Acta Cryst. (1976). A**32**, 43

A Joint Probability Distribution of Seven Structure Factors in $P\bar{1}$*

BY EDWARD A. GREEN AND HERBERT HAUPTMAN

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

(*Received* 16 *June* 1975; *accepted* 28 *July* 1975)

For a fixed crystal structure in $P\bar{1}$, it is assumed that $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$ are random variables (vectors) uniformly and independently distributed over the subset of reciprocal space defined by $\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}=0$. Then the seven structure factors $E_\mathbf{h}, E_\mathbf{k}, E_\mathbf{l}, E_\mathbf{m}, E_{\mathbf{h}+\mathbf{k}}, E_{\mathbf{k}+\mathbf{l}}, E_{\mathbf{l}+\mathbf{h}}$, as functions of the primitive random variables $\mathbf{h}, \mathbf{k}, \mathbf{l}, \mathbf{m}$, are themselves random variables, and their joint probability distribution is found. This distribution plays the central role in the theory and estimation of the four-phase structure invariants $\varphi_\mathbf{h}+\varphi_\mathbf{k}+\varphi_\mathbf{l}+\varphi_\mathbf{m}$.

1. Introduction

The work initiated in two recent papers (Hauptman, 1975*a*, *b*) for the space group $P1$ is continued here for $P\bar{1}$. The similarities between the present papers and this earlier work, with respect to both methods and results, are noteworthy, and these papers are written in such a way as to stress this parallel. For these reasons the present papers are greatly abbreviated and should be read in conjunction with the earlier ones. As shown by recent applications (DeTitta, Edmonds, Langs & Hauptman, 1974, 1975; Einspahr, Gartland, Freeman & Schenk, 1974), the results obtained are important for both the theory and practice of direct methods. However, of far greater significance is the nature of the probabilistic background, in particular the choice of the primitive and dependent random variables. In fact, the idea to select certain sets of reciprocal vectors as primitive random variables distributed over well defined subsets of reciprocal space represents a significant departure from all earlier theories. This idea, coupled with the newly acquired ability to derive the associated conditional probability distributions of the structure invariants, leads to probabilistic estimates for the structure invariants in terms of arbitrary sets of magnitudes. The unexpected results obtained in turn form the basis for the important concept 'neighborhood of a structure invariant' and the 'principle of nested neighborhoods' described in the earlier work. The available evidence already shows that these concepts and their generalizations will surely play a fundamental role in the further development of the theory of the structure invariants and seminvariants and of the techniques for estimating their values. It is already known that the ability to estimate reliably the values of the cosine invariants and seminvariants leads directly and unambiguously to a technique for evaluating the individual phases by means of least-squares (Hauptman, Fisher, Hancock & Norton, 1969). Hence it is anticipated that the present work and its generalizations will find early application to the solution of unknown structures.

* Presented at the Charlottesville meeting of the ACA, March 9–13, 1975, Abstract A1.

2. Joint probability distribution of the seven structure factors $E_\mathbf{h}, E_\mathbf{k}, E_\mathbf{l}, E_\mathbf{m}, E_{\mathbf{h}+\mathbf{k}}, E_{\mathbf{k}+\mathbf{l}}, E_{\mathbf{l}+\mathbf{h}}$ in $P\bar{1}$

Suppose that a crystal structure, consisting of N identical atoms per unit-cell in the space group $P\bar{1}$, is fixed. In much the same way that the Cartesian plane may be defined as the collection of all ordered pairs (x,y) of real numbers x and y, so now the fourfold Cartesian product $W\times W\times W\times W$ of reciprocal space W with itself is defined to be the collection of all ordered quadruples $(\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m})$ of reciprocal vectors $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$. Assume now 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 the Cartesian product $W\times W\times W\times W$ for which

$$\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}=0\,, \tag{2.1}$$

where $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$ are reciprocal vectors. In view of (2.1) 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. Then the seven (real) normalized structure factors $E_\mathbf{h}, E_\mathbf{k}, E_\mathbf{l}, E_\mathbf{m}, E_{\mathbf{h}+\mathbf{k}}, E_{\mathbf{k}+\mathbf{l}}, E_{\mathbf{l}+\mathbf{h}}$, as functions of the primitive random variables $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$, are themselves random variables. Denote by

$$P=P(S_1,S_2,S_3,S_4,S_{12},S_{23},S_{31}) \tag{2.2}$$

the joint probability distribution of the seven structure factors $E_\mathbf{h}, E_\mathbf{k}, E_\mathbf{l}, E_\mathbf{m}, E_{\mathbf{h}+\mathbf{k}}, E_{\mathbf{k}+\mathbf{l}}, E_{\mathbf{l}+\mathbf{h}}$. Then, following Karle & Hauptman (1958), P is given by the sevenfold integral

$$P=\frac{1}{(2\pi)^7}\int_{\sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_{12},\sigma_{23},\sigma_{31}=-\infty}^{\infty}\exp\left\{-i\left[S_1\sigma_1+S_2\sigma_2+S_3\sigma_3+S_4\sigma_4+S_{12}\sigma_{12}+S_{23}\sigma_{23}+S_{31}\sigma_{31}\right]\right\}\times\prod_{\lambda=1}^{N/2}g_\lambda\,d\sigma_1 d\sigma_2 d\sigma_3 d\sigma_4 d\sigma_{12} d\sigma_{23} d\sigma_{31} \tag{2.3}$$

where

$$g_\lambda = g(\mathbf{r}_\lambda;\ \sigma_1,\sigma_2,\sigma_3,\sigma_4,\sigma_{12},\sigma_{23},\sigma_{31})$$
$$= \Big\langle \exp\Big\{\frac{2i}{N^{1/2}}\Big[\sigma_1 \cos 2\pi\mathbf{h}\,.\,\mathbf{r}_\lambda + \sigma_2 \cos 2\pi\mathbf{k}\,.\,\mathbf{r}_\lambda$$
$$+\sigma_3 \cos 2\pi\mathbf{l}\,.\,\mathbf{r}_\lambda + \sigma_4 \cos 2\pi(\mathbf{h}+\mathbf{k}+\mathbf{l})\,.\,\mathbf{r}_\lambda$$
$$+\sigma_{12} \cos 2\pi(\mathbf{h}+\mathbf{k})\,.\,\mathbf{r}_\lambda + \sigma_{23} \cos 2\pi(\mathbf{k}+\mathbf{l})\,.\,\mathbf{r}_\lambda$$
$$+\sigma_{31} \cos 2\pi(\mathbf{l}+\mathbf{h})\,.\,\mathbf{r}_\lambda\Big]\Big\}\Big\rangle_{\mathbf{h},\mathbf{k},\mathbf{l}}, \qquad (2.4)$$

in which $\mathbf{r}_\lambda$ is the position vector of the atom labeled λ and the average is taken over all vectors $\mathbf{h},\mathbf{k},\mathbf{l}$ in reciprocal space. The mathematical Appendixes I–III contain respectively, I. the derivation of g_λ, II. the derivation of $\prod_{\lambda=1}^{N/2} g_\lambda$, and III. a brief description of the techniques required to evaluate the sevenfold integral (2.3). Only the final formula, the chief result of the present paper, is written down here:

$$P = \frac{1}{(2\pi)^{7/2}} \exp\Big\{-\tfrac{1}{2}\Big(S_1^2+S_2^2+S_3^2+S_4^2+S_{12}^2+S_{23}^2+S_{31}^2\Big)$$
$$+\frac{1}{N^{1/2}}\Big(S_1S_2S_{12}+S_3S_4S_{12}+S_2S_3S_{23}+S_1S_4S_{23}$$
$$+S_1S_3S_{31}+S_2S_4S_{31}\Big)$$
$$-\frac{1}{N}\Big(S_1S_3S_{12}S_{23}+S_2S_4S_{12}S_{23}+S_1S_2S_{23}S_{31}$$
$$+S_3S_4S_{23}S_{31}+S_2S_3S_{31}S_{12}+S_1S_4S_{31}S_{12}+2S_1S_2S_3S_4\Big)\Big\}$$
$$\times\Big\{1+O\Big(\frac{1}{N}\Big)\Big\} \qquad (2.5)$$

where $O(1/N)$ represents terms of order $1/N$ or higher in which the terms of order $1/N$ contain only even powers of the S's. Equation (2.5) should be compared with the corresponding distribution for $P1$ [(2.5), Hauptman, 1975*a*]. Note however that in $P1$ the magnitude variables R range from 0 to ∞ whereas the present S variables range from $-\infty$ to ∞. Particularly noteworthy (in view of § 3) is the numerical coefficient, -2, of the last term in the argument of the exponential function of (2.5). It should also be stressed that the exponent in (2.5) is a quartic polynomial in the S's consisting of all the possible structure invariants which can be constructed from the seven variables $S_1, S_2, S_3, S_4, S_{12}, S_{23}, S_{31}$.

Joint probability distribution of the four structure factors $E_\mathbf{h}, E_\mathbf{k}, E_\mathbf{l}, E_\mathbf{m}$

Under the same hypotheses as in § 2 and using similar notation and methods, one readily finds

$$P(S_1,S_2,S_3,S_4)$$
$$= \frac{1}{4\pi^2} \exp\Big\{-\tfrac{1}{2}(S_1^2+S_2^2+S_3^2+S_4^2) + \frac{1}{N} S_1S_2S_3S_4\Big\}$$
$$\times\Big\{1+O\Big(\frac{1}{N}\Big)\Big\}, \qquad (3.1)$$

where again $O(1/N)$ represents terms of order $1/N$ or higher in which the terms of order $1/N$ contain only even powers of the S's. The numerical coefficient, $+1$, of the last term in the exponent of (3.1) is to be compared with the corresponding coefficient, -2, of the last term in the exponent of (2.5). It follows, as shown in the accompanying paper (Hauptman & Green, 1976), that (3.1) can lead only to the positive estimate, dependent on four magnitudes, for the cosine invariant $\cos(\varphi_\mathbf{h}+\varphi_\mathbf{k}+\varphi_\mathbf{l}+\varphi_\mathbf{m})$ but that (2.5) may lead to either estimate, ± 1, for the value of this cosine; in the latter case the estimate is dependent on the presumed known values of the seven (rather than only four) related structure-factor magnitudes.

4. Concluding remarks

The work initiated in a previous paper (Hauptman, 1975*a*) for the space group $P1$ is here carried over without essential change in order to derive joint probability distributions of seven and four structure factors in $P\bar{1}$ on the basis that four reciprocal vectors $\mathbf{h},\mathbf{k},\mathbf{l},\mathbf{m}$, subject to (2.1), are the primitive random variables. As shown in the following paper (Hauptman & Green, 1976), (2.5) leads directly to the conditional probability distribution of the structure invariant $\varphi = \varphi_\mathbf{h}+\varphi_\mathbf{k}+\varphi_\mathbf{l}+\varphi_\mathbf{m}$, give the seven magnitudes $|E_\mathbf{h}|, |E_\mathbf{k}|, |E_\mathbf{l}|, |E_\mathbf{m}|, |E_{\mathbf{h}+\mathbf{k}}|, |E_{\mathbf{k}+\mathbf{l}}|, |E_{\mathbf{l}+\mathbf{h}}|$.

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

APPENDIX I
The derivation of g_λ

The present equation (2.4) is obtainable from equation (2.4) of the earlier paper (Hauptman, 1975*a*) by replacing each θ of the latter by zero and doubling the argument of the exponential function (aside from the trivial replacement of each ϱ by σ). Hence the present g_λ is obtained from equation (II.29) of the earlier paper by replacing each θ in (II.29) by zero, doubling the argument of each Bessel Function, and replacing each ϱ by σ.

APPENDIX II
The derivation of $\prod_{\lambda=1}^{N/2} g_\lambda$

In view of Appendix I, refer to Appendix III of the earlier paper to infer that

$$\prod_{\lambda=1}^{N/2} g_\lambda = \exp\left\{-\tfrac{1}{2}(\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2+\sigma_{12}^2+\sigma_{23}^2+\sigma_{31}^2) - \frac{i}{N^{1/2}}(\sigma_1\sigma_4\sigma_{23}+\sigma_1\sigma_3\sigma_{31}+\sigma_1\sigma_2\sigma_{12}+\sigma_2\sigma_3\sigma_{23} + \sigma_2\sigma_4\sigma_{31}+\sigma_3\sigma_4\sigma_{12}) + \frac{1}{N}(\sigma_1\sigma_2\sigma_3\sigma_4+\sigma_2\sigma_3\sigma_{31}\sigma_{12}+\sigma_2\sigma_4\sigma_{12}\sigma_{23}+\sigma_3\sigma_4\sigma_{23}\sigma_{31} + \sigma_1\sigma_4\sigma_{31}\sigma_{12}+\sigma_1\sigma_3\sigma_{12}\sigma_{23}+\sigma_1\sigma_2\sigma_{23}\sigma_{31})\right\} \times\left\{1+O\left(\frac{1}{N}\right)\right\} \quad \text{(II.1)}$$

where $O(1/N)$ represents the terms of order $1/N$ or higher in which the terms of order $1/N$ contain only even powers of the σ's.

APPENDIX III

Evaluating the sevenfold integral (2.3)

III. 1. *The σ_1 integration*

Substitute for $\prod_{\lambda=1}^{N/2} g_\lambda$ from (II.I) into (2.3), combine the terms in the exponent involving σ_1, complete the square, and perform the σ_1 integration.

III. 2. *The remaining integrations*

One continues in this way, carrying out the successive integrations with respect to $\sigma_2, \sigma_3, \ldots$, until finally (2.5) is obtained.

It is instructive to compare these integrations with those of the earlier paper (Appendix IV).

References

DeTitta, G., Edmonds, J., Langs, D. & Hauptman, H. (1974). Amer. Cryst. Assoc. Summer Meeting at Penn State Univ. Abstract D3.
DeTitta, G., Edmonds, J., Langs, D. & Hauptman, H. (1975). *Acta Cryst.* A**31**, 472–479.
Einspahr, H., Gartland, G., Freeman, G. & Schenck, H. (1974). Amer. Cryst. Assoc. Summer Meeting at Penn State Univ. Abstract B6.
Hauptman, H. (1975*a*). *Acta Cryst.* A**31**, 671–679.
Hauptman, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.
Hauptman, H., Fisher, J., Hancock, H. & Norton, D. (1969). *Acta Cryst.* B**25**, 811–814.
Hauptman, H. & Green, E. A. (1976). *Acta Cryst.* A**32**, 45–49.
Karle, J. J. & Hauptman, H. (1958). *Acta Cryst.* **11**, 264–269.

Acta Cryst. (1976). A**32**, 45

Conditional Probability Distributions of the Four-Phase Structure Invariant $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$ in $P\bar{1}$*

By Herbert Hauptman and Edward A. Green
Medical Foundation of Buffalo, 73 *High Street*, *Buffalo*, *New York* 14203, *U.S.A.*

(*Received* 16 *June* 1975; *accepted* 28 *July* 1975)

A crystal structure in $P\bar{1}$ is assumed to be fixed and the seven non-negative numbers R_1, R_2, R_3, R_4, R_{12}, R_{23}, R_{31} are also given. It is assumed that **h**,**k**,**l**,**m** are random variables uniformly and independently distributed over the subsets of reciprocal space defined by

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

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

and

$$\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}=0\,. \quad (3)$$

Then the structure invariant

$$\varphi=\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}} \quad (4)$$

is a function of the primitive random variables **h**, **k**, **l**, **m**. The conditional probability distribution of φ, given (1) and (2), is obtained and compared with the conditional probability distribution of φ when only (1) is given. Some calculations are presented which show the usefulness of the distribution, given (1) and (2), in estimating the value of φ.

1. Introduction

The methods introduced in two previous papers (Hauptman 1975*a*, *b*) for $P1$ are applied here to the space group $P\bar{1}$. Again, as in the earlier work, the joint probability distribution of seven structure factors [Green & Hauptman, 1976, equation (2.5)] leads directly to the conditional probability distribution of the four-phase structure invariant $\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}+\varphi_{\mathbf{m}}$, given the seven magnitudes $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$, $|E_{\mathbf{l}}|$, $|E_{\mathbf{m}}|$, $|E_{\mathbf{h}+\mathbf{k}}|$, $|E_{\mathbf{k}+\mathbf{l}}|$, $|E_{\mathbf{l}+\mathbf{h}}|$. However, in contrast to the earlier distri-

* Presented at the Charlottesville meeting of the ACA, March 9–13, 1975, Abstract A1.