that these distributions will have the potential to yield reliable estimates for the four-phase structure seminvariants $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, *i.e.* that many of the distributions will have a very small variance. However, comparison of the different rows of Table 5 reveals significant contradictions. It is therefore not recommended that the conditional probability distribution of $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, given all 24 magnitudes in its second neighborhood, be derived. Although the 24-magnitude estimate may be somewhat superior to the 12 or the three 16-magnitude estimates, it is anticipated that the improvement will be at best marginal and hardly worth the additional effort to derive or the time to calculate.

VI. Concluding remarks

The first two neighborhoods of each of the structure seminvariants, $\varphi_{\mathbf{h}}, \varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}, \varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, in the space group $P\bar{1}$ have been found. In this way those magnitudes are identified on which the value of the structure seminvariant chiefly depends, and the qualitative relation between the seminvariant and the magnitudes in the appropriate neighborhood (or subset) is derived. The task of determining the more precise relation, *i.e.* the conditional probability distribution of the structure seminvariant, given the magnitudes in the neighborhood, or an appropriate subset, remains to be completed. For the structure seminvariants $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}$ this task has been done for the first neighborhood and is described in the accompanying paper (Green & Hauptman, 1976). In view of this work it is anticipated that the remaining task, though time consuming, will not present insurmountable obstacles.

Next, there remains the problem of identifying the neighborhoods of the structure seminvariants in other space groups, in particular $P2_1$ and $P2_12_12_1$. It is anticipated that the methods described here will carry over to these space groups without essential change. Once this is done the derivation of the appropriate probability distributions in these space groups is called for. In view of our limited experience, it seems impossible to evaluate now the magnitude or difficulty of this task or the extent to which present methods, successful in the space groups $P1$ and $P\bar{1}$, will be applicable to the remaining space groups. However, some preliminary work along these lines suggests that the task will not present insuperable difficulties.

This research was supported by NSF Grant No. MPS73-04992.

References

GRANT, D., HOWELLS, R. & ROGERS, D. (1957). *Acta Cryst.* **10**, 489–497.

GREEN, E. & HAUPTMAN, H. (1976). *Acta Cryst.* A**32**, 940–944.

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

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

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

HAUPTMAN, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.

HAUPTMAN, H. & GREEN, E. (1976). *Acta Cryst.* A**32**, 45–49.

Acta Cryst. (1976). A**32**, 940

A Conditional Probability Distribution of the Structure Seminvariant $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}$ in $P\bar{1}$: Effects of Higher-Order Terms

BY EDWARD A. GREEN AND HERBERT HAUPTMAN

Medical Foundation of Buffalo, 73 *High Street, Buffalo, New York* 14203, *U.S.A.*

(*Received* 1 *December* 1975; *accepted* 12 *May* 1976)

A crystal structure in $P\bar{1}$ consisting of N identical atoms in the whole unit cell is fixed, and the four non-negative numbers $R_1, R_2, R_{12/2}, R_{1\bar{2}/2}$ are also specified. The random variables (vectors) $\mathbf{h},\mathbf{k}$ are assumed to be uniformly and independently distributed in the regions of reciprocal space defined by $|E_{\mathbf{h}}|=R_1, |E_{\mathbf{k}}|=R_2, |E_{(\mathbf{h}+\mathbf{k})/2}|=R_{12/2}, |E_{(\mathbf{h}-\mathbf{k}/2)}|=R_{1\bar{2}/2}$, (1), and $\mathbf{h}+\mathbf{k}\equiv 0 \pmod{\boldsymbol{\omega}_s}$, (2), where $\boldsymbol{\omega}_s$, the seminvariant modulus for $P\bar{1}$, is the three-dimensional vector $\boldsymbol{\omega}_s=(2,2,2)$, (3). Then the components of each of $(\mathbf{h}\pm\mathbf{k})/2$ are integers. In view of (2) and (3) the linear combination of the phases $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}$, (4), is a structure seminvariant which, as a function of the primitive random variables $\mathbf{h},\mathbf{k}$, is itself a random variable. Two approximations $Q_{\pm}, P_{\pm}$, of respective orders $1/N$, $1/N^2$, to the conditional probability distribution of φ, given the four magnitudes (1), are derived and compared. In favorable cases, *i.e.* when the variance of the distribution happens to be small, they yield a reliable estimate (0 or π) for the structure seminvariant φ.

I. Introduction

Recently secured methods in the probabilistic theory of the structure invariants (Hauptman, 1975*a*,*b*; Green & Hauptman, 1976; Hauptman & Green, 1976) are applied here to the determination of the conditional distribution of the two-phase structure seminvariant

$$\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}} \tag{1.1}$$

in the space group $P\bar{1}$ which is based on the first neighborhood of φ. As anticipated in this earlier work and described at greater length in the previous paper (Hauptman, 1976), the neighborhood concept is seen to play the central role in the probabilistic theory of the structure seminvariants, just as it did for the structure invariants.

The seminvariant modulus $\boldsymbol{\omega}_s$ in the space group $P\bar{1}$ is defined by means of

$$\boldsymbol{\omega}_s=(2,2,2)\,. \qquad (1.2)$$

Then φ (1.1) is a structure seminvariant if and only if

$$\mathbf{h}+\mathbf{k}\equiv 0 \pmod{\boldsymbol{\omega}_s}\,, \qquad (1.3)$$

i.e. if and only if the three components of $\mathbf{h}+\mathbf{k}$ are even integers. As shown in the previous paper (Hauptman, 1976), the first neighborhood of φ consists of the four magnitudes

$$|E_{\mathbf{h}}|,|E_{\mathbf{k}}|,|E_{(\mathbf{h}+\mathbf{k})/2}|,|E_{(\mathbf{h}-\mathbf{k})/2}|\,, \qquad (1.4)$$

where, in view of (1.3), $(\mathbf{h}\pm\mathbf{k})/2$ both have integer components. In this way one is led to investigate the joint probability distribution described in the following section.

Only the major results are cited in this paper and it is assumed that the reader is familiar with earlier work (Hauptman, 1975*a*, *b*; Green & Hauptman, 1976; Hauptman & Green, 1976). However, the long Appendix I contains complete details of the derivations.*

II. The joint probability distribution of the four structure factors $E_{\mathbf{h}}$, $E_{\mathbf{k}}$, $E_{(\mathbf{h}+\mathbf{k})/2}$, $E_{(\mathbf{h}-\mathbf{k})/2}$

The present paper is heavily dependent on the previous work (Hauptman, 1975*a*, *b*; Green & Hauptman, 1976; Hauptman & Green, 1976) and the assumptions made here are the same as those described earlier. Thus, a crystal structure in $P\bar{1}$, consisting of N identical atoms in the unit cell, is supposed to be fixed. The twofold Cartesian product $W\times W$ of reciprocal space W is defined to be the collection of all ordered pairs $(\mathbf{h},\mathbf{k})$ where $\mathbf{h}$ and $\mathbf{k}$ are reciprocal vectors. The ordered pair $(\mathbf{h},\mathbf{k})$ is assumed to be the primitive random variable uniformly distributed over the subset of $W\times W$ defined by (1.3). Note that $\mathbf{h}$ and $\mathbf{k}$, the components of $(\mathbf{h},\mathbf{k})$, are therefore not independently distributed in reciprocal space. Then the four structure factors

$$E_{\mathbf{h}},E_{\mathbf{k}},E_{(\mathbf{h}+\mathbf{k})/2},E_{(\mathbf{h}-\mathbf{k})/2}\,, \qquad (2.1)$$

as functions of the primitive random variables $\mathbf{h},\mathbf{k}$, are themselves random variables. Denote by

$$P=P(S_1,S_2,S_{12/2},S_{1\bar{2}/2}) \qquad (2.2)$$

the joint probability distribution of the four structure factors (2.1) which, because the space group is $P\bar{1}$, are all real. Then, following the methods described in the earlier work already referred to, it is found that

$$\begin{aligned}
P=\frac{1}{2\pi^2}\exp\Big[&-\tfrac{1}{2}(S_1^2+S_2^2+S_{12/2}^2+S_{1\bar{2}//2}^2)-\frac{13}{8N}\\
&+\frac{1}{N}(5S_1^2+5S_2^2+7S_{12/2}^2+7S_{1\bar{2}/2}^2)-\frac{1}{8N}(S_1^4+S_2^4\\
&+S_{12/2}^4+S_{1\bar{2}/2}^4+4S_1^2S_{12/2}^2+4S_1^2S_{1\bar{2}/2}^2+4S_2^2S_{12/2}^2\\
&+4S_2^2S_{1\bar{2}/2}^2+8S_{12/2}^2S_{1\bar{2}/2}^2)+\frac{S_1S_{12/2}S_{1\bar{2}/2}}{N^{1/2}}\\
&+\frac{S_2S_{12/2}S_{1\bar{2}/2}}{N^{1/2}}-\frac{S_1S_2S_{12/2}^2}{2N}-\frac{S_1S_2S_{1\bar{2}/2}^2}{2N}\\
&+\frac{S_1^3S_{12/2}S_{1\bar{2}/2}}{N^{3/2}}+\frac{S_2^3S_{12/2}S_{1\bar{2}/2}}{N^{3/2}}+\frac{3S_1S_{12/2}^3S_{1\bar{2}/2}}{2N^{3/2}}\\
&+\frac{3S_2S_{12/2}^3S_{1\bar{2}/2}}{2N^{3/2}}+\frac{3S_1S_{12/2}S_{1\bar{2}/2}^3}{2N^{3/2}}+\frac{3S_2S_{12/2}S_{1\bar{2}/2}^3}{2N^{3/2}}\\
&-\frac{5S_1S_{12/2}S_{1\bar{2}/2}}{N^{3/2}}-\frac{5S_2S_{12/2}S_{1\bar{2}/2}}{N^{3/2}}+\frac{S_1^2S_2S_{12/2}S_{1\bar{2}/2}}{N^{3/2}}\\
&+\frac{S_1S_2^2S_{12/2}S_{1\bar{2}/2}}{N^{3/2}}-\frac{S_1^3S_2S_{12/2}^2}{2N^2}-\frac{S_1S_2^3S_{12/2}^2}{2N^2}\\
&-\frac{S_1^3S_2S_{1\bar{2}/2}^2}{2N^2}-\frac{S_1S_2^3S_{1\bar{2}/2}^2}{2N^2}-\frac{5S_1S_2S_{12/2}^2S_{1\bar{2}/2}^2}{N^2}\\
&-\frac{5S_1S_2S_{12/2}^4}{6N^2}-\frac{5S_1S_2S_{1\bar{2}/2}^4}{6N^2}+\frac{11S_1S_2S_{12/2}^2}{4N^2}\\
&+\frac{11S_1S_2S_{1\bar{2}/2}^2}{4N^2}+\frac{S_1S_2}{16N^2}\Big]\left[1+\mathrm{O}\left(\frac{1}{N^2}\right)\right] \qquad (2.3)
\end{aligned}$$

where $\mathrm{O}(1/N^2)$ denotes terms of order $1/N^2$ or higher in which terms of order $1/N^2$ contain only even powers of the S's. Observe that the S variables of (2.3) range over all real values from $-\infty$ to $+\infty$. The distribution (2.3) leads directly to the joint conditional probability distribution of the pair $\varphi_{\mathbf{h}},\varphi_{\mathbf{k}}$, given the magnitudes of the four structure factors (2.1), as is shown next.

III. The joint conditional probability distribution of the pair of phases $\varphi_{\mathbf{h}}$, $\varphi_{\mathbf{k}}$, given the four magnitudes $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$, $|E_{(\mathbf{h}+\mathbf{k})/2}|$, $|E_{\mathbf{h}-\mathbf{k})/2}|$

Suppose again that a crystal structure consisting of N identical atoms per unit cell in the space group $P\bar{1}$ is fixed and the four non-negative numbers $R_1,R_2,R_{12/2}$, $R_{1\bar{2}/2}$ are also specified. Assume now that the ordered pair $(\mathbf{h},\mathbf{k})$ of reciprocal vectors is a random variable which is uniformly distributed over the subset of the Cartesian product $W\times W$ defined by

$$|E_{\mathbf{h}}|=R_1,|E_{\mathbf{k}}|=R_2,|E_{(\mathbf{h}+\mathbf{k})/2}|=R_{12/2},|E_{(\mathbf{h}-\mathbf{k})/2}|=R_{1\bar{2}/2} \qquad (3.1)$$

and (1.3). In view of (3.1) and (1.3), the random variables $\mathbf{h},\mathbf{k}$, the components of the ordered pair $(\mathbf{h},\mathbf{k})$,

* Appendix I has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 31891 (41 pp., 1 microfiche). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 13 White Friars, Chester CH1 1NZ, England; or from the authors as a technical report of the Medical Foundation of Buffalo.

are not independently distributed in reciprocal space. Then $\varphi_{\mathbf{h}}$ and $\varphi_{\mathbf{k}}$, the phases of the normalized structure factors $E_{\mathbf{h}}$ and $E_{\mathbf{k}}$, as functions of the primitive random variables $\mathbf{h}$ and $\mathbf{k}$, are themselves random variables. Denote by $P(\Phi_1,\Phi_2|R_1,R_2,R_{12/2},R_{1\bar{2}/2})$ the joint conditional probability distribution of the two phases, $\varphi_{\mathbf{h}},\varphi_{\mathbf{k}}$, given (3.1) and (1.3). Then $P(\Phi_1,\Phi_2|R_1,R_2,R_{12/2},R_{1\bar{2}/2})$ is found from (2.3) by fixing the magnitudes of $S_1,S_2,S_{12/2},S_{1\bar{2}/2}$ in accordance with the scheme

$$|S_1|=R_1,|S_2|=R_2,|S_{12/2}|=R_{12/2},|S_{1\bar{2}/2}|=R_{1\bar{2}/2}\,, \quad (3.2)$$

i.e.

$$S_1=R_1\cos\Phi_1,\ S_2=R_2\cos\Phi_2,$$
$$S_{12/2}=R_{12/2}\cos\Phi_{12/2},\ S_{1\bar{2}/2}=R_{1\bar{2}/2}\cos\Phi_{1\bar{2}/2}\,, \quad (3.3)$$

where $\Phi_{12/2}$ and $\Phi_{1\bar{2}/2}$ are the variables associated with the phases $\varphi_{(\mathbf{h}+\mathbf{k})/2}$ and $\varphi_{(\mathbf{h}-\mathbf{k})/2}$ respectively, summing with respect to $S_{12/2},S_{1\bar{2}/2}$ over their two possible signs (+ and −) or, equivalently, summing with respect to $\Phi_{12/2},\Phi_{1\bar{2}/2}$ over their two possible values (0 and π), and multiplying the result by a suitable normalizing factor. Carrying out these summations one finally obtains, correct up to and including terms of order $1/N^2$ [since $O(1/N^2)$ of (2.3) consists of all terms of order $1/N^2$ or higher in which the terms of order $1/N^2$ contain only even powers of S],

$$\begin{aligned}P(\Phi_1,\Phi_2|R_1,R_2,R_{12/2},R_{1\bar{2}/2})\\ \simeq \frac{1}{K}\exp\Big\{-\frac{R_1R_2\cos(\Phi_1+\Phi_2)}{N}\Big[\Big(\tfrac{1}{2}-\frac{11}{4N}\Big)\\ \times(R_{12/2}^2+R_{1\bar{2}/2}^2)+\frac{1}{2N}(R_1^2+R_2^2)(R_{12/2}^2+R_{1\bar{2}/2}^2)\\ +\frac{5}{6N}(R_{12/2}^4+6R_{12/2}^2R_{1\bar{2}/2}^2+R_{1\bar{2}/2}^4)+\frac{1}{16N}\Big]\Big\}\\ \times\cosh\Big\{\frac{R_{12/2}R_{1\bar{2}/2}}{N^{1/2}}(R_1\cos\Phi_1+R_2\cos\Phi_2)\\ \times\Big[1+\frac{1}{N}(R_1^2+R_2^2)+\frac{3}{2N}(R_{12/2}^2+R_{1\bar{2}/2}^2)-\frac{5}{N}\Big]\Big\}\end{aligned} \quad (3.4)$$

where K is a suitable normalizing constant, independent of Φ_1 and Φ_2. Although K is readily found by summing (3.4) over the four possible values of Φ_1,Φ_2 (each takes on the two values $0,\pi$) and setting the result equal to unity, the value of this normalizing parameter is not needed for the present purpose and is therefore not derived explicitly. Although (3.4) depends on the values of Φ_1 and Φ_2, it is readily confirmed that it actually is a function of $\Phi_1+\Phi_2$ only, since $|\cos\Phi_1|=|\cos\Phi_2|=1$ and the hyperbolic cosine is an even function of its argument. Hence (3.4) leads directly to the conditional distribution, given (3.1) and (1.3), of the sum $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}$, as is shown next.

IV. The conditional probability distribution of the structure seminvariant $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}$, given the four magnitudes $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$, $|E_{(\mathbf{h}+\mathbf{k})/2}|$, $|E_{(\mathbf{h}-\mathbf{k})/2}|$

Under the same hypotheses as in §3, the structure seminvariant

$$\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}} \quad (4{\cdot}1)$$

is a random variable whose conditional probability distribution, given (3.1), $P(\Phi|R_1,R_2,R_{12/2},R_{1\bar{2}/2})$, is readily found from (3.4). Thus, correct up to and including terms of order $1/N^2$, the major result of this paper is given by

$$\begin{aligned}P(\Phi|R_1,R_2,R_{12/2},R_{1\bar{2}/2})\\ \simeq\frac{1}{L}\exp(-C\cos\Phi)\cosh[D(R_1+R_2\cos\Phi)]\end{aligned} \quad (4.2)$$

where

$$\begin{aligned}C=\frac{R_1R_2}{2N}\Big[\Big(1-\frac{11}{2N}\Big)(R_{12/2}^2+R_{1\bar{2}/2}^2)\\ +\frac{1}{N}(R_1^2++R_2^2)(R_{12/2}^2+R_{1\bar{2}/2}^2)\\ +\frac{5}{3N}(R_{12/2}^4+6R_{12/2}^2R_{1\bar{2}/2}^2+R_{1\bar{2}/2}^4)+\frac{1}{8N}\Big]\,,\end{aligned} \quad (4.3)$$

$$\begin{aligned}D=\frac{R_{12/2}R_{1\bar{2}/2}}{N^{1/2}}\Big[1+\frac{1}{N}(R_1^2+R_2^2)\\ +\frac{3}{2N}(R_{12/2}^2+R_{1\bar{2}/2}^2)-\frac{5}{N}\Big]\,,\end{aligned} \quad (4.4)$$

and

$$\begin{aligned}L=\exp(-C)\cosh[D(R_1+R_2)]\\ +\exp(C)\cosh[D(R_1-R_2)]\,.\end{aligned} \quad (4.5)$$

If one denotes by $P_+(P_-)$ the conditional probability, given (3.1), that

$$\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}=0\ (\pi)\,, \quad (4.6)$$

or that

$$\cos\varphi=+1\ (-1)\,, \quad (4.7)$$

or that

$$E_{\mathbf{h}}E_{\mathbf{k}}\ \text{be positive (negative)}\,, \quad (4.8)$$

then (4.2) is replaced by the more suggestive

$$P_\pm\simeq\frac{1}{L}\exp(\mp C)\cosh[D(R_1\pm R_2)] \quad (4.9)$$

where C, D and L are given by (4.3)–(4.5) and the upper (lower) signs go together.

If one retains terms out to $1/N$ only, then (4.9) reduces to

$$\begin{aligned}Q_\pm\simeq\frac{1}{M}\exp\Big[\mp\frac{R_1R_2(R_{12/2}^2+R_{1\bar{2}/2}^2)}{2N}\Big]\\ \times\cosh\Big[\frac{R_{12/2}R_{1\bar{2}/2}(R_1\pm R_2)}{N^{1/2}}\Big]\,,\end{aligned} \quad (4.10)$$

where

$$M = \exp\left[-\frac{R_1R_2(R^2_{12/2}+R^2_{1\bar{2}/2})}{2N}\right] \times\cosh\left[\frac{R_{12/2}R_{1\bar{2}/2}(R_1+R_2)}{N^{1/2}}\right] + \exp\left[+\frac{R_1R_2(R^2_{12/2}+R^2_{1\bar{2}/2})}{2N}\right] \times\cosh\left[\frac{R_{12/2}R_{1\bar{2}/2}(R_2-R_1)}{N^{1/2}}\right], \quad (4.11)$$

which should be compared with (3.3) of an earlier paper (Hauptman & Green, 1976). Equation (4.10) shows that if $R_{12/2}$ and $R_{1\bar{2}/2}$ are both large then $Q_+ > \frac{1}{2}$ and φ is probably equal to 0. If, on the other hand, one of $R_{12/2}, R_{1\bar{2}/2}$ is large and the other small then $Q_+ < \frac{1}{2}$ as shown next.

IV. 1. *The case that* $R_{12/2}R_{1\bar{2}/2} \simeq 0$

In the case that

$$|E_{(\mathbf{h}+\mathbf{k})/2}E_{(\mathbf{h}-\mathbf{k})/2}| = R_{12/2}R_{1\bar{2}/2} \simeq 0\,, \quad (4.12)$$

then (4.10) and (4.11) reduce to

$$Q_\pm \simeq \frac{1}{M}\exp\left[\mp\frac{R_1R_2(R^2_{12/2}+R^2_{1\bar{2}/2})}{2N}\right], \quad (4.13)$$

$$M = 2\cosh\left[\frac{R_1R_2(R^2_{12/2}+R^2_{1\bar{2}/2})}{2N}\right] \quad (4.14)$$

so that, in this special case, $Q_+ < \frac{1}{2}$, φ is probably equal to π, and the larger the value of $1/2NR_1R_2(R^2_{12/2}+R^2_{1\bar{2}/2})$ the more likely it is that $\varphi = \pi$. Thus the results obtained here bear out the qualitative prediction made in the previous paper (Hauptman, 1976).

V. The applications

A crystal structure in $P\bar{1}$ consisting of $N=90$ identical atoms in the unit cell was constructed [the same as in a previous paper (Hauptman & Green, 1976)] and 6701 normalized structure factors $E_\mathbf{h}$ calculated. Using the 299 $|E|$'s greater than 2 (*i.e.* $|E_\mathbf{h}|>2, |E_\mathbf{k}|>2$), the 237 pairs $\mathbf{h},\mathbf{k}$ satisfying (1.3) and having the largest products $|E_\mathbf{h}E_\mathbf{k}|$ were found. The 237 values of each of P_+, Q_+, equations (4.9) and (4.10) respectively, and the standard deviation σ of (4.9),

$$\sigma = 2\sqrt{(P_+P_-)}\,, \quad (5.1)$$

Table 1. 50 *values of* $P_+(GHR), Q_+, P_+$ *for a structure in* $P\bar{1}$ *with* $N=90$ *atoms in the unit cell, arranged in ascending order of the standard deviation* (5.1) (*SIG*)

	Indices of Main Terms		Observed Magnitudes, $\|E\|$								
Number	$\vec{h}$	$\vec{k}$	$\vec{h}$	$\vec{k}$	$\frac{1}{2}(\vec{h}+\vec{k})$	$\frac{1}{2}(\vec{h}-\vec{k})$	P_+(GHR)	Q_+	P_+	COS(T)	SIG
1	5 -16 3	11 -4 -3	3.00	2.24	3.68	2.95	0.993	0.964	0.986	0.999	0.227
2	7 -10 4	5 6 -10	2.49	2.44	3.00	3.29	0.988	0.956	0.983	0.999	0.251
3	2 7 -8	4 -9 6	4.17	3.06	2.67	3.29	0.996	0.952	0.979	1.000	0.281
4	3 -4 -10	9 0 4	2.57	2.36	3.00	3.10	0.984	0.946	0.977	0.999	0.295
5	4 -6 8	10 -6 6	2.81	2.08	3.36	2.99	0.985	0.947	0.976	0.999	0.300
6	2 -11 0	4 13 0	3.46	2.40	2.61	3.66	0.991	0.945	0.975	0.999	0.311
7	2 8 -8	4 -8 6	2.38	2.02	2.99	3.29	0.977	0.939	0.973	0.999	0.321
8	13 -9 -2	7 -7 0	3.15	2.10	3.80	2.67	0.987	0.942	0.973	-1.000	0.323
9	13 -1 1	1 3 7	2.34	2.06	3.12	3.00	0.973	0.931	0.968	-1.000	0.350
10	5 -11 0	11 -9 0	2.77	2.00	3.68	2.61	0.979	0.931	0.967	1.000	0.356
11	6 -3 -7	6 5 7	2.57	2.49	3.44	2.45	0.978	0.926	0.966	1.000	0.360
12	1 12 0	5 -10 0	3.66	2.28	2.61	3.46	0.986	0.926	0.959	0.999	0.392
13	6 -6 7	10 -4 7	3.57	2.49	3.06	2.58	0.981	0.917	0.955	1.000	0.410
14	3 -3 -8	9 -3 -10	2.91	2.15	2.77	2.99	0.971	0.914	0.953	1.000	0.420
15	2 7 -9	4 -9 5	2.79	2.14	2.53	3.29	0.969	0.911	0.952	1.000	0.423
16	3 4 8	3 -4 -10	2.73	2.57	2.99	2.51	0.970	0.908	0.952	0.999	0.424
17	7 -6 7	5 ´10 -7	3.36	2.94	2.06	3.29	0.976	0.890	0.944	0.999	0.458
18	1 11 0	11 -9 0	2.76	2.00	3.44	2.28	0.955	0.882	0.931	1.000	0.506
19	8 -10 0	2 -12 0	3.68	2.33	2.77	2.61	0.968	0.886	0.929	1.000	0.511
20	4 -9 6	10 -11 2	3.06	2.77	2.49	2.53	0.959	0.877	0.929	1.000	0.512
21	1 -10 0	5 10 -2	2.20	2.15	2.99	2.51	0.940	0.875	0.925	1.000	0.525
22	2 7 -8	10 -11 2	4.17	2.77	3.00	2.14	0.973	0.866	0.918	1.000	0.546
23	5 -6 7	1 6 -9	3.02	2.18	2.99	2.30	0.948	0.866	0.915	1.000	0.556
24	2 10 7	4 -6 7	2.46	2.00	3.10	2.35	0.935	0.863	0.914	1.000	0.559
25	5 0 6	9 0 4	3.04	2.36	4.25	1.73	0.966	0.843	0.911	0.999	0.568
26	5 -11 0	1 11 0	2.77	2.76	1.79	3.46	0.948	0.836	0.901	1.000	0.595
27	4 14 0	2 -12 0	2.69	2.33	2.61	2.40	0.930	0.848	0.900	1.000	0.599
28	4 2 7	0 4 7	2.62	2.45	2.41	2.55	0.929	0.846	0.899	-1.000	0.601
29	5 -11 0	1 13 0	2.77	2.40	2.61	2.33	0.930	0.844	0.897	1.000	0.606
30	3 0 -1	9 -2 -5	2.99	2.67	2.24	2.53	0.934	0.839	0.895	1.000	0.611
31	2 3 7	2 -5 -7	2.41	2.18	2.55	2.45	0.910	0.830	0.882	-1.000	0.643
32	2 14 3	6 14 -1	2.45	2.31	2.35	2.55	0.909	0.825	0.878	1.000	0.653
33	4 11 -1	10 9 -5	3.02	2.44	2.19	2.53	0.919	0.820	0.876	1.000	0.658
34	7 -6 7	5 -6 7	3.36	3.02	3.57	0.17	0.519	0.203	0.132	-1.000	0.678
35	1 -14 1	3 -8 -1	2.28	2.05	3.46	1.85	0.901	0.804	0.867	1.000	0.678
36	2 -10 -3	8 8 -5	2.20	2.14	2.39	2.56	0.891	0.812	0.864	1.000	0.683
37	3 1 0	9 1 0	2.61	2.04	3.44	1.79	0.905	0.797	0.859	1.000	0.694
38	8 -10 0	4 14 0	3.68	2.69	2.06	2.33	0.917	0.793	0.855	1.000	0.703
39	1 9 -7	11 -5 7	3.54	2.39	2.06	2.51	0.912	0.795	0.852	0.999	0.709
40	0 4 9	6 2 5	2.51	2.06	2.26	2.53	0.883	0.797	0.848	0.999	0.716
41	5 -3 8	3 -7 6	2.99	2.08	4.18	1.57	0.932	0.777	0.847	-1.000	0.719
42	2 -11 0	2 -13 -2	3.46	2.35	3.70	0.02	0.500	0.225	0.153	-1.000	0.720
43	6 1 0	8 -15 0	3.44	2.44	2.10	2.35	0.902	0.786	0.844	0.999	0.725
44	1 5 -10	3 9 -6	2.54	2.42	4.17	0.07	0.502	0.235	0.162	-1.000	0.737
45	1 13 0	11 -9 0	2.40	2.00	2.06	2.77	0.872	0.784	0.837	1.000	0.738
46	6 -15 1	0 13 -5	2.56	2.05	2.53	2.18	0.875	0.785	0.836	1.000	0.739
47	7 -11 0	11 -9 0	2.07	2.00	2.25	2.58	0.859	0.780	0.830	1.000	0.749
48	4 14 0	8 12 0	2.69	2.08	2.04	2.55	0.865	0.769	0.821	1.000	0.766
49	2 6 7	8 -10 5	3.33	2.09	2.45	2.05	0.878	0.766	0.818	1.000	0.770
50	10 -5 -1	8 -15 1	3.26	2.56	3.68	0.24	0.533	0.245	0.181	-1.000	0.771

were calculated, and the first 50 of these, arranged in ascending order of the standard deviation (SIG), are shown in Table 1. The column headed COS (T) lists the true values of $\cos\varphi$. Each entry in the column headed P_+ (GHR) is the probability that $\cos\varphi=1$ as obtained by Grant, Howells & Rogers (1957) who employed the conditional probability distributions of the two three-phase structure invariants

$$\varphi_{\mathbf{h}}+\varphi_{-(\mathbf{h}+\mathbf{k})/2}+\varphi_{-(\mathbf{h}-\mathbf{k})/2} \tag{5.2}$$

$$\varphi_{\mathbf{k}}+\varphi_{-(\mathbf{h}+\mathbf{k})/2}+\varphi_{(\mathbf{h}-\mathbf{k})/2} \tag{5.3}$$

and, assuming independence, derived their P_+ (GHR) from these,

$$P_{\pm}\text{ (GHR)}=\tfrac{1}{2}\left[1\pm\tanh\left(\frac{R_1R_{12/2}R_{1\bar{2}/2}}{N^{1/2}}\right)\times\tanh\left(\frac{R_2R_{12/2}R_{1\bar{2}/2}}{N^{1/2}}\right)\right]. \tag{5.4}$$

Inspection of Table 1 shows first that, when σ is small,

$$P_+\text{ (GHR)}>P_+>Q_+>0\cdot5 \tag{5.5}$$

or

$$P_+\text{ (GHR)}>0.5>Q_+>P_+\,. \tag{5.6}$$

Comparison with the true cosine values shows further that P_+(GHR) has a small positive bias, *i.e.* P_+ (GHR) is too large. In fact, since

$$P_+\text{ (GHR)}\geq\tfrac{1}{2}\,, \tag{5.7}$$

this distribution is not able to identify the negative cosines. The distribution Q_+, on the other hand, not only reliably identifies many cosines, which are in fact positive, but its negative indications are quite reliable and the positive bias is virtually eliminated. Comparison of P_+ with COS (T) shows P_+ to be essentially unbiased, and its negative indications are most reliable. The comparison of P_+ (GHR) with Q_+ and P_+ clearly shows how risky it may be to assume independence when not justified and the improvement which results from the ability to take into account all correlations among the structure factors.

Comparison of the present Table 1 with Table 1 of an earlier paper (Hauptman & Green, 1976), shows that the second (seven-magnitude) neighborhood of the four-phase structure invariant is more useful in the applications than the first (four-magnitude) neighborhood of the two-phase structure seminvariant described here. This result is not surprising since one naturally anticipates that the seven-magnitude neighborhood contains more information than the four-magnitude neighborhood in accordance with the principle of nested neighborhoods recently formulated (Hauptman, 1975). Nevertheless, the present Table 1 does yield reliable estimates for a few of the two-phase structure seminvariants (particularly those which are indicated to be π) and these may well prove useful in supplementing the more reliably determined four-phase invariants. For very complex structures, however, it will be necessary to employ the higher neighborhoods of both the structure invariants and seminvariants, *e.g.* the 13 or 21-magnitude neighborhood of the four-phase invariant or the 5, 6 or 12-magnitude neighborhood of the two-phase seminvariant, *etc.*

VI. Concluding remarks

The conditional probability distribution of the two-phase structure seminvariant $\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}$ in $P\bar{1}$, given the four magnitudes $|E_{\mathbf{h}}|,|E_{\mathbf{k}}|,|E_{(\mathbf{h}+\mathbf{k})/2}|,|E_{(\mathbf{h}-\mathbf{k})/2}|$, has been found. This derivation shows that the methods initiated recently (Hauptman 1975*a*, *b*) may be carried over without essential change to structure invariants and seminvariants in general. As anticipated in the earlier work, the neighborhood concept plays an essential role. It is necessary now to derive the distributions appropriate to the higher neighborhoods of the structure invariants and seminvariants in the various space groups if the most effective application to very complex crystal structures is to be made.

This research was supported by NSF Grant No. MPS73-04992 and NIH Grant No. GM-19684.

References

Grant, D., Howells, R. & Rogers, D. (1957). *Acta Cryst.* **10**, 489–497.

Green, E. A. & Hauptman, H. (1976). *Acta Cryst.* A**32**, 43–45.

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

Hauptman, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.

Hauptman, H. (1976). *Acta Cryst.* A**32**, 934–940.

Hauptman, H. & Green, E. A. (1976). *Acta Cryst.* A**32**, 45–49.