

On the Identity and Estimation of those Cosine Invariants, $\text{Cos}(\varphi_m + \varphi_n + \varphi_p + \varphi_q)$, which are Probably Negative

BY HERBERT HAUPTMAN*

Medical Foundation of Buffalo, Buffalo, New York 14203, U.S.A.

(Received 26 November 1973; accepted 18 February 1974)

If $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ are fixed reciprocal vectors which satisfy $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = \mathbf{0}$, and if \mathbf{k} is the primitive, uniformly distributed random variable, then, under the assumption that each of $|E_{\mathbf{h}_1}|, |E_{-\mathbf{h}_3+\mathbf{k}}|$ is sufficiently small, the conditional probability distribution of the cosine invariant $\text{cos}(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}_1-\mathbf{k}} + \varphi_{-\mathbf{h}_2} + \varphi_{-\mathbf{h}_3})$, given $|E_{-\mathbf{h}_3+\mathbf{k}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}_1+\mathbf{k}}|$, is found. The distribution leads to the surprising result that the conditional expected value of this cosine invariant is always negative and approaches -1 with increasing $|E_{\mathbf{k}}E_{\mathbf{h}_1+\mathbf{k}}E_{\mathbf{h}_2}E_{\mathbf{h}_3}|$. If $\mathbf{m}, \mathbf{n}, \mathbf{p}, \mathbf{q}$ are fixed reciprocal vectors satisfying $\mathbf{m} + \mathbf{n} + \mathbf{p} + \mathbf{q} = \mathbf{0}$, suitable sampling of reciprocal space then leads to a formula for the cosine invariant $\text{cos}(\varphi_{\mathbf{m}} + \varphi_{\mathbf{n}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}})$ having probabilistic validity in the case that $|E_{\mathbf{m}+\mathbf{n}}|, |E_{\mathbf{m}+\mathbf{p}}|$ and $|E_{\mathbf{m}+\mathbf{q}}|$ are sufficiently small. It follows, in particular, that under the stated conditions the value of the cosine is probably negative and the larger the value of $|E_{\mathbf{m}}E_{\mathbf{n}}E_{\mathbf{p}}E_{\mathbf{q}}|$ the more negative the cosine is likely to be.

1. Introduction

Explicit formulas for the cosine seminvariants $\text{cos} \varphi$ and $\text{cos}(\varphi_1 + \varphi_2)$, having exact validity under certain conditions, are now known for a number of space groups, and the algebraic techniques for deriving similar formulas in most of the other space groups have been described (Hauptman & Karle, 1953; Hauptman, 1972*a, b*). Both algebraic and probabilistic methods are available for estimating the value of the cosine invariants $\text{cos}(\varphi_1 + \varphi_2 + \varphi_3)$. Thus it is known that the conditional expected value of this cosine, given $|E_1E_2E_3|$, is always positive and approaches unity with increasing $|E_1E_2E_3|$. However, except for some recent semi-empirical observations on invariants of special type by Schenk & de Jong (1973), no theoretical estimate has hitherto been known for the general cosine invariants, $\text{cos}(\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4)$, which are dependent on four phases. A major goal of the present paper is to derive an estimate for the cosine invariant $\text{cos}(\varphi_{\mathbf{m}} + \varphi_{\mathbf{n}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}})$ under the condition that each of $|E_{\mathbf{m}+\mathbf{n}}|, |E_{\mathbf{m}+\mathbf{p}}|, |E_{\mathbf{m}+\mathbf{q}}|$ is very small, and it is shown, in particular, that the conditional expected value of this cosine, given $|E_{\mathbf{m}}E_{\mathbf{n}}E_{\mathbf{p}}E_{\mathbf{q}}|$, is always negative and approaches -1 with increasing $|E_{\mathbf{m}}E_{\mathbf{n}}E_{\mathbf{p}}E_{\mathbf{q}}|$. Since the identity of those cosine invariants which are small or negative is of crucial importance in direct methods of phase determination, it is anticipated that the unexpected results obtained here will have important application in the further development of these procedures.

2. For fixed \mathbf{h}_1 and \mathbf{h}_3 , the conditional distribution of the pair $\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}_1+\mathbf{k}}$, given $|E_{-\mathbf{h}_3+\mathbf{k}}|, |E_{\mathbf{k}}|$ and $|E_{\mathbf{h}_1+\mathbf{k}}|$

Fix the reciprocal vectors $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ subject to

$$\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = \mathbf{0}, \quad (2.1)$$

* Part of this work was done while the author was a visiting fellow in Italy under the auspices of the Consiglio Nazionale delle Ricerche, March 15–May 15, 1973.

and assume that a crystal structure, in the space group $P1$, is also fixed. As usual, denote by N the number of atoms, assumed identical, in the unit cell and by φ the phase of the normalized structure factor E , and introduce the notation

$$E_{\mathbf{h}_j} = E_j, |E_{\mathbf{h}_j}| = |E_j|, \varphi_{\mathbf{h}_j} = \varphi_j, \quad j=1,2,3. \quad (2.2)$$

Suppose that the vector \mathbf{k} is the primitive random variable which is assumed to be uniformly distributed throughout reciprocal space. Then $E_{-\mathbf{h}_3+\mathbf{k}}, E_{\mathbf{k}}, E_{\mathbf{h}_1+\mathbf{k}}$, as functions of the random variable \mathbf{k} , are themselves random variables with joint probability distribution $P(R_1, R_2, R_3; \Phi_1, \Phi_2, \Phi_3)$ where R_1 is associated with $|E_{-\mathbf{h}_3+\mathbf{k}}|$, R_2 with $|E_{\mathbf{k}}|$, R_3 with $|E_{\mathbf{h}_1+\mathbf{k}}|$, Φ_1 with $\varphi_{-\mathbf{h}_3+\mathbf{k}}$, Φ_2 with $\varphi_{\mathbf{k}}$, and Φ_3 with $\varphi_{\mathbf{h}_1+\mathbf{k}}$. An expression for $P(R_1, R_2, R_3; \Phi_1, \Phi_2, \Phi_3)$ sufficiently accurate for all values of the parameters E_1, E_2, E_3 and the whole range of values of the variables $R_1, R_2, R_3, \Phi_1, \Phi_2, \Phi_3$ to be useful here has been obtained recently (Tsoucaris, 1970; Hauptman, 1971, 1972*a*, p. 165), and, correct to terms of order $1/N$, is given by

$$\begin{aligned} P(R_1, R_2, R_3; \Phi_1, \Phi_2, \Phi_3) \simeq & \frac{R_1 R_2 R_3}{\pi^3 \Delta} \\ & \times \exp \left\{ -\frac{1}{\Delta} \left[R_1^2 \left(1 - \frac{|E_1|^2}{N} \right) + R_2^2 \left(1 - \frac{|E_2|^2}{N} \right) \right. \right. \\ & \left. \left. + R_3^2 \left(1 - \frac{|E_3|^2}{N} \right) \right] \right\} \\ & \times \exp \left\{ \frac{2}{N^{1/2} \Delta} \left[R_1 R_2 |E_3| \cos(\Phi_1 - \Phi_2 + \varphi_3) \right. \right. \\ & + R_2 R_3 |E_1| \cos(\Phi_2 - \Phi_3 + \varphi_1) \\ & \left. \left. + R_3 R_1 |E_2| \cos(\Phi_3 - \Phi_1 + \varphi_2) \right] \right\} \\ & \times \exp \left\{ -\frac{2}{N \Delta} \left[R_1 R_2 |E_1 E_2| \cos(\Phi_1 - \Phi_2 - \varphi_1 - \varphi_2) \right. \right. \\ & \left. \left. + R_2 R_3 |E_2 E_3| \cos(\Phi_2 - \Phi_3 - \varphi_2 - \varphi_3) \right. \right. \end{aligned}$$

$$\begin{aligned}
& + R_3 R_1 |E_3 E_1| \cos(\Phi_3 - \Phi_1 - \varphi_3 - \varphi_1) \Big\} \\
& \times \left\{ 1 - \frac{1}{4N} (R_1^4 + R_2^4 + R_3^4 + 4R_1^2 R_2^2 + 4R_2^2 R_3^2 \right. \\
& \left. + 4R_3^2 R_1^2 - 12R_1^2 - 12R_2^2 - 12R_3^2 + 18) \right\} \quad (2.3)
\end{aligned}$$

where

$$\begin{aligned}
A = 1 - \frac{1}{N} (|E_1|^2 + |E_2|^2 + |E_3|^2) + \frac{2}{N^{3/2}} \\
\times |E_1 E_2 E_3| \cos(\varphi_1 + \varphi_2 + \varphi_3). \quad (2.4)
\end{aligned}$$

Next, denote by $P(\Phi_2, \Phi_3 | R_1, R_2, R_3)$ the conditional joint probability distribution of the pair of phases $\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}_1 + \mathbf{k}}$, given that R_1, R_2, R_3 have fixed, specified values. Then $P(\Phi_2, \Phi_3 | R_1, R_2, R_3)$ is obtained from $P(R_1, R_2, R_3; \Phi_1, \Phi_2, \Phi_3)$ by integrating the latter with respect to Φ_1 from 0 to 2π , fixing R_1, R_2, R_3 , and multiplying the result by a suitable normalizing factor. This integration has already been carried out in a different context (Hauptman, 1971, 1972a, pp. 167–170). Refer to equations (4.3) and (4.6) on pages 168 and 170 of the latter reference and employ the Bessel Function expansion

$$I_0(z) \simeq 1 + \frac{z^2}{4} \simeq \exp\left(\frac{z^2}{4}\right)$$

if z is small. Since R_1, R_2, R_3 are now regarded as fixed parameters rather than variables, the conditional distribution, correct to terms of order $1/N$, is readily found to be (if R_1 is not too large)

$$\begin{aligned}
& P(\Phi_2, \Phi_3 | R_1, R_2, R_3) \\
& \simeq \frac{1}{K} \exp \left\{ \frac{2R_2 R_3 |E_1| \cos(\Phi_2 - \Phi_3 + \varphi_1)}{\Delta N^{1/2}} \right. \\
& \quad \left. \frac{2R_2 R_3 \left(1 - \frac{R_1^2}{A}\right) |E_2 E_3|}{\Delta N} \right. \\
& \quad \left. \times \cos(\Phi_2 - \Phi_3 - \varphi_2 - \varphi_3) \right\} \quad (2.5)
\end{aligned}$$

where K is a suitable normalizing constant. Assume next that R_1^2 is small compared to unity and that $|E_1|$ is small compared to $|E_2 E_3|/N^{1/2}$, *i.e.*

$$R_1^2 \ll 1, \quad |E_1| \ll |E_2 E_3|/N^{1/2}. \quad (2.6)$$

Then (2.5) becomes (*cf.* Hauptman, 1972a, p. 144)

$$\begin{aligned}
& P(\Phi_2, \Phi_3 | R_1, R_2, R_3) \simeq \frac{1}{K} \\
& \times \exp \left\{ - \frac{2R_2 R_3 |E_2 E_3|}{\Delta N} \cos(\Phi_2 - \Phi_3 - \varphi_2 - \varphi_3) \right\} \quad (2.7)
\end{aligned}$$

where

$$\begin{aligned}
& K = 4\pi^2 I_0(B_{23}), \\
& B_{23} = \frac{2R_2 R_3 |E_2 E_3|}{\Delta N} \simeq \frac{2R_2 R_3 |E_2 E_3|}{N} \quad (\text{for large } N), \quad (2.8)
\end{aligned}$$

and I is the modified Bessel function. Thus, for fixed $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ satisfying (2.1), (2.7) is the conditional joint probability distribution of the pair of phases $\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}_1 + \mathbf{k}}$, given that the primitive random variable \mathbf{k} is uniformly distributed over that region of reciprocal space for which $|E_{-\mathbf{h}_3 + \mathbf{k}}|$, $|E_{\mathbf{k}}|$ and $|E_{\mathbf{h}_1 + \mathbf{k}}|$ have the specified values R_1, R_2 and R_3 respectively, provided, of course, that (2.6) holds.

3. For fixed \mathbf{h}_1 and \mathbf{h}_3 , the conditional distribution, expectation value and variance of $\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3)$, given $|E_{-\mathbf{h}_3 + \mathbf{k}}|$, $|E_{\mathbf{k}}|$, and $|E_{\mathbf{h}_1 + \mathbf{k}}|$

3.1. The conditional distribution

Denote by $P(x | B_{23})$ the conditional probability distribution of $\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3)$, given $|E_{-\mathbf{h}_3 + \mathbf{k}}|$, $|E_{\mathbf{k}}|$, $|E_{\mathbf{h}_1 + \mathbf{k}}|$, under the assumption that (2.6) holds. Then, in view of (2.7), $P(x | B_{23})$ depends only on the parameter B_{23} and by standard techniques (Hauptman, 1972a, p. 146), is found to be

$$P(x | B_{23}) \simeq \frac{\exp(-B_{23}x)}{\pi I_0(B_{23}) \sqrt{1-x^2}}. \quad (3.1)$$

Thus, for fixed $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ satisfying (2.1), (3.1) is the conditional probability distribution of the cosine invariant, $\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3)$, given that the primitive random variable \mathbf{k} is uniformly distributed over that region of reciprocal space for which $|E_{-\mathbf{h}_3 + \mathbf{k}}|$, $|E_{\mathbf{k}}|$, and $|E_{\mathbf{h}_1 + \mathbf{k}}|$ have the specified values R_1, R_2 , and R_3 respectively, provided that (2.6) holds.

3.2. The conditional expectation value

Denote by $\varepsilon\{\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3) | B_{23}\}$ the conditional expectation of $\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3)$, given B_{23} , again under the assumption (2.6). One then readily finds (*cf.* Hauptman, 1972a, p. 155)

$$\varepsilon\{\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3) | B_{23}\} \simeq - \frac{I_1(B_{23})}{I_0(B_{23})}. \quad (3.2)$$

Under the assumption (2.6) then, the conditional expected value of the cosine invariant (3.2) is always negative.

3.3. The conditional variance

The conditional variance of $\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3)$, given B_{23} , is also found in the standard way (Hauptman, 1972a, p. 156) and is given by

$$\begin{aligned}
& \text{Var}\{\cos(\varphi_{\mathbf{k}} - \varphi_{\mathbf{h}_1 + \mathbf{k}} - \varphi_2 - \varphi_3) | B_{23}\} \\
& \simeq 1 - \frac{I_1(B_{23})}{B_{23} I_0(B_{23})} - \frac{I_1^2(B_{23})}{I_0^2(B_{23})}. \quad (3.3)
\end{aligned}$$

The conditional distribution, expectation values, and variances have been tabulated (Hauptman, 1972a, pp. 148–150).

4. Estimation of certain cosine invariants, $\cos(\varphi_m + \varphi_n + \varphi_p + \varphi_q)$, by suitable sampling of reciprocal space

Fix the reciprocal vectors $\mathbf{m}, \mathbf{n}, \mathbf{p}, \mathbf{q}$ subject to

$$\mathbf{m} + \mathbf{n} + \mathbf{p} + \mathbf{q} = 0. \quad (4.1)$$

Note that (4.1) implies $|E_{\mathbf{m}+\mathbf{n}}| = |E_{\mathbf{p}+\mathbf{q}}|$, etc. Assume that

$$\left. \begin{aligned} |E_{\mathbf{m}+\mathbf{n}}|^2 &= |E_{\mathbf{p}+\mathbf{q}}|^2 \leq 1, \\ |E_{\mathbf{m}+\mathbf{p}}|^2 &= |E_{\mathbf{n}+\mathbf{q}}|^2 \leq 1, \\ |E_{\mathbf{m}+\mathbf{q}}|^2 &= |E_{\mathbf{n}+\mathbf{p}}|^2 \leq 1, \end{aligned} \right\} \quad (4.2)$$

and

$$\left. \begin{aligned} |E_{\mathbf{m}+\mathbf{n}}| &\ll |E_{\mathbf{p}}E_{\mathbf{q}}|/N^{1/2}, & |E_{\mathbf{p}+\mathbf{q}}| &\ll |E_{\mathbf{m}}E_{\mathbf{n}}|/N^{1/2}, \\ |E_{\mathbf{m}+\mathbf{p}}| &\ll |E_{\mathbf{n}}E_{\mathbf{q}}|/N^{1/2}, & |E_{\mathbf{n}+\mathbf{q}}| &\ll |E_{\mathbf{m}}E_{\mathbf{p}}|/N^{1/2}, \\ |E_{\mathbf{m}+\mathbf{q}}| &\ll |E_{\mathbf{n}}E_{\mathbf{p}}|/N^{1/2}, & |E_{\mathbf{n}+\mathbf{p}}| &\ll |E_{\mathbf{m}}E_{\mathbf{q}}|/N^{1/2}. \end{aligned} \right\} \quad (4.3)$$

Roughly speaking then, (4.2) and (4.3) imply that each of $|E_{\mathbf{m}}|, |E_{\mathbf{n}}|, |E_{\mathbf{p}}|, |E_{\mathbf{q}}|$ is relatively large and each of $|E_{\mathbf{m}+\mathbf{n}}|, |E_{\mathbf{m}+\mathbf{p}}|, |E_{\mathbf{m}+\mathbf{q}}|$ is relatively small. In actual application, if N is at least moderately large, say $N > 100$, then (4.3) would imply that each of $|E_{\mathbf{m}+\mathbf{n}}|, |E_{\mathbf{m}+\mathbf{p}}|, |E_{\mathbf{m}+\mathbf{q}}|$ is quite small, about 0.2 or so, and (4.2) would then be automatically satisfied if (4.3) holds.

Now, define $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ by means of

$$\mathbf{h}_1 = -\mathbf{m} - \mathbf{n}, \quad \mathbf{h}_2 = -\mathbf{p}, \quad \mathbf{h}_3 = -\mathbf{q} \quad (4.4)$$

so that, in view of (4.1), (2.1) is satisfied. Choose a sample of size two from reciprocal space by means of

$$\mathbf{k} = \mathbf{m} \text{ or } \mathbf{n}. \quad (4.5)$$

Then

$$\mathbf{h}_1 + \mathbf{k} = -\mathbf{n} \text{ or } -\mathbf{m} \quad (4.6)$$

respectively,

$$-\mathbf{h}_3 + \mathbf{k} = \mathbf{m} + \mathbf{q} \text{ or } \mathbf{n} + \mathbf{q} \quad (4.7)$$

respectively and, in view of (4.2) and (4.3),

$$\left. \begin{aligned} R_1^2 &= |E_{-\mathbf{h}_3+\mathbf{k}}|^2 = |E_{\mathbf{m}+\mathbf{q}}|^2 \ll 1, \\ |E_1| &= |E_{\mathbf{m}+\mathbf{n}}| \ll |E_{\mathbf{p}}E_{\mathbf{q}}|/N^{1/2} = |E_2E_3|/N^{1/2}, \end{aligned} \right\} \quad (4.8)$$

or

$$\left. \begin{aligned} R_1^2 &= |E_{-\mathbf{h}_3+\mathbf{k}}|^2 = |E_{\mathbf{n}+\mathbf{q}}|^2 \ll 1, \\ |E_1| &= |E_{\mathbf{m}+\mathbf{n}}| \ll |E_{\mathbf{p}}E_{\mathbf{q}}|/N^{1/2} = |E_2E_3|/N^{1/2}, \end{aligned} \right\} \quad (4.9)$$

in the respective cases. In both cases then, (2.6) holds and one obtains the following estimate of the conditional expectation (3.2) by means of the sample (4.5):

$$\begin{aligned} &\frac{1}{2} \sum_{\mathbf{k}=\mathbf{m},\mathbf{n}} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}_1-\mathbf{k}} + \varphi_{-\mathbf{h}_2} + \varphi_{-\mathbf{h}_3}) \\ &= \cos(\varphi_{\mathbf{m}} + \varphi_{\mathbf{n}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}}) \simeq -\frac{I_1(B)}{I_0(B)} \end{aligned} \quad (4.10)$$

where, from (2.8), and recalling that $R_2 = |E_{\mathbf{k}}|$, $R_3 = |E_{\mathbf{n}_1+\mathbf{k}}|$,

$$B = B_{23} \simeq \frac{2|E_{\mathbf{m}}E_{\mathbf{n}}E_{\mathbf{p}}E_{\mathbf{q}}|}{N}. \quad (4.11)$$

Next, define $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ by means of

$$\mathbf{h}_1 = -\mathbf{m} - \mathbf{p}, \quad \mathbf{h}_2 = -\mathbf{n}, \quad \mathbf{h}_3 = -\mathbf{q} \quad (4.12)$$

so that, in view of (4.1), (2.1) is satisfied. Choose a sample of size two from reciprocal space by means of

$$\mathbf{k} = \mathbf{m} \text{ or } \mathbf{p}. \quad (4.13)$$

Then

$$\mathbf{h}_1 + \mathbf{k} = -\mathbf{p} \text{ or } -\mathbf{m} \quad (4.14)$$

respectively,

$$-\mathbf{h}_3 + \mathbf{k} = \mathbf{m} + \mathbf{q} \text{ or } \mathbf{p} + \mathbf{q} \quad (4.15)$$

respectively and, in view of (4.2) and (4.3),

$$\left. \begin{aligned} R_1^2 &= |E_{-\mathbf{h}_3+\mathbf{k}}|^2 = |E_{\mathbf{n}+\mathbf{q}}|^2 \ll 1, \\ |E_1| &= |E_{\mathbf{m}+\mathbf{p}}| \ll |E_{\mathbf{n}}E_{\mathbf{q}}|/N^{1/2} = |E_2E_3|/N^{1/2}, \end{aligned} \right\} \quad (4.16)$$

or

$$\left. \begin{aligned} R_1^2 &= |E_{-\mathbf{h}_3+\mathbf{k}}|^2 = |E_{\mathbf{p}+\mathbf{q}}|^2 \ll 1, \\ |E_1| &= |E_{\mathbf{m}+\mathbf{p}}| \ll |E_{\mathbf{n}}E_{\mathbf{q}}|/N^{1/2} = |E_2E_3|/N^{1/2}, \end{aligned} \right\} \quad (4.17)$$

in the respective cases. In both cases then, (2.6) holds and one obtains the following estimate of the conditional expectation (3.2) by means of the sample (4.13):

$$\begin{aligned} &\frac{1}{2} \sum_{\mathbf{k}=\mathbf{m},\mathbf{p}} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}_1-\mathbf{k}} + \varphi_{-\mathbf{h}_2} + \varphi_{-\mathbf{h}_3}) \\ &= \cos(\varphi_{\mathbf{m}} + \varphi_{\mathbf{n}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}}) \simeq -\frac{I_1(B)}{I_0(B)} \end{aligned} \quad (4.18)$$

where B is again defined by (4.11).

One continues in this way, defining $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ successively by means of

$$\mathbf{h}_1 = -\mathbf{m} - \mathbf{q}, \quad \mathbf{h}_2 = -\mathbf{n}, \quad \mathbf{h}_3 = -\mathbf{p}, \quad (4.19)$$

$$\mathbf{h}_1 = -\mathbf{n} - \mathbf{p}, \quad \mathbf{h}_2 = -\mathbf{m}, \quad \mathbf{h}_3 = -\mathbf{q}, \quad (4.20)$$

$$\mathbf{h}_1 = -\mathbf{n} - \mathbf{q}, \quad \mathbf{h}_2 = -\mathbf{m}, \quad \mathbf{h}_3 = -\mathbf{p}, \quad (4.21)$$

$$\mathbf{h}_1 = -\mathbf{p} - \mathbf{q}, \quad \mathbf{h}_2 = -\mathbf{m}, \quad \mathbf{h}_3 = -\mathbf{n}, \quad (4.22)$$

and respective samples of size two from reciprocal space by means of

$$\mathbf{k} = \mathbf{m} \text{ or } \mathbf{q}, \quad (4.23)$$

$$\mathbf{k} = \mathbf{n} \text{ or } \mathbf{p}, \quad (4.24)$$

$$\mathbf{k} = \mathbf{n} \text{ or } \mathbf{q}, \quad (4.25)$$

$$\mathbf{k} = \mathbf{p} \text{ or } \mathbf{q}. \quad (4.26)$$

As before, one is led in every case to the same sample estimate of the expectation value of the cosine invariant $\cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}_1-\mathbf{k}} + \varphi_{-\mathbf{h}_2} + \varphi_{-\mathbf{h}_3})$, given by (4.10) or (4.18) with B defined by (4.11). Averaging these six equations one obtains the first major result of this paper that, subject to (4.2) and (4.3), and based on an overall sample of size twelve from reciprocal space,

$$\cos(\varphi_{\mathbf{m}} + \varphi_{\mathbf{n}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}}) \simeq -\frac{I_1(B)}{I_0(B)} \quad (4.27)$$

in which

$$B = \frac{2|E_{\mathbf{m}}E_{\mathbf{n}}E_{\mathbf{p}}E_{\mathbf{q}}|}{N}. \quad (4.28)$$

5. Improved estimate for the cosine invariant

If N is only moderately large, it is not justified to replace Δ by unity in (2.8) as has been done in the derivation of (4.27). It is necessary instead, as reference to (2.4) shows, to use

$$\Delta \simeq 1 - \frac{1}{N} (|E_1|^2 + |E_2|^2 + |E_3|^2) \quad (5.1)$$

since the last term of (2.4) is relatively small. In this case the sampling procedure of § 4 leads to six different estimates for the expectation value of the cosine invariant, rather than just the single estimate (4.27). Averaging over these six estimates one obtains the improved formula, again based on an overall sample of size twelve from reciprocal space,

$$\cos(\varphi_m + \varphi_n + \varphi_p + \varphi_q) \simeq - \left\langle \frac{I_1(B_\mu)}{I_0(B_\mu)} \right\rangle \mu, \quad (5.2)$$

in which the average is taken over the six values of B_μ :

$$B_\mu = \frac{2|E_m E_n E_p E_q|}{\Delta_\mu N}, \quad \mu=1, \dots, 6. \quad (5.3)$$

$$\Delta_1 = 1 - \frac{1}{N} (|E_m|^2 + |E_n|^2), \quad (5.4)$$

$$\Delta_2 = 1 - \frac{1}{N} (|E_m|^2 + |E_p|^2), \quad (5.5)$$

$$\Delta_3 = 1 - \frac{1}{N} (|E_m|^2 + |E_q|^2), \quad (5.6)$$

$$\Delta_4 = 1 - \frac{1}{N} (|E_n|^2 + |E_p|^2), \quad (5.7)$$

$$\Delta_5 = 1 - \frac{1}{N} (|E_n|^2 + |E_q|^2), \quad (5.8)$$

$$\Delta_6 = 1 - \frac{1}{N} (|E_p|^2 + |E_q|^2). \quad (5.9)$$

Clearly (5.2), the second major result of this paper, reduces to (4.27) in the case that N is very large. A still further improvement over (4.27) is presumably possible if one replaces the average (5.2) by a weighted average employing the variance (3.3).

6. The applications

An idealized structure consisting of $N=29$ identical point atoms in the space group $P1$ was constructed and normalized structure factors and cosine invariants were calculated as shown in the Tables. The structure was chosen to simulate a real crystal structure; in particular the Patterson function exhibited a great deal of overlap. As before, m, n, p, q satisfy

$$m + n + p + q = 0. \quad (6.1)$$

Those quartets m, n, p, q corresponding to the 30 largest values of B were selected for which the inequalities

$$|E_{m+n}| < 0.4, \quad |E_{m+p}| < 0.4, \quad |E_{m+q}| < 0.4 \quad (6.2)$$

and

$$|E_{m+n}| + |E_{m+p}| + |E_{m+q}| < 1 \quad (6.3)$$

also held. Hence most of the inequalities (4.2) and (4.3) were satisfied and it was therefore expected that (4.27) and (5.2) would hold, at least approximately. The tenth column of Table 1 shows the true values of the 30 cosine invariants $\cos(\varphi_m + \varphi_n + \varphi_p + \varphi_q)$. Column eleven gives the estimate (4.27) and the penultimate column the improved estimate (5.2). The discrepancies between the true values and the improved estimate, shown in the last column, are due in part to the probabilistic nature of the estimates. However the estimates tend to be too large, *i.e.* not sufficiently negative, and this bias must be attributed to the omission of terms of higher order in $1/N$ in the probability distribution (2.3) or to the excessive overlap in the Patterson function which destroys the exact validity of (2.3), the theoretical basis of (4.27) and (5.2). An important problem for future research then would be to determine the form of the improved probability distribution which takes into account higher-order terms in $1/N$ and the existing overlap in the Patterson function. Nevertheless, although quantitative agreement has not yet been

Table 1. Thirty cosines predicted to be negative

Serial Number	$\frac{m}{ E_m }$	$\frac{n}{ E_n }$	$\frac{p}{ E_p }$	$\frac{q}{ E_q }$	$\frac{m+n}{ E_{m+n} }$	$\frac{m+p}{ E_{m+p} }$	$\frac{m+q}{ E_{m+q} }$	B	$\cos(\varphi_m + \varphi_n + \varphi_p + \varphi_q)$			Discrepancy actual
									True	Calc. (4.27)	Calc. (5.2)	
1	2 3 4	9 1 1	3 3 5	8 9 0	11 2 5	1 6 1	8 4 4	2.16	-0.9853	-0.722	-0.846	0.139
2	2.830	2.566	2.001	2.152	0.349	0.176	0.374					
2	6 9 0	6 1 1	8 7 1	1 1 1	15 8 1	3 0 0	1 8 1	1.96	-0.9646	-0.691	-0.820	0.145
2	6.626	2.566	2.152	1.957	0.179	0.133	0.176					
3	1 1 2	6 1 4	6 1 4	1 5 2	5 4 6	7 2 2	0 4 0	1.29	-0.7950	-0.542	-0.692	0.103
3	2.170	1.793	1.559	3.087	0.094	0.258	0.385					
4	1 8 7	3 3 2	6 4 3	4 1 2	2 5 3	7 4 4	3 7 5	1.15	-0.9998	-0.495	-0.644	0.356
4	1.039	3.034	1.604	1.771	0.090	0.235	0.241					
5	1 2 1	3 4 5	0 7 1	4 5 1	4 2 2	1 8 2	5 3 2	1.07	-0.8467	-0.469	-0.615	0.232
5	2.862	2.275	1.811	1.325	0.222	0.229	0.387					
6	1 5 2	3 3 1	6 1 1	2 4 3	2 8 3	7 8 4	3 8 3	1.03	-0.8312	-0.454	-0.607	0.224
6	3.087	1.329	1.713	1.136	0.186	0.275	0.282					
7	4 1 1	2 5 4	4 7 2	2 1 1	2 4 3	8 8 3	8 10 2	0.99	-0.9442	-0.640	-0.574	0.370
7	1.713	2.034	1.899	1.598	0.040	0.222	0.354					
8	3 3 2	4 1 2	3 6 1	1 0 2	3 7 4	6 5 1	7 5 1	0.94	-0.6571	-0.425	-0.562	0.095
8	5.034	1.871	1.764	1.367	0.210	0.268	0.356					
9	3 4 3	1 2 1	7 8 0	1 1 6	2 2 2	10 2 5	8 10 1	0.92	-0.9274	-0.417	-0.553	0.374
9	2.275	2.862	1.654	1.240	0.222	0.266	0.290					
10	4 2 4	2 2 5	1 9 0	4 3 0	1 8 9	2 3 1	3 2 7	0.88	-0.5911	-0.402	-0.501	0.081
10	1.486	1.770	2.201	2.217	0.189	0.231	0.327					
11	5 3 2	2 3 5	10 2 3	6 3 4	4 1 1	7 5 1	5 0 6	0.86	-0.6627	-0.395	-0.523	0.140
11	3.034	1.686	1.367	1.793	0.166	0.356	0.371					
12	4 1 2	0 0 3	9 1 1	3 2 2	2 1 1	5 2 1	9 1 4	0.84	-0.5480	-0.387	-0.496	0.052
12	1.871	1.555	2.566	1.645	0.166	0.287	0.379					
13	2 5 3	2 4 2	3 5 1	3 3 3	9 4 2	5 0 2	1 1 0	0.84	-0.9500	-0.387	-0.493	0.457
13	2.216	1.815	1.344	2.275	0.277	0.325	0.395					
14	5 3 2	1 1 5	8 1 1	8 3 4	2 3 3	5 2 1	9 0 6	0.81	-0.9974	-0.383	-0.508	0.489
14	5.034	1.464	1.513	1.793	0.040	0.287	0.371					
15	3 2 6	5 2 2	2 2 4	10 6 2	8 4 2	5 4 0	7 2 6	0.78	-0.9959	-0.363	-0.454	0.542
15	1.770	1.645	2.162	1.817	0.290	0.308	0.374					
16	8 3 2	7 2 8	2 2 2	1 1 4	1 5 6	1 0 1	4 2 6	0.78	+0.1856	-0.343	-0.430	0.616
16	2.107	1.429	1.758	2.000	0.169	0.180	0.371					
17	1 2 1	0 7 1	7 0 1	8 9 3	1 9 2	8 2 2	7 7 2	0.72	-0.5704	-0.338	-0.444	0.126
17	2.862	1.811	1.433	1.407	0.229	0.253	0.399					
18	7 8 2	4 3 1	10 7 1	2 2 4	8 5 3	6 1 5	6 10 2	0.72	-0.6766	-0.338	-0.430	0.247
18	2.183	1.271	1.741	2.142	0.179	0.206	0.361					
19	6 5 4	5 3 1	6 5 1	3 3 2	5 2 3	0 8 3	3 0 2	0.72	-0.7571	-0.338	-0.453	0.304
19	1.681	1.344	1.535	3.034	0.154	0.172	0.398					
20	4 1 1	2 1 1	6 8 5	6 6 5	2 0 0	2 7 6	8 5 4	0.71	-0.8703	-0.334	-0.423	0.447
20	1.731	1.361	2.259	1.936	0.133	0.299	0.310					
21	5 6 3	4 1 2	9 3 2	2 3 5	7 5 5	6 3 1	5 4 0	0.70	-0.9963	-0.330	-0.408	0.588
21	1.665	1.871	2.092	1.572	0.166	0.269	0.308					
22	2 2 4	1 4 2	7 1 3	2 2 4	5 4 6	3 3 1	6 5 1	0.68	-0.3366	-0.322	-0.408	0.071
22	1.278	1.878	1.814	2.289	0.094	0.263	0.376					
23	1 5 2	3 6 1	2 2 1	6 3 2	4 1 3	3 3 3	5 8 2	0.68	-0.7498	-0.322	-0.435	0.315
23	3.087	1.664	1.476	1.229	0.221	0.303	0.355					
24	1 1 3	3 5 1	4 1 2	1 0 3	0 4 2	2 1 1	8 4 3	0.67	-0.6159	-0.317	-0.423	0.193
24	1.741	1.207	1.588	2.951	0.225	0.283	0.392					
25	1 8 6	3 5 1	1 1 2	5 2 3	2 3 5	0 9 4	4 4 3	0.67	-0.9489	-0.317	-0.404	0.545
25	1.477	1.344	2.170	2.275	0.247	0.277	0.388					
26	6 5 1	3 3 2	6 1 2	3 3 1	9 2 3	0 6 1	3 2 0	0.67	-0.9800	-0.317	-0.425	0.555
26	1.535	1.324	1.588	1.379	0.154	0.259	0.361					
27	2 2 4	2 2 5	10 7 0	3 1 1	8 8 1	8 5 4	4 1 5	0.66	-0.6345	-0.313	-0.397	0.217
27	2.142	2.259	1.454	1.761	0.145	0.310	0.369					
28	8 7 0	0 1 1	10 2 1	2 2 4	8 8 1	2 3 1	1 0 3 2	0.65	-0.2375	-0.309	-0.386	0.148
28	2.152	1.803	1.386	1.758	0.145	0.240	0.331					
29	3 0 4	3 5 2	5 2 2	5 3 0	0 5 2	8 2 2	2 3 4	0.65	-0.5748	-0.309	-0.381	0.194
29	1.893	1.742	1.645	1.749	0.238	0.293	0.382					
30	2 2 1	7 4 5	9 1 1	0 5 5	9 4 2	7 1 2	5 3 4	0.64	-0.5041	-0.305	-0.394	0.110
30	1.476	2.014	2.566	1.222	0.090	0.284	0.352					

quite attained, the qualitative agreement between the estimates and the true cosine values is noteworthy. In particular, the average value of the magnitude of the discrepancy is 0.282 and the true value of only one of the thirty cosines in Table 1, all of which are predicted to be negative by (4.27) and (5.2), is in fact not negative. A further improvement can be realized by introducing a scaling parameter which forces the distribution of calculated cosines to be in better agreement with the observed distribution of cosine values.

Table 2 lists the true values of those cosines corresponding to the thirty largest B values. As it happens the inequalities (6.2) and (6.3) were satisfied for none of the quartets of Table 2. In strong contrast to the entries of Table 1, not a single cosine in Table 2 is negative. Thus the criteria described here serve effectively to identify the small fraction of cosines which are negative, at least for the larger values of B .

Experience has shown (e.g. Duax & Hauptman, 1972) that the ability to identify even a small number of negative cosine invariants enhances greatly the power of the direct method of phase determination. It is therefore expected that the results secured here will find early application especially if, by constructing quartets of special type, one exploits systematically the space group symmetries which may be present. In particular, some negative cosines whose values are required by the space-group symmetries to be ± 1 may well be readily identified.

The methods and results described in the present paper were secured by the author during the first two weeks (March 15–30, 1973) of the two month period, March 15–May 15, during which the author held a NATO Senior Fellowship Award under the auspices of the Consiglio Nazionale delle Ricerche. He is indebted to Drs Paolo Gallitelli and Lodovico Riva di Sanseverino for making this fellowship possible. In addition, Drs Giovanni Andreotti, Luigi Cavalca, and Mario Nardelli organized a lecture series (1–15 April, 1973) at the University of Parma during which the author had the opportunity to discuss his recent research, to lecture and to consult with, among others, Drs G. Andreotti, H. Krabbendam, D. Rogers, H. Schenk, T. Spek, and D. Viterbo. Rogers and Krabbendam, in particular, shared with the author and others some of their preliminary ideas concerned with the algebraic approach to the problem treated here from the probabilistic point of view. Subsequently Andreotti (1973) reported a preliminary calculation in the space group $P\bar{1}$ which confirmed the results obtained in this paper. The author is grateful to all of these people for the benefits he derived from these stimulating discussions. Finally, grateful acknowledgment is made to Dr David Langs who performed the calculations summarized in the Tables.

In the recent past Drs Henk Schenk and Jan de Jong, motivated by the Harker–Kasper inequalities, made a number of empirical observations and applica-

Table 2. Thirty cosines having largest values of B

Serial Number	$\frac{h}{ h }$	$\frac{k}{ k }$	$\frac{l}{ l }$	$\frac{m}{ m }$	$\frac{n}{ n }$	$\frac{p}{ p }$	$\frac{q}{ q }$	$\frac{r}{ r }$	$\frac{s}{ s }$	$\frac{t}{ t }$	B	cos (signature) True
101	5 5 2	7 1 2	2 5 4	6 7 0	4 2 4	1 6 2	9 4 2				4.540	0.9984
	3.034	2.917	2.830	2.626	2.765	1.406	2.176					
	1 5 2	0 7 2	11 6 1	10 4 1	1 2 0	10 11 1	11 1 3					
	3.087	3.053	2.492	2.292	2.217	2.074	2.074				4.260	0.9936
102	0 7 2	7 1 2	4 4 4	5 4 4	7 8 0	4 5 6	5 5 2					
	3.053	2.917	2.765	2.488	1.654	2.924	3.034				4.230	0.9971
103	5 5 2	7 1 2	1 2 1	9 4 1	10 2 0	2 5 3	6 1 1					
	3.034	2.917	2.862	2.408	1.987	2.216	1.631				4.210	0.5570
104	0 7 2	10 3 0	1 2 1	11 2 3	10 4 3	1 5 5	11 5 1					
	3.053	2.951	1.862	2.351	2.525	2.073	1.928				4.010	0.6428
105	1 5 2	3 3 2	7 1 2	5 3 2	2 5 4	6 6 0	10 5 0					
	3.087	3.034	2.917	2.092	2.142	1.517	1.987				3.940	0.8596
106	0 7 2	1 2 1	10 4 2	1 5 5	1 5 5	10 5 0						
	3.053	2.862	2.566	2.525	2.073	1.958	2.951				3.910	0.9666
107	1 5 2	3 3 2	5 2 7	3 4 5	2 5 4	6 7 5	2 1 5					
	3.087	3.034	2.643	2.275	2.142	1.109	1.643				3.890	0.6861
108	0 7 2	4 4 4	6 7 0	5 10 6	4 3 5	6 14 5	2 3 4					
	3.053	2.765	2.626	2.516	2.924	1.588	2.830				3.850	0.9949
109	0 7 2	5 3 2	7 1 2	10 9 2	3 10 0	7 6 4	10 2 0					
	3.053	3.034	2.917	2.059	1.989	1.325	1.987				3.840	0.9419
110	1 5 2	3 3 2	4 4 4	8 4 4	4 6 0	5 1 6	7 1 2					
	3.087	3.034	2.765	2.135	1.905	1.634	2.917				3.820	0.9980
111	10 3 0	4 4 4	6 7 0	8 0 4	14 7 4	4 4 0	7 0 0					
	2.951	2.765	2.626	2.547	1.921	2.242	2.830				3.770	0.9802
112	1 5 2	0 7 2	5 3 2	4 1 3	1 2 0	8 0 0	3 4 4					
	3.087	3.053	3.034	1.871	2.217	1.805	2.488				3.690	0.9977
113	1 5 2	10 3 0	9 1 1	2 4 3	11 2 2	8 6 3	1 2 1					
	3.087	2.951	2.566	2.289	2.184	1.802	2.062				3.690	0.9690
114	1 5 2	7 1 2	4 4 4	2 2 4	6 6 0	5 1 6	5 5 2					
	3.087	2.917	2.765	2.142	1.517	1.614	3.034				3.680	0.9725
115	2 2 4	2 3 4	9 1 1	8 0 4	1 1 5	10 3 0	7 2 3					
	2.862	2.830	2.566	2.547	1.464	2.951	1.220				3.650	0.9899
116	10 3 0	7 1 2	2 3 4	1 1 2	5 2 2	6 0 4	9 4 2					
	2.951	2.917	2.830	2.170	1.146	2.547	2.176				3.650	0.9024
117	1 5 2	0 7 2	10 3 0	9 1 0	1 2 0	9 8 2	10 4 2					
	3.087	3.053	2.951	1.874	2.217	0.876	2.525				3.600	0.9223
118	10 3 0	2 3 4	4 4 4	4 4 0	8 0 4	6 1 0	6 7 0					
	2.951	2.830	2.765	2.242	2.547	1.559	2.625				3.570	0.9449
119	3 3 2	2 3 4	4 4 4	9 4 2	5 0 6	7 1 2	6 7 0					
	3.034	2.830	2.765	2.176	2.152	2.917	2.626				3.570	0.9756
120	1 5 2	3 3 2	5 4 4	1 3 0	2 2 4	1 1 2	7 2 5					
	3.087	3.034	2.488	2.217	2.142	1.871	2.053				3.570	0.9291
121	7 1 2	1 2 1	4 4 4	2 5 5	6 1 1	5 5 2	5 6 5					
	2.917	2.862	2.765	2.216	1.631	3.034	1.805				3.530	0.9136
122	3 3 2	10 5 0	6 7 0	1 1 2	7 6 2	9 4 2	4 4 0					
	3.034	2.951	2.626	2.170	1.404	2.176	2.242				3.520	0.8773
123	0 7 2	10 3 0	2 5 4	8 7 6	10 4 1	2 10 8	8 0 4					
	3.053	2.951	2.830	1.999	2.525	2.516	2.547				3.520	0.9950
124	3 3 2	10 5 0	9 1 1	2 5 5	7 6 2	11 3 3	1 1 1					
	3.034	2.951	2.566	2.216	1.404	2.206	2.862				3.510	0.9002
125	0 7 2	10 3 0	1 5 1	9 6 1	10 4 2	1 9 1	9 1 1					
	3.053	2.951	2.862	1.958	2.525	1.887	2.566				3.490	0.8710
126	1 5 2	0 7 2	2 1 0	0 0 1	1 2 0	0 7 1	1 5 3					
	3.087	3.053	2.862	1.875	2.217	1.064	2.023				3.490	0.8537
127	1 5 2	0 7 2	3 4 4	2 2 4	1 2 0	2 5 6	5 5 2					
	3.087	3.053	2.488	2.142	2.217	1.491	3.034				3.470	0.9873
128	5 3 2	10 3 0	10 4 1	3 4 5	13 0 2	7 7 3	0 1 1					
	3.034	2.951	2.454	2.275	2.368	2.023	1.803				3.450	0.9961
129	1 5 2	1 2 1	9 1 1	11 2 2	2 5 5	10 4 1	10 1 0					
	3.087	2.862	2.566	2.184	2.289	2.671	2.951				3.420	0.9987

tions of some special cases of the negative cosine invariants (4.27) and (5.2), particularly in the space groups $P\bar{1}$ and $P1$ (Schenk & de Jong, 1973; Schenk, 1973). Following the lectures and discussions in Parma (April, 1973), Dr Schenk made applications of the more general invariants studied here, and these are described in the accompanying paper (Schenk, 1974).

This work was supported in part by U.S.P.H. Grant No. RR 05716.

References

- ANDREOTTI, G. (1973). Private Communication.
 DUAX, W. & HAUPTMAN, H. (1972). *Acta Cryst.* B28, 2912–2916.
 HAUPTMAN, H. (1971). *Z. Kristallogr.* 134, 28–43.
 HAUPTMAN, H. (1972a). *Crystal Structure Determination: The Role of the Cosine Seminvariants*. New York: Plenum Press.
 HAUPTMAN, H. (1972b). *Acta Cryst.* B28, 2337–2340.
 HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem I. The Centrosymmetric Crystal*. A.C.A. Monograph No. 3. Pittsburgh: Polycrystal Book Service.
 SCHENK, H. & DE JONG, J. G. H. (1973). *Acta Cryst.* A29, 31–34.
 SCHENK, H. (1973). *Acta Cryst.* A29, 480–481.
 SCHENK, H. (1974). *Acta Cryst.* A30, 477–481.
 TSOUCARIS, G. (1970). *Acta Cryst.* A26, 492–499.