Acta Cryst. (1958). **11**, 264

Phase Determination from New Joint Probability Distributions: Space Group *P*1

By J. Karle and H. Hauptman

U.S. Naval Research Laboratory, Washington 25, D.C., U.S.A.

(*Received* 4 *September* 1957)

A procedure for calculating phases in space group *P*1 is described. The new formulas constitute an improvement over earlier ones in that they make more effective use of the available data. In addition, two fundamental identities are presented which play an important role in the procedure.

The derivation of the new formulas is based on the joint probability distributions of several normalized structure factors, obtained by fixing the crystal structure and allowing the indices to range uniformly but not independently over the vectors in reciprocal space. From these, the expected values of the products of powers of several structure factors are found. The expected values, together with a subsequent integrating process, lead to explicit formulas for the magnitudes of certain structure invariants in terms of the observed intensities alone. By means of the first fundamental identity, the initial values for the magnitudes of the structure invariants are improved, thus facilitating the problem of selecting one of the two enantiomorphous structures permitted by the given set of observed intensities. Once an enantiomorph has been chosen the second identity leads to refined values (both signs and magnitudes) of the invariants. Initial values of the individual phases are then readily obtainable, and a final least-squares adjustment procedure yields improved values for the phases.

1. Introduction

In this paper formulas are derived which yield the phases of the structure factors explicitly in terms of the measured experimental intensities. The method of derivation is the same as that described in a previous paper devoted to space group $P\bar{1}$ (Hauptman & Karle, 1958). The joint probability distributions of several structure factors, which are employed, are obtained on the basis that the crystal structure is fixed and that the indices range uniformly but not independently over the vectors in reciprocal space. These distributions are used to answer the question: 'What is the expected value or average over indices of specific combinations of the structure factors or their magnitudes?'. It is the answer to this question which constitutes the phase-determining formulas.

By contrast, the joint probability distributions of several structure factors obtained in Monograph I (Hauptman & Karle, 1953*a*) were derived on the basis that the indices were fixed and the atomic coordinates ranged uniformly and independently through all possible values. These distributions afforded an answer to the question: 'What is the probability that the sign of a structure factor be plus?'. From the probability distributions so obtained, a procedure for phase determination was inferred. The distributions derived in this paper constitute another application of the methods introduced in Monograph I, yielding, however, formulas which have exact rather than probable validity.

Formulas for phase determination have already been given for space group *P*1 by Vaughan (1956, 1958) and by Karle & Hauptman (1957). Vaughan obtained an approximate formula for computing the magnitudes of invariants by means of the Patterson superposition method. His formula for large N asymptotically approaches our algebraicly exact one (1957, eq. (2·2)), corresponding to (2·1·3) of this paper when $p=q=r=2$. The new formulas presented here constitute a generalization of those obtained for space group *P*1 by means of the unified algebraic approach (Karle & Hauptman, 1957). The new basic formulas will be seen to reduce to the older algebraic ones for specific values of the parameters, i.e. $p = q = r = 0$ or 2. In addition, the integration over non-negative values of p, q and r gives rise to formulas having important practical significance. It is expected that the integrated formulas will make more efficient use of the available experimental data, although the computations are of the same kind and no more complicated than those required for the algebraic results. In fact, the same program applied by us (1957) to a ten-atom test problem can be used with the integrated formulas.

In addition to the formulas derived by probability methods, we introduce two fundamental identities. The first of these plays the important role of refining the value of the magnitude of a structure invariant (Hauptman & Karle, 1956) using the values of the magnitudes of other invariants, initially obtained from (2·1·3) or (2·2·3). The second identity is used to refine the value (magnitude and sign) of an invariant, using the values of other invariants. The second of these identities has already been noted in different forms by Vaughan (1956, 1958) and by us (Karle & Hauptman, 1957; Hauptman & Karle, 1958). We present it here in a form well adapted to the procedure for phase determination which we propose. Since the identities are valid, no matter what the chemical composition of the unit cell may be, they will be especially useful in the case that the structure consists of dis-similar atoms. A final least-squares refinement for the phases redundantly obtained from the invariants, which is also valid for dis-similar atoms, will also be presented.

2. Phase-determining formulas

We list here for ready reference all formulas derived in this paper. In these formulas the E's are the normalized structure factors, the φ's are the corresponding phases, p, q, r, and t are arbitrary real non-negative numbers, N is the number of atoms (assumed identical) in the unit cell, and Γ is the Gamma function. Although only the space group $P1$ is considered here, it will be clear that the same methods are applicable to all the space groups.

It is known that the magnitudes of the structure factors uniquely determine the magnitudes of the structure invariants (Hauptman & Karle, 1956). Equations (2·1·3) and (2·2·3) are mathematical realizations of this fact in that they express the magnitudes of certain structure invariants (namely those of the form $\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{-\mathbf{h}_1-\mathbf{h}_2}$) explicitly in terms of the magnitudes of the structure factors.

2·1. *The basic formulas, $B_{m,n}$*

$$B_{1,0}:\ \langle|E_{\mathbf{k}}|^p\rangle_{\mathbf{k}} = \Gamma\left(\frac{p+2}{2}\right)\left\{1-\frac{1}{16N}\,p(p-2)+\frac{1}{4608N^2}\times p(p-2)(p-4)(9p+10)+\ldots\right\}^{*}. \qquad (2\cdot1\cdot1)$$

$$B_{2,0}:\ |E_{\mathbf{h}}|^2-1 = \frac{4N}{pq\Gamma\left(\frac{p+2}{2}\right)\Gamma\left(\frac{q+2}{2}\right)} \times\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})(|E_{\mathbf{h}-\mathbf{k}}|^q-\overline{|E|^q})\rangle_{\mathbf{k}}+R_2\,, \qquad (2\cdot1\cdot2)$$

where

$$\overline{|E|^p} = \langle|E_{\mathbf{k}}|^p\rangle_{\mathbf{k}}$$

and

$$R_2 = -\frac{1}{64N}\,[4(p-2)(q-2)|E_{\mathbf{h}}|^4 -4(p^2+4pq+q^2-6p-6q)|E_{\mathbf{h}}|^2 +4(p+q+2)(p+q-4)]+\ldots.$$

$$B_{3,0}:\ |E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}|\cos(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{-\mathbf{h}_1-\mathbf{h}_2}) = \frac{4N^{3/2}}{pqr\Gamma\left(\frac{p+2}{2}\right)\Gamma\left(\frac{q+2}{2}\right)\Gamma\left(\frac{r+2}{2}\right)}\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p}) \times(|E_{\mathbf{h}_1+\mathbf{k}}|^q-\overline{|E|^q})(|E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^r-\overline{|E|^r})\rangle_{\mathbf{k}}+R_3\,, \qquad (2\cdot1\cdot3)$$

where

$$R_3 = -\frac{1}{4N^{\frac{1}{2}}}\,[(q-2)|E_{\mathbf{h}_1}E_{\mathbf{h}_2}|^2+(r-2)|E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}|^2 +(p-2)|E_{\mathbf{h}_1+\mathbf{h}_2}E_{\mathbf{h}_1}|^2-(p+q)|E_{\mathbf{h}_1}|^2-(q+r)|E_{\mathbf{h}_2}|^2 -(r+p)|E_{\mathbf{h}_1+\mathbf{h}_2}|^2+(p+q+r+2)]+\ldots.$$

$$B_{2,2}:\ E_{\mathbf{h}} = \frac{N^{\frac{1}{2}}\langle E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|E_{\mathbf{k}}|^p|E_{\mathbf{h}-\mathbf{k}}|^q\rangle_{\mathbf{k}}}{\langle|E_{\mathbf{k}}|^{p+2}|E_{\mathbf{h}-\mathbf{k}}|^{q+2}\rangle_{\mathbf{k}}}+\ldots. \qquad (2\cdot1\cdot4)$$

The notation $B_{m,n}$ means that each contributor to the average which appears in the corresponding formula requires a knowledge of the magnitudes of m normalized structure factors and of the phases of n of them.

The formulas $B_{1,0}$, $B_{2,0}$, $B_{3,0}$ and $B_{2,2}$ may be compared with the corresponding equations for $P\overline{1}$ of Hauptman & Karle (1958). It is also seen that these equations are generalizations of equations (2·1), (2·2), (2·4) and (2·5) of Karle & Hauptman (1957) which were obtained by algebraic means. The newer formulas reduce to the previous ones for special values of p, q, and r.

Since the coefficients of the averages appearing in (2·1·1)–(2·1·3) decrease with increasing p, q, and r, it is desirable to use as large values of p, q, and r as the number and accuracy of the data permit.

2·2. *The integrated formulas, $I_{m,n}$*

We make the definition

$$A_n = A_n(t) = \int_0^t x^n\Gamma\left(\frac{x+2}{2}\right)dx\,, \qquad (2\cdot2\cdot0)$$

and, using Simpson's Rule, find the entries listed in Table 1. Corresponding to the basic formulas $B_{m,n}$ are the integrated formulas, $I_{m,n}$.

Table 1. *The values of* $A_n(t) = \int_0^t x^n\Gamma\left(\frac{x+2}{2}\right)dx$ *for various values of t and n*

t	$n = 0$	$n = 1$	$n = 2$	$n = 3$
0	0·000	0·000	0·000	0·000
1	0·918	0·450	0·298	0·223
2	1·85	1·85	2·49	3·77
3	2·99	4·74	9·87	22·9
4	4·62	10·5	30·3	96·1
5	7·20	22·2	83·9	341
6	11·7	47·2	223	1117

$$I_{1,0}:\ \left\langle\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|}\right\rangle_{\mathbf{k}} = A_0-\frac{A_2-2A_1}{16N}+\ldots. \qquad (2\cdot2\cdot1)$$

$$I_{2,0}:\ |E_{\mathbf{h}}|^2-1 = \frac{4N}{A_1^2}\left\langle\left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|}-M\right) \times\left(\frac{|E_{\mathbf{h}-\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}-\mathbf{k}}|}-M\right)\right\rangle_{\mathbf{k}}+R_2'\,, \qquad (2\cdot2\cdot2)$$

where

---

* This formula should be compared with (30) of Hauptman & Karle (1953*b*). The discrepancy in the third terms of these two expressions is due to a numerical error in the computation of the third term of (24) in the earlier paper.

$$M = M(t) = \left\langle \frac{|E_{\mathbf{k}}|^t-1}{\log |E_{\mathbf{k}}|} \right\rangle_{\mathbf{k}},$$

and

$$\begin{aligned} R_2' = -\frac{1}{16NA_1^2}\,[&(2A_1-A_2)^2|E_{\mathbf{h}}|^4 \\ &+2(6A_1A_2-A_1A_3-2A_2^2)|E_{\mathbf{h}}|^2 \\ &-2(4A_1^2+2A_1A_2-A_1A_3-A_2^2)]+\ldots . \end{aligned}$$

The value of M is computed from the experimentally determined $|E_{\mathbf{k}}|$ (rather than from the theoretical (2·2·1)), since this is expected to compensate partially for the limitations inherent in the experimental data. The values of the A_n are computed from the entries in Table 1. Since the expression $(|E|^t-1)/\log |E|$ becomes indeterminate when $|E| = 1$, it is to be replaced by its limit, t, as $|E|$ approaches unity.

$$\begin{aligned} I_{3,0}\colon\ |E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}| \cos (\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{-\mathbf{h}_1-\mathbf{h}_2}) &= \frac{4N^{3/2}}{A_1^3} \\ \times \Bigg\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log |E_{\mathbf{k}}|}-M\right)\left(\frac{|E_{\mathbf{h}_1+\mathbf{k}}|^t-1}{\log |E_{\mathbf{h}_1+\mathbf{k}}|}-M\right) \\ \times \left(\frac{|E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^t-1}{\log |E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|}-M\right)\Bigg\rangle_{\mathbf{k}} + R_3' , \qquad (2{\cdot}2{\cdot}3) \end{aligned}$$

where

$$\begin{aligned} R_3' = \frac{1}{4N^{\frac{1}{2}}A_1}\,[&(2A_1-A_2)(|E_{\mathbf{h}_1}E_{\mathbf{h}_2}|^2 \\ &+|E_{\mathbf{h}2}E_{\mathbf{h}_1+\mathbf{h}_2}|^2+|E_{\mathbf{h}_1+\mathbf{h}_2}E_{\mathbf{h}_1}|^2)+2A_2(|E_{\mathbf{h}_1}|^2 \\ &+|E_{\mathbf{h}_2}|^2+|E_{\mathbf{h}_1+\mathbf{h}_2}|^2)-(2A_1+3A_2)]+\ldots . \end{aligned}$$

$$I_{2,2}\colon\ E_{\mathbf{h}} =$$

$$\frac{N^{\frac{1}{2}}\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log |E_{\mathbf{k}}|}\right)\left(\frac{|E_{\mathbf{h}-\mathbf{k}}|^t-1}{\log |E_{\mathbf{h}-\mathbf{k}}|}\right) E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}} \right\rangle_{\mathbf{k}}}{\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log |E_{\mathbf{k}}|}\right)\left(\frac{|E_{\mathbf{h}-\mathbf{k}}|-1}{\log |E_{\mathbf{h}-\mathbf{k}}|}\right) E_{\mathbf{k}}^2E_{\mathbf{h}-\mathbf{k}}^2 \right\rangle_{\mathbf{k}}} + \ldots . \qquad (2{\cdot}2{\cdot}4)$$

Each formula consists of a main term plus remainder terms of higher order in $1/N$. The remainder terms are thus negligible if N is sufficiently large. Only in (2·2·4) is an explicit expression for the first remainder term omitted. If t is chosen to be large (e.g. 5 or 6) and N is small, it may be necessary to obtain this remainder term by the methods to be described. Note that the averages in the denominators of (2·1·4) and (2·2·4) are to be taken over the same vectors $\mathbf{k}$ which occur in the numerators.

3. Analysis

3·1. *The probability distributions*

The method employed here is the same as that introduced in Monograph I. Here, however, the mixed moments are computed by averaging over the indices rather than over the coordinates. Thus, although the methods for obtaining the present joint distributions have already been described in Monograph I, these distributions constitute a new application of the general theory.

In addition to employing the mixed moments, an alternative procedure involving the introduction of Bessel functions, mentioned by us (Karle & Hauptman, 1958, § 3), was also used. It was found that this procedure, together with the introduction of polar coordinates, effected an essential simplification of the mathematical manipulations. This is the method which will be indicated below.

For a structure consisting of N identical point atoms in space group $P1$, the normalized structure factor $E_{\mathbf{k}}$ is defined by

$$E_{\mathbf{k}} = \frac{1}{N^{\frac{1}{2}}} \sum_{j=1}^{N} \exp (2\pi i\mathbf{k}.\mathbf{r}_j) , \qquad (3{\cdot}1{\cdot}1)$$

$$E_{\mathbf{k}} = A_{\mathbf{k}}+iB_{\mathbf{k}} = |E_{\mathbf{k}}| \exp [i\varphi_{\mathbf{k}}] , \qquad (3{\cdot}1{\cdot}2)$$

where $\mathbf{r}_j$ is the position vector of the jth atom. Let $\mathbf{k}_0, \mathbf{k}_1, \ldots, \mathbf{k}_m$ be $m+1$ arbitrary vectors in reciprocal space which are not necessarily independent. Denote by $P'(X_0, X_1, \ldots, X_m;\ Y_0, Y_1, \ldots, Y_m)$ the joint probability distribution of the $2m+2$ real and imaginary parts, $A_0, A_1, \ldots, A_m; B_0, B_1, \ldots, B_m$, of the normalized structure factors, as the $\mathbf{k}_\nu$ range uniformly (but not necessarily independently) through reciprocal space, i.e.

$$P'(X_0, X_1, \ldots, X_m;\ Y_0, Y_1, \ldots, Y_m) \times dX_0dX_1\ldots dX_mdY_0dY_1\ldots dY_m$$

is the probability that the A_ν lie between X_ν and $X_\nu+dX_\nu$ and that the B_ν lie between Y_ν and $Y_\nu+dY_\nu$, $\nu = 0, 1, \ldots, m$. Denote by $p(\xi_0, \xi_1, \ldots, \xi_m;\ \eta_0, \eta_1, \ldots, \eta_m)$ the joint probability distribution of $\xi_{\mathbf{k}_\nu} = \cos 2\pi\mathbf{k}_\nu.\mathbf{r}_j$ and $\eta_{\mathbf{k}_\nu} = \sin 2\pi\mathbf{k}_\nu.\mathbf{r}_j$, $\nu = 0, 1, \ldots, m$, where $\mathbf{r}_j$ is a fixed vector and the $\mathbf{k}_\nu$ range uniformly through reciprocal space. Then, by a well known fundamental theorem of probability theory (e.g. pp. 30, 31 of Monograph I),

$$\begin{aligned} P'(X_0, X_1, \ldots, X_m;\ Y_0, Y_1, \ldots, Y_m) \\ = \frac{1}{(2\pi)^{2m+2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left[-i \sum_{\nu=0}^{m} (X_\nu x_\nu + Y_\nu y_\nu) \right] \\ \times \prod_{j=1}^{N} q'(x_0, x_1, \ldots, x_m;\ y_0, y_1, \ldots, y_m) \\ \times dx_0dx_1\ldots dx_mdy_0dy_1\ldots dy_m , \qquad (3{\cdot}1{\cdot}3) \end{aligned}$$

where

$$\begin{aligned} q'(x_0, x_1, \ldots, x_m;\ y_0, y_1, \ldots, y_m) \\ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\xi_0, \xi_1, \ldots, \xi_m;\ \eta_0, \eta_1, \ldots, \eta_m) \times \\ \exp \left[\frac{i}{N^{\frac{1}{2}}} \sum_{\nu=0}^{m} (\xi_\nu x_\nu + \eta_\nu y_\nu) \right] d\xi_0d\xi_1\ldots d\xi_md\eta_0d\eta_1\ldots d\eta_m . \qquad (3{\cdot}1{\cdot}4) \end{aligned}$$

For the purposes of this paper, we have derived three different joint probability distributions:

(1) The distribution of a single normalized structure factor $E_{\mathbf{k}}$ as $\mathbf{k}$ ranges uniformly throughout reciprocal space.

(2) The joint distribution of the pair of normalized structure factors, $E_{\mathbf{k}}$, $E_{\mathbf{h}-\mathbf{k}}$, where $\mathbf{h}$ is a fixed vector and $\mathbf{k}$ ranges uniformly throughout reciprocal space.

(3) The joint distribution of the three normalized structure factors $E_{\mathbf{k}}$, $E_{\mathbf{h}_1+\mathbf{k}}$, $E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}$, where $\mathbf{h}_1$ and $\mathbf{h}_2$ are fixed vectors and $\mathbf{k}$ ranges uniformly throughout reciprocal space.

In order to illustrate the mathematical method, which involves the introduction of Bessel functions and the transformation to polar coordinates in (3·1·3) and (3·1·4), we consider the three-variable case. Denote by $P(R_0, R_1, \ldots, R_m; \Phi_0, \Phi_1, \ldots, \Phi_m)$ the joint probability distribution of the $2m+2$ magnitudes and phases of the normalized structure factors as the $\mathbf{k}_\nu$ range uniformly (but not necessarily independently) through reciprocal space, i.e.

$$P(R_0, R_1, \ldots, R_m; \Phi_0, \Phi_1, \ldots, \Phi_m) \times dR_0 dR_1 \ldots dR_m d\Phi_0 d\Phi_1 \ldots d\Phi_m$$

is the probability that the $|E_\nu|$ lie between R_ν and $R_\nu + dR_\nu$ and that the φ_ν lie between Φ_ν and $\Phi_\nu + d\Phi_\nu$, $\nu = 0, 1, \ldots, m$. By means of the transformations

$$X_\nu = R_\nu \cos \Phi_\nu, \quad Y_\nu = R_\nu \sin \Phi_\nu\,, \tag{3·1·5}$$

and

$$x_\nu = \varrho_\nu \cos \theta_\nu, \quad y_\nu = \varrho_\nu \sin \theta_\nu\,, \tag{3·1·6}$$

we obtain from (3·1·3) and (3·1·4), for $m = 2$,

$$\begin{aligned} &P(R_0, R_1, R_2;\ \Phi_0, \Phi_1, \Phi_2) \\ &= \frac{R_0 R_1 R_2}{(2\pi)^6} \int_{\varrho_0=0}^{\infty} \int_{\varrho_1=0}^{\infty} \int_{\varrho_2=0}^{\infty} \int_{\theta_0=0}^{2\pi} \int_{\theta_1=0}^{2\pi} \int_{\theta_2=0}^{2\pi} \\ &\times \exp\left[-i \sum_{\nu=0}^{2} R_\nu \varrho_\nu \cos(\theta_\nu - \Phi_\nu)\right] \\ &\times \varrho_0 \varrho_1 \varrho_2 \prod_{j=1}^{N} q(\varrho_0, \varrho_1, \varrho_2;\ \theta_0, \theta_1, \theta_2)\, d\varrho_0 d\varrho_1 d\varrho_2 d\theta_0 d\theta_1 d\theta_2\,, \end{aligned} \tag{3·1·7}$$

where

$$\begin{aligned} &q(\varrho_0, \varrho_1, \varrho_2;\ \theta_0, \theta_1, \theta_2) \\ &= \left\langle \exp\left\{\frac{i}{N^{\frac{1}{2}}}\left[\varrho_0 \cos(2\pi \mathbf{k}.\mathbf{r}_j - \theta_0)\right.\right.\right. \\ &\quad + \varrho_1 \cos(2\pi(\mathbf{h}_1+\mathbf{k}).\mathbf{r}_j - \theta_1) \\ &\quad \left.\left.\left. + \varrho_2 \cos(2\pi(\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}).\mathbf{r}_j - \theta_2)\right]\right\}\right\rangle_{\mathbf{k}}. \end{aligned} \tag{3·1·8}$$

Making use of the Bessel function expansions for $\exp(iz \cos \theta)$ (Watson, 1945, p. 23), it is found that (3·1·8) becomes

$$\begin{aligned} &q(\varrho_0, \varrho_1, \varrho_2;\ \theta_0, \theta_1, \theta_2) \\ &= J_0\left(\frac{1}{N^{\frac{1}{2}}}\left[\varrho_0^2 + \varrho_1^2 + \varrho_2^2 + 2\varrho_0\varrho_1 \cos(2\pi \mathbf{h}_1.\mathbf{r}_j + \theta_0 - \theta_1)\right.\right. \\ &\quad + 2\varrho_0\varrho_2 \cos(2\pi(\mathbf{h}_1+\mathbf{h}_2).\mathbf{r}_j + \theta_0 - \theta_2) \\ &\quad \left.\left. + 2\varrho_1\varrho_2 \cos(2\pi \mathbf{h}_2.\mathbf{r}_j + \theta_1 - \theta_2)\right]^{\frac{1}{2}}\right). \end{aligned} \tag{3·1·9}$$

It should be noted that in deriving (3·1·9) the subscripts 0, 1, 2 have been associated with the vector subscripts $\mathbf{k}$, $\mathbf{h}_1+\mathbf{k}$, $\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}$, respectively, of the normalized structure factors. The substitution of (3·1·9) into (3·1·7) and the evaluation of the integrals in (3·1·7) lead after tedious manipulation to the desired probability distribution,

$$\begin{aligned} &P(R_0, R_1, R_2;\ \Phi_0, \Phi_1, \Phi_2) \\ &= \frac{R_0 R_1 R_2}{\pi^3} \exp(-R_0^2 - R_1^2 - R_2^2)\left\{1 + \frac{2}{N^{\frac{1}{2}}}\left[R_0 R_1 |E_{\mathbf{h}_1}|\right.\right. \\ &\quad \times \cos(\varphi_{\mathbf{h}_1} + \Phi_0 - \Phi_1) + R_0 R_2 |E_{\mathbf{h}_1+\mathbf{h}_2}| \cos(\varphi_{\mathbf{h}_1+\mathbf{h}_2} \\ &\quad \left.\left. + \Phi_0 - \Phi_2) + R_1 R_2 |E_{\mathbf{h}_2}| \cos(\varphi_{\mathbf{h}_2} + \Phi_1 - \Phi_2)\right] + \ldots\right\}. \end{aligned} \tag{3·1·10}$$

We have carried out the probability distribution to terms of the order $1/N^2$ as is required for the purposes of this paper. However, owing to their extreme length, these terms are not included here. Similar probability distributions were obtained for the structure factor $E_{\mathbf{k}}$ and for the pair of structure factors $E_{\mathbf{k}}$, $E_{\mathbf{h}-\mathbf{k}}$. It is anticipated that the probability distributions and the expected values (see § 3·2) will be made available at a future date.

3·2. *The average values*

By means of the following formula

$$\begin{aligned} &\langle |E_{\mathbf{k}}|^p |E_{\mathbf{h}_1+\mathbf{k}}|^q |E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^r \rangle_{\mathbf{k}} \\ &= \int_{R_0=0}^{\infty} \int_{R_1=0}^{\infty} \int_{R_2=0}^{\infty} \int_{\Phi_0=0}^{2\pi} \int_{\Phi_1=0}^{2\pi} \int_{\Phi_2=0}^{2\pi} R_0^p R_1^q R_2^r \\ &\times P(R_0, R_1, R_2;\ \Phi_0, \Phi_1, \Phi_2)\, dR_0 dR_1 dR_2 d\Phi_0 d\Phi_1 d\Phi_2\,, \end{aligned} \tag{3·2·1}$$

where P is given by (3·1·10), the average value of $|E_{\mathbf{k}}|^p |E_{\mathbf{h}_1+\mathbf{k}}|^q |E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^r$ is obtained, where k ranges uniformly throughout reciprocal space. In a similar way the average values of $|E_{\mathbf{k}}|^p$, $|E_{\mathbf{k}}|^p |E_{\mathbf{h}-\mathbf{k}}|^q$, and $|E_{\mathbf{k}}|^p |E_{\mathbf{h}-\mathbf{k}}|^q E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}$ may be found.

3·3. *The basic formulas*

The basic formula $B_{3,0}$ is obtained by expanding

$$\langle (|E_{\mathbf{k}}|^p - \overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{k}}|^q - \overline{|E|^q})(|E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^r - \overline{|E|^r}) \rangle_{\mathbf{k}}\,. \tag{3·3·1}$$

It is found that (3·3·1) is expressible in terms of several of the averages obtained in § 3·2. We are then immediately led to (2·1·3). In a similar way the remaining basic formulas are found.

Since the basic formulas are valid for all nonnegative values of the parameters, p, q and r, they constitute a generalization of the corresponding algebraic formulas given by us (Karle & Hauptman, 1957). Clearly, for different choices of the values of p, q, and r, we have different relationships for computing the invariants. For the case of an infinite set of accurate data any one set of values of p, q and r would suffice to give accurate values for the invariants. In the practical case of limited experimental data, the possibility of using various sets of values for p, q and r makes for more effective use of the data. In particular, the larger values of p, q, and r are especially useful when the data are sufficiently accurate and extensive.

3·4. *The integrated formulas*

Instead of computing the basic formulas for many different sets of p, q, and r, it is desirable to obtain a single formula which is equivalent to computing these formulas over a continuous range of values from 0 to t for p, q, and r. It is advantageous to use as large a value of t as the accuracy and extent of the experimental data warrant.

In order to prove the typical integrated formulas (2·2·3), we multiply (2·1·3) by

$$pqr\Gamma\left(\frac{p+2}{2}\right)\Gamma\left(\frac{q+2}{2}\right)\Gamma\left(\frac{r+2}{2}\right),$$

and integrate between the limits 0 to t for p, q, and r (using $\int a^x dx = a^x/\log a$). The remaining integrated formulas are derived in the same way where reference should be made to Table 1.

4. Dis-similar atoms

In the case that the structure consist of N atoms per unit cell, not all identical, we suggest, on the basis of results previously obtained in Monograph I (1953), that N, in formulas (2·1·1)–(2·1·4), (2·2·1)–(2·2·4), and (5·1·1) written below, be replaced as follows:

$$N \rightarrow \sigma_2^3/\sigma_3^2\,, \qquad (4{\cdot}1)$$

where

$$\sigma_n = \sum_{j=1}^{N} Z_j^n\,, \qquad (4{\cdot}2)$$

and Z_j is the atomic number of the jth atom. Evidently σ_2^3/σ_3^2 reduces to N in the case that all atoms are identical. The above mentioned equations, for the case of dis-similar atoms, will then no longer have exact validity, but merely probable validity. In the next two sections, however, the limitation imposed by the presence of unlike atoms will be removed. In fact, once approximate values for invariants or phases have been found, it will be seen that the validity of the refinement procedures to be described is independent of the chemical composition of the unit cell.

5. Two fundamental identities

5·1. *The first fundamental identity*

The identity

$$|E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}|\cos(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{-\mathbf{h}_1-\mathbf{h}_2}) \approxeq \frac{\begin{array}{r} 4\sum\limits_{\mathbf{k}'}[|E_{\mathbf{h}_1}E_{\mathbf{k}'}E_{\mathbf{h}_1+\mathbf{k}'}|\cos(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{k}'}+\varphi_{-\mathbf{h}_1-\mathbf{k}'})][|E_{\mathbf{h}_2}E_{\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'}|\cos(\varphi_{\mathbf{h}_2}+\varphi_{-\mathbf{k}'}+\varphi_{-\mathbf{h}_2+\mathbf{k}'})] \\ [|E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'}|\cos(\varphi_{-\mathbf{h}_1-\mathbf{h}_2}+\varphi_{\mathbf{h}_1+\mathbf{k}'}+\varphi_{\mathbf{h}_2-\mathbf{k}'})] \end{array}}{\left[1+\frac{1}{2N}(|E_{\mathbf{h}_1}|^2+|E_{\mathbf{h}_2}|^2+|E_{\mathbf{h}_1+\mathbf{h}_2}|^2)\right]\sum\limits_{\mathbf{k}'}|E_{\mathbf{k}'}E_{\mathbf{h}_1+\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'}|^2}\,, \qquad (5{\cdot}1{\cdot}1)$$

where $\mathbf{k}'$ ranges over all vectors in reciprocal space, is a useful supplement to (2·1·3) and (2·2·3). In practice something of the order of fifty values for $\mathbf{k}'$ may suffice to make the application of (5·1·1) worth while. The symbol $\approxeq$ indicates that the identity is only approximate owing to the omission of correction terms higher than order of $1/N$ in the bracketed term of the denominator of (5·1·1). It seems unlikely that the terms of higher order than $1/N$ would ever be required in practice. They may be obtained however (as is (5·1·1) itself) from a computation of the expected value of the numerator of (5·1·1), making use of the joint probability distribution indicated in (3·1·10), with the higher-order terms included. It should be noted that each of the bracketed terms in the numerator of (5·1·1) is obtained directly from a calculation of formulas (2·1·3) or (2·2·3). We thus have a means for improving the initially calculated value of the magnitude of an invariant by making use of the initially calculated values of the magnitudes of other invariants.

5·2. *The second fundamental identity*

Once refined values for the magnitudes of the structure invariants have been found by means of § 5·1, one of the two enantiomorphs permitted by the magnitudes of the structure factors may be selected by the method described by us (Karle & Hauptman, 1957). This is equivalent to selecting those signs for the structure invariants which belong to this choice of enantiomorph. Having a set of values for the structure invariants, improved values may be obtained by means of the exact identity,

$$E_{\mathbf{h}_1}E_{\mathbf{h}_1}E_{-\mathbf{h}_1-\mathbf{h}_2} \equiv \frac{\begin{array}{r} \sum\limits_{\mathbf{k}'}(E_{\mathbf{h}_1}E_{\mathbf{k}'}E_{-\mathbf{h}_1-\mathbf{k}'})(E_{\mathbf{h}_2}E_{-\mathbf{k}'}E_{-\mathbf{h}_2+\mathbf{k}'}) \\ (E_{-\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'}) \end{array}}{\sum\limits_{\mathbf{k}'}|E_{\mathbf{k}'}E_{\mathbf{h}_1+\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'}|^2}\,, \qquad (5{\cdot}2{\cdot}1)$$

where $\mathbf{k}'$ ranges over any chosen set of vectors. In fact, (5·2·1) is an identity for a single value of $\mathbf{k}'$. Clearly

(5·2·1) is valid regardless of the chemical composition of the unit cell. Each of the three products to be used in the parentheses occurring in the numerator of (5·2·1) is obtained from § 5·1 and the procedure for choosing an enantiomorph referred to above.

6. Least-squares refinement of phases

From the refined values of the invariants obtained in § 5·2, initial values of the phases may be determined by the procedure outlined by us (Karle & Hauptman, 1957). To this end the values of a primitive linearly independent triple of phases, $\varphi_{\mathbf{h}_1}, \varphi_{\mathbf{h}_2}, \varphi_{\mathbf{h}_3}$ must first be specified, thus uniquely fixing the origin. From these phase specifications and the values of the invariants, initial values for all phases may be obtained by a simple arithmetic process. Improved values for the phases may then be obtained by means of a least-squares procedure (which is valid regardless of the chemical composition of the unit cell) described by us (Karle & Hauptman, 1957). The final formulas may be written in the following form

$$\cos\varphi_{\mathbf{h}} = C/(C^2+S^2)^{\frac{1}{2}}\,, \qquad (6{\cdot}1)$$

$$\sin\varphi_{\mathbf{h}} = S/(C^2+S^2)^{\frac{1}{2}}\,, \qquad (6{\cdot}2)$$

where

$$C = \sum_{i,j} |E_{ij}| \cos(b_{ij}-\varphi_{ij})\,, \qquad (6{\cdot}3)$$

$$S = \sum_{i,j} |E_{ij}| \sin(b_{ij}-\varphi_{ij})\,, \qquad (6{\cdot}4)$$

$$E_{ij} = E_{\mathbf{h}_i}E_{\mathbf{h}_j}\,, \qquad (6{\cdot}5)$$

$$\varphi_{ij} = \varphi_{\mathbf{h}_i}+\varphi_{\mathbf{h}_j}; \qquad (6{\cdot}6)$$

and b_{ij} is the value of the invariant $\varphi_{\mathbf{h}}+\varphi_{\mathbf{h}_i}+\varphi_{\mathbf{h}_j}$ (obtained from § 5·2), where

$$\mathbf{h}+\mathbf{h}_i+\mathbf{h}_j = 0\,, \qquad (6{\cdot}7)$$

and the sums in (6·3) and (6·4) are computed over all, i and j satisfying (6·7). Thus equations (6·1)–(6·7) yield an improved value for the phase $\varphi_{\mathbf{h}}$ once initial values $\varphi_{\mathbf{h}_i}$ and $\varphi_{\mathbf{h}_j}$ of the phases are known. Equations (6·3) and (6·4) are to be identified with (4·40) and (4·41) of our previous paper on space group $P1$ (Karle & Hauptman, 1957).

7. Concluding remarks

By means of the joint probability distribution, the basic and integrated formulas of this paper were obtained. It is apparent that these formulas constitute a considerable generalization over those obtained by us (Karle & Hauptman, 1957) by means of the unified algebraic method. They thus may be expected to make more efficient use of the available data. This is accomplished without any additional complication in the program for computing. The required computations may be easily and rapidly carried out on present-day computing facilities, and the basic program has already been worked out for a medium-speed machine at the Bureau of Standards.

The procedure for phase determination presented in this paper is an improvement over our previous one not only as a consequence of improved formulas for computing the magnitudes of invariants but also as a consequence of the subsequent refinements effected by the two fundamental identities. The importance of the first fundamental identity is that it yields the accurate values for the magnitudes of the invariants which are needed in order to fix an enantiomorph unambiguously. Once one of the two enantiomorphs has been thus chosen (i.e. signs as well as magnitudes of the structure invariants have been determined) the second fundamental identity enables one to compute accurately the values (signs and magnitudes) of all the important invariants. These, together with the arbitrarily chosen values of three phases constituting a primitive linearly independent set, lead to initial values of the phases. Finally, the least-squares adjustment then yields improved values for the phases.

It is to be noted that in the procedure for phase determination which we suggest, equations (2·1·4) and (2·2·4) do not play an important role. They are to be regarded, instead, as supplementary formulas which may be occasionally useful, particularly in the case that the structure consists of identical atoms.

The details of the mathematical analyses in this paper are so long and tedious that they have been omitted altogether. Only the briefest sketch of the analysis, already presented for space group $P\bar{1}$ at somewhat greater length in a previous paper (Hauptman & Karle, 1958), and a statement of the final results are included here. It is expected that the missing details will be published at a later date.

References

Hauptman, H. & Karle, J. (1953*a*). *Solution of the Phase Problem I. The Centrosymmetric Crystal.* A. C. A. Monograph No. 3. Brooklyn: Polycrystal Book Service.

Hauptman, H. & Karle, J. (1953*b*). *Acta Cryst.* **6**, 136.

Hauptman, H. & Karle, J. (1956). *Acta Cryst.* **9**, 45.

Hauptman, H. & Karle, J. (1958). *Acta Cryst.* **11**, 149.

Karle, J. & Hauptman, H. (1957). *Acta Cryst.* **10**, 515.

Vaughan, P. A. (1956). A. C. A. Annual Meeting, French Lick, Ind.

Vaughan, P. A. (1958). *Acta Cryst.* **11**, 111.

Watson, G. N. (1945). *A Treatise on the Theory of Bessel Functions.* Cambridge: University Press.