Acta Cryst. (1985). A**41**, 454–456

Estimating the Three-Phase Structure Invariants *via* Their Second Neighborhoods

BY HERBERT HAUPTMAN

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(*Received* 10 *January* 1984; *accepted* 15 *April* 1985)

Abstract

The quintet extension of the three-phase structure invariant is used to define the second neighborhood of the invariant, which in turn serves to define a new kind of probability distribution. The latter leads to novel estimates of the three-phase structure invariants and the magnitudes of the normalized structure factors for the 'squared' structure.

1. Introduction

Suppose that $\mathbf{h}$, $\mathbf{k}$, $\mathbf{l}$ are reciprocal-lattice vectors satisfying

$$\mathbf{h}+\mathbf{k}+\mathbf{l}=0, \tag{1.1}$$

so that the linear combination of three phases

$$\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}} \tag{1.2}$$

is a structure invariant. The first neighborhood of the invariant (1.2) consists of the three magnitudes

$$|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}| \tag{1.3}$$

and the second of the three magnitudes (1.3) plus the additional seven

$$|E_{\mathbf{H}}|; \tag{1.4}$$

$$|E_{\mathbf{h}\pm\mathbf{H}}|, |E_{\mathbf{k}\pm\mathbf{H}}|, |E_{\mathbf{l}\pm\mathbf{H}}|, \tag{1.5}$$

where $\mathbf{H}$ is an arbitrary reciprocal-lattice vector (Hauptman, 1978). Since $\mathbf{H}$ is arbitrary, there are many second neighborhoods. Estimates of the three-phase invariant in terms of some or all of the magnitudes in the second neighborhood have been known for many years [*e.g.* Messager & Tsoucaris (1972), Hauptman (1972), Viterbo & Woolfson (1973), Giacovazzo (1976, 1977), Karle (1979, 1980)], but methods for combining the different estimates, corresponding to different vectors $\mathbf{H}$, have not been completely satisfactory because the estimates are not independent. In the present paper a different kind of estimate is obtained and a more satisfactory theoretical basis for combining the different estimates derived.

The normalized structure factor $E_{\mathbf{H}}$ is defined by

$$E_{\mathbf{H}}=|E_{\mathbf{H}}|\exp(i\varphi_{\mathbf{H}}) = \sigma_2^{-1/2}\sum_{j=1}^{N} f_j \exp(2\pi i\mathbf{H}\,.\,\mathbf{r}_j), \tag{1.6}$$

where N is the number of atoms in the unit cell, $\mathbf{r}_j$ and f_j are the position vector and zero-angle atomic scattering factor, respectively, of the atom labelled j, and

$$\sigma_n=\sum_{j=1}^{N} f_j^n,\ n=1,2,3,\ldots. \tag{1.7}$$

The normalized structure factor $G_{\mathbf{H}}$ for the 'squared' structure is defined by

$$G_{\mathbf{H}}=|G_{\mathbf{H}}|\exp(i\psi_{\mathbf{H}})=\sigma_4^{-1/2}\sum_{j=1}^{N} f_j^2\exp(2\pi i\mathbf{H}\,.\,\mathbf{r}_j). \tag{1.8}$$

2. The probabilistic background

A crystal structure in $P1$, consisting of N, not necessarily identical, atoms per unit cell is supposed to be fixed. The three reciprocal-lattice vectors $\mathbf{h}$, $\mathbf{k}$, $\mathbf{l}$ satisfying (1.1) are also specified so that the structure invariant

$$T=\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{l}} \tag{2.1}$$

is also fixed. The primitive random variable is the reciprocal-lattice vector $\mathbf{H}$, which is assumed to be uniformly distributed over a certain well-defined subset of reciprocal space (more precisely specified later). Then the magnitude of the normalized structure factor $E_{\mathbf{H}}$, as a function of the primitive random variable $\mathbf{H}$, is itself a random variable. The conditional probability distribution of $|E_{\mathbf{H}}|$, (II.4),* assuming as known the six magnitudes $|E|$, (1.5), in the second neighborhood of the invariant (1.2), and its implications, in particular the system of linear equations (3.21)-

* Appendices I-III have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 42170 (10pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

0108-7673/85/050454-03$01.50

(3.30), are the major goals of this paper. The distribution turns out to depend on the parameter cos T [where T, (2.1), is the three-phase invariant for the 'squared' structure], among others, so that the conditional expectation value of $|E_{\mathbf{H}}|^2-1$, given the six magnitudes (1.5), (III.3),* also depends on cos T. By suitable choice of the six magnitudes (1.5), equivalent to choosing different sets of reciprocal-lattice vectors $\mathbf{H}$, one is thus led to a system of linear equations, one of whose unknowns is cos T. In this way cos T is expressed in terms of known magnitudes $|E|$, and all, or many, reciprocal-lattice vectors $\mathbf{H}$ enter into the estimate.

In order to obtain the required conditional probability distribution of $|E_{\mathbf{H}}|$, it is necessary first to derive the joint probability distribution of the seven structure factors $E_{\mathbf{H}}$, $E_{\mathbf{h}\pm\mathbf{H}}$, $E_{\mathbf{k}\pm\mathbf{H}}$, $E_{\mathbf{l}\pm\mathbf{H}}$ and the method for doing this is briefly sketched in Appendix I.†

Owing to limitations of space, no details of the derivations are given here; only the barest outline of the method is presented in Appendices I–III.‡ The final result is briefly described in § 3, (3.21)–(3.30).

3. The estimation of $|G_{\mathbf{h}}G_{\mathbf{k}}G_{\mathbf{l}}|\cos(\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{l}})$ and $|G|^2-1$

Fix the reciprocal-lattice vectors $\mathbf{h}$, $\mathbf{k}$, $\mathbf{l}$, subject to

$$\mathbf{h}+\mathbf{k}+\mathbf{l}=0. \quad (3.1)$$

The reciprocal-lattice vector $\mathbf{H}$ ranges over subsets of reciprocal space to be defined. The six non-negative numbers R_1, R_2, R_3, $R_{\bar{1}}$, $R_{\bar{2}}$, $R_{\bar{3}}$ are specified. Make the definitions

$$R_1=|E_{\mathbf{h}+\mathbf{H}}|,\ R_2=|E_{\mathbf{k}+\mathbf{H}}|,\ R_3=|E_{\mathbf{l}+\mathbf{H}}|, \quad (3.2)$$

$$R_{\bar{1}}=|E_{\mathbf{h}-\mathbf{H}}|,\ R_{\bar{2}}=|E_{\mathbf{k}-\mathbf{H}}|,\ R_{\bar{3}}=|E_{\mathbf{l}-\mathbf{H}}|, \quad (3.3)$$

$$C=|E_{\mathbf{H}}|^2-1, \quad (3.4)$$

$$B_1=R_1^2+R_{\bar{1}}^2-2,\ B_2=R_2^2+R_{\bar{2}}^2-2,\ B_3=R_3^2+R_{\bar{3}}^2-2, \quad (3.5)$$

$$D=(R_1^2-1)(R_{\bar{2}}^2+R_{\bar{3}}^2-2)+(R_2^2-1)(R_{\bar{3}}^2+R_{\bar{1}}^2-2)+(R_3^2-1)(R_{\bar{1}}^2+R_{\bar{2}}^2-2), \quad (3.6)$$

$$D_{12}=(R_1^2-1)(R_2^2-1)+(R_{\bar{1}}^2-1)(R_{\bar{2}}^2-1), \quad (3.7)$$

$$D_{23}=(R_2^2-1)(R_3^2-1)+(R_{\bar{2}}^2-1)(R_{\bar{3}}^2-1), \quad (3.8)$$

$$D_{31}=(R_3^2-1)(R_1^2-1)+(R_{\bar{3}}^2-1)(R_{\bar{1}}^2-1), \quad (3.9)$$

$$D_1=(R_1^2-1)(R_{\bar{1}}^2-1), \quad (3.10)$$

$$D_2=(R_2^2-1)(R_{\bar{2}}^2-1), \quad (3.11)$$

$$D_3=(R_3^2-1)(R_{\bar{3}}^2-1). \quad (3.12)$$

The unknowns are defined by

$$x_1=\sigma_4\sigma_2^{-2}(|G_{\mathbf{h}}|^2-1),\quad x_2=\sigma_4\sigma_2^{-2}(|G_{\mathbf{k}}|^2-1),\quad x_3=\sigma_4\sigma_2^{-2}(|G_{\mathbf{l}}|^2-1); \quad (3.13)$$

$$\xi=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{h}}G_{\mathbf{k}}G_{\mathbf{l}}|\cos(\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{l}}), \quad (3.14)$$

$$\xi_{12}=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{h}}G_{\mathbf{k}}G_{\mathbf{h}-\mathbf{k}}|\cos(\psi_{\mathbf{h}}-\psi_{\mathbf{k}}-\psi_{\mathbf{h}-\mathbf{k}}), \quad (3.15)$$

$$\xi_{23}=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{k}}G_{\mathbf{l}}G_{\mathbf{k}-\mathbf{l}}|\cos(\psi_{\mathbf{k}}-\psi_{\mathbf{l}}-\psi_{\mathbf{k}-\mathbf{l}}), \quad (3.16)$$

$$\xi_{31}=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{l}}G_{\mathbf{h}}G_{\mathbf{l}-\mathbf{h}}|\cos(\psi_{\mathbf{l}}-\psi_{\mathbf{h}}-\psi_{\mathbf{l}-\mathbf{h}}), \quad (3.17)$$

$$\xi_1=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{h}}^2G_{2\mathbf{h}}|\cos(2\psi_{\mathbf{h}}-\psi_{2\mathbf{h}}), \quad (3.18)$$

$$\xi_2=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{k}}^2G_{2\mathbf{k}}|\cos(2\psi_{\mathbf{k}}-\psi_{2\mathbf{k}}), \quad (3.19)$$

$$\xi_3=2\sigma_4^{3/2}\sigma_2^{-3}|G_{\mathbf{l}}^2G_{2\mathbf{l}}|\cos(2\psi_{\mathbf{l}}-\psi_{2\mathbf{l}}). \quad (3.20)$$

Then (III.3) leads to ten linear equations:

$$x_1\overline{B_1^2}+x_2\overline{B_2B_1}+x_3\overline{B_3B_1}+\xi\overline{DB_1}+\xi_{12}\overline{D_{12}B_1}+\xi_{23}\overline{D_{23}B_1}+\xi_{31}\overline{D_{31}B_1}+\xi_1\overline{D_1B_1}+\xi_2\overline{D_2B_1}+\xi_3\overline{D_3B_1}=\overline{CB_1}, \quad (3.21)$$

$$x_1\overline{B_1B_2}+x_2\overline{B_2^2}+x_3\overline{B_3B_2}+\xi\overline{DB_2}+\xi_{12}\overline{D_{12}B_2}+\xi_{23}\overline{D_{23}B_2}+\xi_{31}\overline{D_{31}B_2}+\xi_1\overline{D_1B_2}+\xi_2\overline{D_2B_2}+\xi_3\overline{D_3B_2}=\overline{CB_2}, \quad (3.22)$$

$$x_1\overline{B_1B_3}+x_2\overline{B_2B_3}+x_3\overline{B_3^2}+\xi\overline{DB_3}+\xi_{12}\overline{D_{12}B_3}+\xi_{23}\overline{D_{23}B_3}+\xi_{31}\overline{D_{31}B_3}+\xi_1\overline{D_1B_3}+\xi_2\overline{D_2B_3}+\xi_3\overline{D_3B_3}=\overline{CB_3}, \quad (3.23)$$

$$x_1\overline{B_1D}+x_2\overline{B_2D}+x_3\overline{B_3D}+\xi\overline{D^2}+\xi_{12}\overline{D_{12}D}+\xi_{23}\overline{D_{23}D}+\xi_{31}\overline{D_{31}D}+\xi_1\overline{D_1D}+\xi_2\overline{D_2D}+\xi_3\overline{D_3D}=\overline{CD}, \quad (3.24)$$

$$x_1\overline{B_1D_{12}}+x_2\overline{B_2D_{12}}+x_3\overline{B_3D_{12}}+\xi\overline{DD_{12}}+\xi_{12}\overline{D_{12}^2}+\xi_{23}\overline{D_{23}D_{12}}+\xi_{31}\overline{D_{31}D_{12}}+\xi_1\overline{D_1D_{12}}+\xi_2\overline{D_2D_{12}}+\xi_3\overline{D_3D_{12}}=\overline{CD_{12}}, \quad (3.25)$$

$$x_1\overline{B_1D_{23}}+x_2\overline{B_2D_{23}}+x_3\overline{B_3D_{23}}+\xi\overline{DD_{23}}+\xi_{12}\overline{D_{12}D_{23}}+\xi_{23}\overline{D_{23}^2}+\xi_{31}\overline{D_{31}D_{23}}+\xi_1\overline{D_1D_{23}}+\xi_2\overline{D_2D_{23}}+\xi_3\overline{D_3D_{23}}=\overline{CD_{23}}, \quad (3.26)$$

$$x_1\overline{B_1D_{31}}+x_2\overline{B_2D_{31}}+x_3\overline{B_3D_{31}}+\xi\overline{DD_{31}}+\xi_{12}\overline{D_{12}D_{31}}+\xi_{23}\overline{D_{23}D_{31}}+\xi_{31}\overline{D_{31}^2}+\xi_1\overline{D_1D_{31}}+\xi_2\overline{D_2D_{31}}+\xi_3\overline{D_3D_{31}}=\overline{CD_{31}}, \quad (3.27)$$

* See deposition footnote.
† See deposition footnote.
‡ See deposition footnote.

$$x_1\overline{B_1D_1}+x_2\overline{B_2D_1}+x_3\overline{B_3D_1}+\xi\overline{DD_1}$$
$$+\xi_{12}\overline{D_{12}D_1}+\xi_{23}\overline{D_{23}D_1}+\xi_{31}\overline{D_{31}D_1}$$
$$+\xi_1\overline{D_1^2}+\xi_2\overline{D_2D_1}+\xi_3\overline{D_3D_1}=\overline{CD_1}, \quad (3.28)$$

$$x_1\overline{B_1D_2}+x_2\overline{B_2D_2}+x_3\overline{B_3D_2}+\xi\overline{DD_2}$$
$$+\xi_{12}\overline{D_{12}D_2}+\xi_{23}\overline{D_{23}D_2}+\xi_{31}\overline{D_{31}D_2}$$
$$+\xi_1\overline{D_1D_2}+\xi_2\overline{D_2^2}+\xi_3\overline{D_3D_2}=\overline{CD_2}, \quad (3.29)$$

$$x_1\overline{B_1D_3}+x_2\overline{B_2D_3}+x_3\overline{B_3D_3}+\xi\overline{DD_3}$$
$$+\xi_{12}\overline{D_{12}D_3}+\xi_{23}\overline{D_{23}D_3}+\xi_{31}\overline{D_{31}D_3}$$
$$+\xi_1\overline{D_1D_3}+\xi_2\overline{D_2D_3}+\xi_3\overline{D_3^2}=\overline{CD_3} \quad (3.30)$$

in the ten unknowns x_1, x_2, x_3, ξ, ξ_{12}, ξ_{23}, ξ_{31}, ξ_1, ξ_2, ξ_3. In view of (II.4) and (III.3) the averages in (3.21) are taken over those reciprocal-lattice vectors **H** for which B_1^2 is large and D_1^2 is small, in (3.22) those for which B_2^2 is large and D_2^2 is small, in (3.23) those for which B_3^2 is large and D_3^2 is small, in (3.24) those for which D^2 is large, in (3.25) those for which D_{12}^2 is large, in (3.26) those for which D_{23}^2 is large, in (3.27) those for which D_{31}^2 is large, in (3.28) those for which D_1^2 is large, in (3.29) those for which D_2^2 is large, and in (3.30) those for which D_3^2 is large. The subsets of reciprocal space so chosen are selected on the basis that they do not overlap and, simultaneously, yield the most reliable estimate of the unknown associated with the element on the main diagonal of the 10×10 matrix of the system (3.21)-(3.30). Thus in (3.21) **H** may be restricted to the n reciprocal-lattice vectors corresponding to the largest values of B_1^2 for which D_1^2 is simultaneously less than some small positive number t, say $t=0{\cdot}5$ *etc.* One may choose different values for n, *e.g.* $n=10$ or 20. Again **H** may be further restricted in such a way that all needed R's lie in the observed sphere of reflections; alternatively, missing R's, *i.e.* those lying outside the observed sphere of reflections, may be replaced by unity, provided they are not too numerous. The system of ten linear equations (3.21)-(3.30) is then solved for the ten unknowns x_1, x_2, x_3, ξ, ξ_{12}, ξ_{23}, ξ_{31}, ξ_1, ξ_2, ξ_3, in this way obtaining estimates for the magnitudes $|G|$ and the three-phase cosine invariants $\cos(\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{l}})$.

In the case that all atoms are identical,

$$G=|G|\exp(i\psi)=E=|E|\exp(i\varphi) \quad (3.31)$$

and the unknowns x_1, x_2 and x_3 are presumably known. In this case the system (3.21)-(3.30) may then be solved by least squares for the remaining seven unknowns.

4. Concluding remarks

Employing the quintet extensions of the three-phase structure invariant one is led to the magnitudes $|E|$ that constitute the second neighborhoods of the invariant and on which its value most sensitively depends. An appropriate conditional probability distribution then yields a system of ten linear equations, the solution of which gives an estimate of the invariant for the 'squared' structure. Applications are described in the following paper (Gilmore & Hauptman, 1985).

The research described here has been supported by Grant No. CHE-8203930 from the National Science Foundation and by a grant from the James H. Cummings Foundation, Inc.

References

GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 967-976.
GIACOVAZZO, C. (1977). *Acta Cryst.* **A33**, 527-531.
GILMORE, C. J. & HAUPTMAN, H. (1985). *Acta Cryst.* **A41**, 457-462.
HAUPTMAN, H. (1972). *Crystal Structure Determination*, pp. 192-203 and 307-326. New York: Plenum Press.
HAUPTMAN, H. (1978). *Acta Cryst.* **A34**, 525-528.
KARLE, J. (1979). *Proc. Natl Acad. Sci. USA*, **76**, 2089-2093.
KARLE, J. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 527-531.
MESSAGER, J. C. & TSOUCARIS, G. (1972). *Acta Cryst.* **A28**, 482-484.
VITERBO, D. & WOOLFSON, M. M. (1973). *Acta Cryst.* **A29**, 205-208.