some preliminary work on the three-phase structure invariants, $\varphi_1+\varphi_2+\varphi_3$, and the two-phase structure seminvariants, $\varphi_1+\varphi_2$, provides additional evidence that the principle holds for structure invariants and seminvariants in general, not merely the quartets (6.1). The central question is

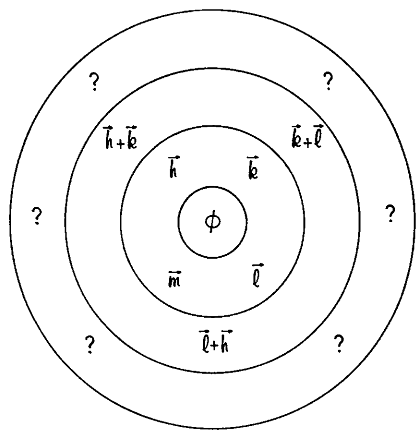

Fig. 4. A sequence of nested neighborhoods for the structure invariant (6·1).

Table 4. *Summary of Tables* 1–3 *with a comparison of some earlier results*

	Average error in cosine			Average magnitude of the error			Average value of B	Number of contributors to the average	Nature of estimated cosine
	$\langle\Delta_1\rangle$	$\langle\Delta_2\rangle$	$\langle\Delta_3\rangle$	$\langle\lvert\Delta_1\rvert\rangle$	$\langle\lvert\Delta_2\rvert\rangle$	$\langle\lvert\Delta_3\rvert\rangle$	$\langle B\rangle$		
From Table 1	−0·229	−0·119	0·194	0·276	0·209	0·194	0·831	31	Negative
From Table 2	−0·118	−0·006	0·193	0·221	0·183	0·193	1·205	32	Negative
From Table 3	0·274	0·188	0·015	0·275	0·206	0·185	2·369	30	Middle range
From Table 3 of Hauptman (1974*b*)	−0·362			0·389			1·231	33	Negative
From Table 4 of Hauptman (1974*b*)	−0·298			0·527			2·044	25	Middle range

how to identify a sequence of nested neighborhoods for a given structure invariant or seminvariant. A beginning only has been made for the structure invariants (6.1).

Using the present terminology then, the results described here partake of the character of an interpolation formula: The value of the structure invariant φ, or cosine invariant cos φ, is estimated by means of the values of the magnitudes $|E|$ 'in the neighborhood' of φ, and the more magnitudes that are used the better, in general, the estimate.

7. Concluding remarks

Conditional distributions of the structure invariant $\varphi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{m}}$, given, in the first instance, only the four magnitudes (2.1) and, in the second instance, all seven magnitudes (2.1) and (2.2), have been found. The distributions lead to estimates for $|\varphi|$ and cos φ dependent on these magnitudes. The importance of the newly acquired ability to take into account the concerted effect of these magnitudes on the value of the invariant has been stressed. It is suggested that conditional distributions, given additional magnitudes, will yield still better estimates for the invariants. It is noteworthy that the first application of the new method is to the four-phase invariant φ rather than to the much more familiar and more intensively studied three-phase invariant $\varphi_1 + \varphi_2 + \varphi_3$. However the available evidence strongly suggests that the same methods will find important application in the study of the three-phase invariants as well as the five-phase invariants $\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 + \varphi_5$, and, owing to the importance of these invariants in the applications, it is anticipated that this study will be vigorously pursued. Further areas of investigation include the space-group-dependent probability distributions, particularly in $P\bar{1}, P2_1, P2_12_12_1$, and distributions of the structure seminvariants. It should be mentioned in conclusion that the existence of bimodal or multimodal distributions in the whole interval $(-\pi, \pi)$, *e.g.* Fig. 2, opens up the possibility that these distributions may serve as the vehicle for sorting out the different members of homometric sets when they exist (in much the same way as a distribution like Fig. 2 serves to distinguish the enantiomorphs).

The author acknowledges with gratitude a number of stimulating conversations with Drs Edward Green and David Langs; Dr Langs also calculated the distributions shown in the figures as well as the data in the tables. This research was supported by DHEW Grants Nos. RR05716 and HL15378 and NSF Grant No. MPS73-04992.

References

DETITTA, G., EDMONDS, J., LANGS, D. & HAUPTMAN, H. (1974). Amer. Cryst. Assoc. Summer Meeting at Penn. State Univ. Abstract D3.

EINSPAHR, H., GARTLAND, G., FREEMAN, G. & SCHENK, H. (1974). Amer. Cryst. Assoc. Summer Meeting at Penn. State Univ. Abstract B6.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421–1433.

HAUPTMAN, H. (1972). *Crystal Structure Determination: The Role of the Cosine Seminvariants.* New York, London: Plenum.

HAUPTMAN, H. (1974*a*). *Acta Cryst.* A**30**, 472–476.

HAUPTMAN, H. (1974*b*). *Acta Cryst.* A**30**, 822–829.

HAUPTMAN, H. (1975). *Acta Cryst.* A**31**, 671–679.

HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* B**25**, 811–814.

KARLE, J. & HAUPTMAN, H. (1958). *Acta Cryst.* **11**, 264–269.