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

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 $P\bar{T}$, it is assumed that $h, k, l, m$ are random variables (vectors) uniformly and independently distributed over the subset of reciprocal space defined by $h + k + l + m = 0$. Then the seven structure factors $E_h, E_k, E_l, E_m, E_{h+k}, E_{k+l}, E_{h+l}$, as functions of the primitive random variables $h, k, l, m$, 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 $\phi_h + \phi_k + \phi_l + \phi_m$.

1. Introduction

The work initiated in two recent papers (Hauptman, 1975a, b) for the space group $P\bar{1}$ is continued here for $P\bar{T}$. 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 find early application to the solution of unknown structures.

2. Joint probability distribution of the seven structure factors $E_h, E_k, E_l, E_m, E_{h+k}, E_{k+l}, E_{h+l}$ in $P\bar{T}$

Suppose that a crystal structure, consisting of $N$ identical atoms per unit-cell in the space group $P\bar{T}$, 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 $(h, k, l, m)$ of reciprocal vectors $h, k, l, m$. Assume now that the ordered quadruple of reciprocal vectors $(h, k, l, m)$ is a random variable which is uniformly distributed over the subset of the Cartesian product $W \times W \times W \times W$ for which

$$h + k + l + m = 0, \quad (2.1)$$

where $h, k, l, m$ are reciprocal vectors. In view of (2.1) the random variables $h, k, l, m$, the components of the ordered quadruple $(h, k, l, m)$, are not independently distributed. Then the seven (real) normalized structure factors $E_h, E_k, E_l, E_m, E_{h+k}, E_{k+l}, E_{h+l}$, as functions of the primitive random variables $h, k, l, m$, are themselves random variables. Denote by

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

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

$$P = \frac{1}{(2\pi)^7} \int_{-\infty}^{\infty} \cdots \exp \left\{ -i \left[ S_1 \sigma_1 + S_2 \sigma_2 + S_3 \sigma_3 + S_4 \sigma_4 + S_{12} \sigma_{12} + S_{23} \sigma_{23} + S_{31} \sigma_{31} \right] \right\}$$

$$\times \prod_{\lambda=1}^{N/2} g_{\lambda} d\sigma_1 d\sigma_2 d\sigma_3 d\sigma_4 d\sigma_{12} d\sigma_{23} d\sigma_{31} \quad (2.3)$$
where

\[ g_\lambda = g(r_\lambda; \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_{12}, \sigma_{23}, \sigma_{31}) \]

\[ = \exp \left\{ \frac{2i}{N^{1/2}} \left[ \sigma_1 \cos 2\pi h \cdot r_\lambda + \sigma_2 \cos 2\pi k \cdot r_\lambda 
+ \sigma_3 \cos 2\pi l \cdot r_\lambda + \sigma_4 \cos 2\pi (h + k + l) \cdot r_\lambda 
+ \sigma_{12} \cos 2\pi (h + k) \cdot r_\lambda + \sigma_{23} \cos 2\pi (k + l) \cdot r_\lambda 
+ \sigma_{31} \cos 2\pi (l + h) \cdot r_\lambda \right] \right\} \]

\[ = \exp \left\{ \sum_{\lambda=1}^{\Pi} \left[ \sigma_\lambda \cos 2\pi (h + k + l) \cdot r_\lambda \right] \right\} \]

\[ = \exp \left\{ \sum_{\lambda=1}^{\Pi} \left[ \sigma_\lambda \cos 2\pi (h + k + l) \cdot r_\lambda \right] \right\} \]

in which \( r_\lambda \) is the position vector of the atom labeled \( \lambda \) and the average is taken over all vectors \( h, k, l \) in reciprocal space. The mathematical Appendixes I–III contain respectively, I. the derivation of \( g_\lambda \), II. the derivation of \( \Pi g_\lambda \), 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:

\[ P(S_1, S_2, S_3, S_4) \]

\[ = \frac{1}{4\pi^2} \exp \left\{ -\frac{1}{2} \left( S_1^2 + S_2^2 + S_3^2 + S_4^2 + S_{12} + S_{23} + S_{31} \right) \right\} \times \left\{ 1 + O\left( \frac{1}{N} \right) \right\} \]

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 \( \cos (\phi_n + \phi_k + \phi_l + \phi_m) \) but that (2.5) may lead to either estimate, \( \pm 1 \), 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, 1975a) 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 \( P1 \) on the basis that four reciprocal vectors \( h, k, l, m \), 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 \( \varphi = \varphi_n + \varphi_k + \varphi_l + \varphi_m \), give the seven magnitudes \( |E_h|, |E_k|, |E_l|, |E_m|, |E_{h+k}|, |E_{k+l}|, |E_{n+k}| \).

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

APPENDIX I

The derivation of \( g_\lambda \)

The present equation (2.4) is obtainable from equation (2.4) of the earlier paper (Hauptman, 1975a) by replacing each \( \theta \) of the latter by zero and doubling the argument of the exponential function (aside from the trivial replacement of each \( \varphi \) by \( \sigma \)). Hence the present \( g_\lambda \) is obtained from equation (II.29) of the earlier paper by replacing each \( \theta \) in (II.29) by zero, doubling the argument of each Bessel Function, and replacing each \( \varphi \) by \( \sigma \).

APPENDIX II

The derivation of \( \Pi g_\lambda \)

In view of Appendix I, refer to Appendix III of the earlier paper to infer that
\[ \prod_{\lambda=1}^{N/2} g_\lambda = \exp \left\{ -\frac{1}{2} \left( \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma_4^2 + \sigma_{12}^2 + \sigma_{23}^2 + \sigma_{34}^2 \right) \right\} \]

\[ \frac{1}{N^{1/2}} \left( \sigma_1 \sigma_3 \sigma_{23} + \sigma_1 \sigma_3 \sigma_{31} + \sigma_1 \sigma_2 \sigma_{12} + \sigma_2 \sigma_3 \sigma_{23} \right) \]

\[ + \frac{1}{N} \left( \sigma_1 \sigma_4 \sigma_{31} + \sigma_2 \sigma_4 \sigma_{32} + \sigma_3 \sigma_4 \sigma_{23} \sigma_{31} \right) \]

\[ \times \left\{ 1 + O \left( \frac{1}{N} \right) \right\} \]

(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 \( \sigma \)'s.

**APPENDIX III**

**Evaluating the sevenfold integral (2.3)**

III. 1. The \( \sigma_1 \) integration

Substitute for \( \prod_{\lambda=1}^{N/2} g_\lambda \) from (II.1) into (2.3), combine the terms in the exponent involving \( \sigma_1 \), complete the square, and perform the \( \sigma_1 \) integration.

---

**Conditional Probability Distributions of the Four-Phase Structure Invariant**

\[ \phi_h + \phi_k + \phi_l + \phi_m \text{ in } P\Gamma^* \]

**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 \( P\Gamma^* \) 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

\[ |E_n| = R_1, \quad |E_k| = R_2, \quad |E_l| = R_3, \quad |E_m| = R_4, \]

(1)

\[ |E_{h+k}| = R_{12}, \quad |E_{k+l}| = R_{23}, \quad |E_{l+h}| = R_{31}, \quad |E_{h+k+l+m}| = 0 \]

(2)

Then the structure invariant

\[ \varphi = \varphi_h + \varphi_k + \varphi_l + \varphi_m \]

(3)

is a function of the primitive random variables \( h, k, l, m \). The conditional probability distribution of \( \varphi \), given (1) and (2), is obtained and compared with the conditional probability distribution of \( \varphi \) 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 \( \varphi \).

**1. Introduction**

The methods introduced in two previous papers (Hauptman 1975a, b) for \( P1 \) are applied here to the space group \( P\Gamma^* \). 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 \( \varphi_h + \varphi_k + \varphi_l + \varphi_m \), given the seven magnitudes \( |E_n|, |E_k|, |E_l|, |E_m|, |E_{h+k}|, |E_{k+l}|, |E_{l+h}|. \) However, in contrast to the earlier distri-