Acta Cryst. (1958). **11**, 149

Phase Determination from New Joint Probability Distributions: Space Group $P\bar{1}$

BY H. HAUPTMAN AND J. KARLE

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

(*Received* 17 *May* 1957)

New joint probability distributions of several normalized structure factors are 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 product of powers of several structure factors are found. A subsequent integrating process, together with an important identity, then yield greatly improved formulas for phase determination by means of which the phases of the structure factors are obtainable from the magnitudes of the structure factors alone. Owing to the greater power of these formulas the data are used more efficiently, and complex structures may be more readily analyzed than heretofore.

1. Introduction

Joint probability distributions of several structure factors were first derived in our Monograph I (Hauptman & Karle, 1953). These were obtained by fixing the vectors $\mathbf{h}_1, \mathbf{h}_2, \ldots$ and allowing the atomic coordinates in the asymmetric unit to range uniformly and independently throughout the unit cell. In this way the probability that the sign of a structure factor be plus was determined on the basis that the magnitudes, and possibly the signs, of a certain set of structure factors are known. From these probabilities phase determining formulas were derived and a procedure for phase determination was inferred. These formulas had, however, only probable validity. The question naturally arose whether there exist formulas which, under suitable circumstances, have exact, rather than merely probable, validity.

A partial affirmative answer to this question had already been given by Hughes (1953) and subsequently by Cochran (1954) and Bullough & Cruickshank (1955). A more complete answer was obtained by us using a unified algebraic approach (Hauptman & Karle, 1957). The algebraic approach leads in general to formulas which are the exact analogues of the probability formulas previously derived. In addition, new formulas were found in this way which surpassed the others in that the phases of all structure factors could be more readily obtained by means of explicit expressions in terms of the magnitudes of the structure factors alone. For their exact validity the algebraic formulas require that the structure consist of N identical point atoms, a restriction of no important consequence for centrosymmetric crystals, and that a mild condition concerning the rational independence of atomic coordinates be fulfilled. Now the question arises whether the probability methods are capable of yielding the more powerful algebraic formulas.

The present paper not only gives an affirmative answer to this question but generalizes the algebraic formulas in a way which should prove to be of important practical significance. Here, too, the condition concerning the rational independence of atomic coordinates (Hauptman & Karle, 1957) must be fulfilled. In this paper we introduce the concept of the joint probability distribution of several structure factors, based upon the conditions that the crystal structure be fixed and that the vector $\mathbf{k}$ be permitted to range uniformly over all vectors in reciprocal space. This is in contrast to the previously obtained joint probability distributions in which the indices were fixed and the atomic coordinates were uniformly and independently distributed. The latter distributions answered the question: 'What is the probability that the sign of a structure factor be plus?'. The distributions here obtained answer the question: 'What is the expected value or average over indices of specific combinations of the structure factors or their magnitudes?'.

All phase-determining relations require for their successful application the collection of a very large number of data, and the formulas to be described in this paper are no exception. Because of their greater power, however, it is to be expected that the present formulas will require fewer data than those derived previously.

2. Phase-determining formulas

We list here for ready reference all important phase-determining formulas derived in this paper. In these formulas the E's are the normalized structure factors; 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 $P\bar{1}$ is considered here, it will be clear that the methods are applicable to all the space groups, non-centrosymmetric as well as centrosymmetric.

2·1. *The basic formulas*, $B_{m,n}$

$$B_{1,0}:\ \langle |E_{\mathbf{k}}|^p\rangle_{\mathbf{k}} = \frac{2^{(p+1)/2}}{(2\pi)^{\frac{1}{2}}}\,\Gamma\left(\frac{p+1}{2}\right) \times\left\{1-\frac{p(p-2)}{8N}+\frac{p(p-2)(p-4)(9p+10)}{1152N^2}+\ldots\right\}^*. \tag{2·1·1}$$

$$B_{2,0}:\ E_{2\mathbf{h}} = -\frac{8\pi N^{3/2}}{2^{(p+q+2)/2}pq\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)} \times\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})(|E_{\mathbf{h}+\mathbf{k}}|^q-\overline{|E|^q})\rangle_{\mathbf{k}} +2N^{\frac{1}{2}}(E_{\mathbf{h}}^2-1)+R_2\,, \tag{2·1·2}$$

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

$$R_2 = \frac{1}{12N^{\frac{1}{2}}}[2(p-2)(q-2)E_{\mathbf{h}}^4-3(p^2+4pq+q^2-6p-6q) \times E_{\mathbf{h}}^2+3(p+q+2)(p+q-4)]+\ldots.$$

$$B_{3,0}:\ E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2} = \frac{(2\pi N)^{3/2}}{2^{(p+q+r+3)/2}pqr\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{r+1}{2}\right)} \times\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{k}}|^q-\overline{|E|^q})(|E_{\mathbf{h}_2+\mathbf{k}}|^r-\overline{|E|^r})\rangle_{\mathbf{k}}+R_3\,, \tag{2·1·3}$$

where

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

$$B_{3,0}:\ (E_{\mathbf{h}}^2-\tfrac{1}{2})E_{2\mathbf{h}} = \frac{(2\pi N)^{3/2}}{2^{(p+q+r+3)/2}pqr\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{r+1}{2}\right)} \times\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})(|E_{\mathbf{h}+\mathbf{k}}|^q-\overline{|E|^q})(|E_{\mathbf{h}-\mathbf{k}}|^r-\overline{|E|^r})\rangle_{\mathbf{k}}+R_{3a}\,, \tag{2·1·3a}$$

where

$$R_{3a} = -\frac{1}{N^{\frac{1}{2}}}\left[\frac{E_{\mathbf{h}}^2}{4}((p-2)E_{\mathbf{h}}^2+(q+r-4)E_{2\mathbf{h}}^2) -\tfrac{1}{4}((2p+q+r)E_{\mathbf{h}}^2+(q+r)E_{2\mathbf{h}}^2-(p+q+r+2)) -E_{\mathbf{h}}E_{3\mathbf{h}}\right]+\ldots.$$

* This formula should be compared with (23) of Karle & Hauptman (1953). The discrepancy in the third terms of these two expressions is due to a numerical error in the computation of the third term of (20) in the earlier paper.

$$B_{2,1}:\ E_{\mathbf{h}} = \frac{(8\pi)^{\frac{1}{2}}N}{2^{(p+1)/2}p\Gamma\left(\frac{p+1}{2}\right)} \times\frac{\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})|E_{\mathbf{h}+2\mathbf{k}}|^qE_{\mathbf{h}+2\mathbf{k}}\rangle_{\mathbf{k}}}{\langle|E_{\mathbf{h}+2\mathbf{k}}|^{q+2}\rangle_{\mathbf{k}}}+\ldots \tag{2·1·4}$$

$$B_{2,1}:\ E_{2\mathbf{h}} = \frac{2(2\pi)^{\frac{1}{2}}N}{2^{\frac{p+1}{2}}p\Gamma\left(\frac{p+1}{2}\right)} \times\frac{\langle(|E_{\mathbf{h}+\mathbf{k}}|^p-\overline{|E|^p})|E_{2\mathbf{k}}|^qE_{2\mathbf{k}}\rangle_{\mathbf{k}}}{\langle|E_{2\mathbf{k}}|^{q+2}\rangle_{\mathbf{k}}}+\ldots. \tag{2·1·4a}$$

$$B_{3,2}:\ E_{\mathbf{h}_1}E_{\mathbf{h}_2} = \frac{(2\pi)^{\frac{1}{2}}N}{2^{\frac{p+1}{2}}p\Gamma\left(\frac{p+1}{2}\right)} \times\frac{\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})|E_{\mathbf{h}_1+\mathbf{k}}|^q|E_{\mathbf{h}_2+\mathbf{k}}|^rE_{\mathbf{h}_1+\mathbf{k}}E_{\mathbf{h}_2+\mathbf{k}}\rangle_{\mathbf{k}}}{\langle|E_{\mathbf{h}_1+\mathbf{k}}|^{q+2}|E_{\mathbf{h}_2+\mathbf{k}}|^{r+2}\rangle_{\mathbf{k}}}+R_5\,, \tag{2·1·5}$$

where

$$R_5 = \frac{E_{\mathbf{h}_1+\mathbf{h}_2}}{2N^{\frac{1}{2}}}-\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{8pN^{\frac{1}{2}}}[4(pq+qr+q+r)E_{\mathbf{h}_1}^2 +4(pr+qr+q+r)E_{\mathbf{h}_2}^2+4(qr+q+r)E_{\mathbf{h}_1-\mathbf{h}_2}^2 -p(p+4q+4r+6)-4q-4r]+\ldots.$$

$$B_{2,2}:\ E_{\mathbf{h}} = \frac{N^{\frac{1}{2}}\langle E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}|E_{\mathbf{k}}|^p|E_{\mathbf{h}+\mathbf{k}}|^q\rangle_{\mathbf{k}}}{\langle|E_{\mathbf{k}}|^{p+2}|E_{\mathbf{h}+\mathbf{k}}|^{q+2}\rangle_{\mathbf{k}}}+\ldots. \tag{2·1·6}$$

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 signs of n of them. When more than one formula has the same $B_{m,n}$ label, it is seen that one is a special case of the other. We note that Σ_2 of Monograph I is related to the special case $p = q = 0$ of $B_{2,2}$ while Σ_3 corresponds to the special case $p = 2$, $q = 0$ of $B_{2,1}$ (2·1·4*a*).

It is evident that if the data are sufficiently accurate and extensive then the larger values of p, q, r yield the more useful formulas, since the coefficients of the respective averages will then be smaller.

Table 1

The values of

$$G_n(t) = \int_0^t x^n 2^{(x+1)/2}\Gamma\left(\frac{x+1}{2}\right)dx$$

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	2·122	1·023	0·673	0·502
2	4·31	4·35	5·91	9·01
3	7·46	12·33	26·44	62·5
4	12·97	31·9	96·5	314
5	24·1	82·7	329	1384
6	49·2	222	1110	5755

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

We make the definition

$$G_n = G_n(t) = \int_0^t x^n 2^{(x+1)/2} \Gamma\left(\frac{x+1}{2}\right) dx\,, \qquad (2{\cdot}2{\cdot}0)$$

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

$$I_{1,0}\colon\ \left\langle \frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} \right\rangle_{\mathbf{k}} = \frac{1}{(2\pi)^{\frac{1}{2}}}\left\{G_0 - \frac{G_2-2G_1}{8N} + \ldots\right\}. \quad (2{\cdot}2{\cdot}1)$$

$$I_{2,0}\colon\ E_{2\mathbf{h}} = -\frac{8\pi N^{3/2}}{G_1^2}\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} - M\right) \times \left(\frac{|E_{\mathbf{h+k}}|^t-1}{\log|E_{\mathbf{h+k}}|} - M\right)\right\rangle_{\mathbf{k}} + 2N^{\frac{1}{2}}(E_{\mathbf{h}}^2-1) + R_2'\,, \qquad (2{\cdot}2{\cdot}2)$$

where

$$R_2' = \frac{1}{G_1^2}\int_0^t\int_0^t 2^{(p+q+2)/2} pq\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right) R_2\, dp\, dq\,,$$

hence readily computed in terms of the entries in Table 1, while

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

is computed from the experimentally determined $|E_{\mathbf{k}}|$ (rather than from the theoretical (2·2·1)). We note that the expression $(|E|^t-1)/\log|E|$ becomes indeterminate when $|E| = 1$, and is therefore to be replaced by its limit, namely t, as $|E|$ approaches unity.

$$I_{3,0}\colon\ E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2} = \frac{(2\pi N)^{3/2}}{G_1^3} \times \left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} - M\right)\left(\frac{|E_{\mathbf{h}_1+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_1+\mathbf{k}}|} - M\right) \times \left(\frac{|E_{\mathbf{h}_2+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_2+\mathbf{k}}|} - M\right)\right\rangle_{\mathbf{k}} + R_3'\,, \qquad (2{\cdot}2{\cdot}3)$$

where

$$R_3' = \frac{1}{G_1^3}\int_0^t\int_0^t\int_0^t 2^{(p+q+r+3)/2} \times pqr\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{r+1}{2}\right) R_3\, dp\, dq\, dr$$

is easily computed from the entries in Table 1.

$$I_{3,0}\colon\ (E_{\mathbf{h}}^2-\tfrac{1}{2})E_{2\mathbf{h}} = \frac{(2\pi N)^{3/2}}{G_1^3} \times \left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} - M\right)\left(\frac{|E_{\mathbf{h+k}}|^t-1}{\log|E_{\mathbf{h+k}}|} - M\right)\left(\frac{|E_{\mathbf{h-k}}|^t-1}{\log|E_{\mathbf{h-k}}|} - M\right)\right\rangle_{\mathbf{k}} + R_{3a}'\,, \qquad (2{\cdot}2{\cdot}3a)$$

where

$$R_{3a}' = \frac{1}{G_1^3}\int_0^t\int_0^t\int_0^t 2^{(p+q+r+3)/2} pqr\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{r+1}{2}\right) R_{3a}\, dp\, dq\, dr$$

is computed by means of Table 1.

$$I_{2,1}\colon\ E_{\mathbf{h}} = \frac{2(2\pi)^{\frac{1}{2}}N}{G_1} \times \frac{\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} - M\right)\left(\frac{|E_{\mathbf{h+2k}}|^t-1}{\log|E_{\mathbf{h+2k}}|}\right)E_{\mathbf{h+2k}}\right\rangle_{\mathbf{k}}}{\left\langle \left(\frac{|E_{\mathbf{h+2k}}|^t-1}{\log|E_{\mathbf{h+2k}}|}\right)E_{\mathbf{h+2k}}^2\right\rangle_{\mathbf{k}}} + \ldots. \qquad (2{\cdot}2{\cdot}4)$$

$$I_{2,1}\colon\ E_{2\mathbf{h}} = \frac{2(2\pi)^{\frac{1}{2}}N}{G_1} \times \frac{\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} - M\right)\left(\frac{|E_{2\mathbf{k}}|^t-1}{\log|E_{2\mathbf{k}}|}\right)E_{2\mathbf{k}}\right\rangle_{\mathbf{k}}}{\left\langle \left(\frac{|E_{2\mathbf{k}}|^t-1}{\log|E_{2\mathbf{k}}|}\right)E_{2\mathbf{k}}^2\right\rangle_{\mathbf{k}}} + \ldots. \qquad (2{\cdot}2{\cdot}4a)$$

$$I_{3,2}\colon\ E_{\mathbf{h}_1}E_{\mathbf{h}_2} = \frac{(2\pi)^{\frac{1}{2}}N}{G_1}\cdot\frac{\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|} - M\right)\frac{|E_{\mathbf{h}_1+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_1+\mathbf{k}}|}\cdot\frac{|E_{\mathbf{h}_2+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_2+\mathbf{k}}|} E_{\mathbf{h}_1+\mathbf{k}}E_{\mathbf{h}_2+\mathbf{k}}\right\rangle_{\mathbf{k}}}{\left\langle \frac{|E_{\mathbf{h}_1+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_1+\mathbf{k}}|}\cdot\frac{|E_{\mathbf{h}_2+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_2+\mathbf{k}}|} E_{\mathbf{h}_1+\mathbf{k}}^2E_{\mathbf{h}_2+\mathbf{k}}^2\right\rangle_{\mathbf{k}}} + R_5'\,, \qquad (2{\cdot}2{\cdot}5)$$

where

$$R_5' = \frac{\left\langle \int_0^t\int_0^t\int_0^t 2^{(p+1)/2} p\Gamma\left(\frac{p+1}{2}\right)|E_{\mathbf{h}_1+\mathbf{k}}|^{q+2}|E_{\mathbf{h}_2+\mathbf{k}}|^{r+2}R_5\, dp\, dq\, dr\right\rangle_{\mathbf{k}}}{G_1\left\langle \frac{|E_{\mathbf{h}_1+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_1+\mathbf{k}}|}\cdot\frac{|E_{\mathbf{h}_2+\mathbf{k}}|^t-1}{\log|E_{\mathbf{h}_2+\mathbf{k}}|} E_{\mathbf{h}_1+\mathbf{k}}^2E_{\mathbf{h}_2+\mathbf{k}}^2\right\rangle_{\mathbf{k}}}$$

is computed from Table 1.

$I_{2,2}$:

$$E_{\mathbf{h}} = \frac{N^{\frac{1}{2}}\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|}\right)\left(\frac{|E_{\mathbf{h+k}}|^t-1}{\log|E_{\mathbf{h+k}}|}\right)E_{\mathbf{k}}E_{\mathbf{h+k}}\right\rangle_{\mathbf{k}}}{\left\langle \left(\frac{|E_{\mathbf{k}}|^t-1}{\log|E_{\mathbf{k}}|}\right)\left(\frac{|E_{\mathbf{h+k}}|^t-1}{\log|E_{\mathbf{h+k}}|}\right)E_{\mathbf{k}}^2E_{\mathbf{h+k}}^2\right\rangle_{\mathbf{k}}} + \ldots. \qquad (2{\cdot}2{\cdot}6)$$

We note that each formula consists of a main term plus a remainder term of higher order in $1/N$. Thus the remainder term is negligible if N is sufficiently large. In several of the formulas explicit expressions for the remainder terms have not been derived. In the case that t is chosen to be large (e.g. 5 or 6) and N is small, it may be necessary to obtain these remainder terms. This may be readily accomplished by the methods to be described.

The formulas obtained by the unified algebraic approach are the special cases of the basic formulas in which $p, q, r = 0$ or 2. Equations (2·1·5) and (2·2·5) are likely to be most useful if $E_{\mathbf{h}_1}$ and $E_{\mathbf{h}_2}$ are linearly dependent (modulo 2), in which case $\mathbf{k}$ ranges over all vectors congruent modulo 2 to $\mathbf{h}_1$ (or $\mathbf{h}_2$). It is understood that, in practice, the averages in the denominators of (2·1·4)–(2·1·6) and (2·2·4)–(2·2·6) are to be taken over the same vectors $\mathbf{k}$ which occur in the respective numerators.

3. The joint probability distributions

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

For a structure consisting of N identical point atoms in space group $P\bar{1}$ the normalized structure factor $E_{\mathbf{k}}$ is defined by means of

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

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 $P(X_0, X_1, \ldots, X_m)$ the joint probability distribution of the $m+1$ normalized structure factors $E_{\mathbf{k}_0}, E_{\mathbf{k}_1}, \ldots E_{\mathbf{k}_m}$ as the $\mathbf{k}_\nu$ range uniformly (but not necessarily independently) through reciprocal space, i.e. $P(X_0, X_1, \ldots, X_m)dX_0dX_1\ldots dX_m$ is the probability that $E_{\mathbf{k}_\nu}$ lie between X_ν and $X_\nu+dX_\nu$, $\nu = 0, 1, \ldots, m$. Denote by $p(\xi_0, \xi_1, \ldots, \xi_m)$ the joint probability distribution of $\xi_{\mathbf{k}_\nu} = 2\cos 2\pi\mathbf{k}_\nu.\mathbf{r}_j$, $\nu = 0, 1, \ldots, m$, 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), we have

$$P(X_0, X_1, \ldots, X_m) = \frac{1}{(2\pi)^{m+1}} \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \times \exp\left(-i\sum_{\nu=0}^{m} X_\nu x_\nu\right) \prod_{j=1}^{N/2} q(x_0, x_1, \ldots, x_m)dx_0dx_1\ldots dx_m \,, \qquad (3·2)$$

where

$$q(x_0, x_1, \ldots, x_m) = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} p(\xi_0, \xi_1, \ldots, \xi_m) \times \exp\left(\frac{i}{N^{\frac{1}{2}}}\sum_{\nu=0}^{m}\xi_\nu x_\nu\right) d\xi_0 d\xi_1\ldots d\xi_m \,. \qquad (3·3)$$

Following Monograph I, p. 32, we find

$$q(x_0, x_1, \ldots, x_m) = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} p(\xi_0, \xi_1, \ldots, \xi_m)\left\{1+\frac{i}{N^{\frac{1}{2}}}\times \sum_{\nu=0}^{m}\xi_\nu x_\nu - \frac{1}{2!N}\left(\sum_{\nu=0}^{m}\xi_\nu x_\nu\right)^2 - \frac{i}{3!N^{3/2}}\left(\sum_{\nu=0}^{m}\xi_\nu x_\nu\right)^3 + \ldots\right\}, \qquad (3·4)$$

each term of which has the form of a mixed moment

$$\mu_{\lambda_0\lambda_1\ldots\lambda_m} = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} p(\xi_0, \xi_1, \ldots, \xi_m) \times \xi_0^{\lambda_0}\xi_1^{\lambda_1}\ldots\xi_m^{\lambda_m} d\xi_0 d\xi_1\ldots d\xi_m \,. \qquad (3·5)$$

Interpreting (3·5) as the expected value, or average, of $\xi_0^{\lambda_0}\xi_1^{\lambda_1}\ldots\xi_m^{\lambda_m}$, we conclude that

$$\mu_{\lambda_0\lambda_1\ldots\lambda_m} = \langle \xi_0^{\lambda_0}\xi_1^{\lambda_1}\ldots\xi_m^{\lambda_m}\rangle_{\mathbf{k}_0, \mathbf{k}_1, \ldots, \mathbf{k}_m} \,, \qquad (3·6)$$

where

$$\xi_\nu = 2\cos 2\pi\mathbf{k}_\nu.\mathbf{r}_j, \quad \nu = 0, 1, \ldots, m \,. \qquad (3·7)$$

We note that, in contrast to the mixed moments obtained previously in Monograph I, (3·6) is computed by fixing the atomic coordinates $\mathbf{r}_j$ and averaging over the vectors $\mathbf{k}_\nu$, $\nu = 0, 1, \ldots, m$. The mixed moments in this paper are evaluated on the basis of the rational independence of atomic coordinates previously described (Hauptman & Karle, 1957). The significance of (3·6) is due to the fact that in evaluating (3·2) from (3·3) and (3·4) it is not necessary to obtain an explicit expression for $p(\xi_0, \xi_1, \ldots, \xi_m)$. It is sufficient merely to evaluate the averages in (3·6) where, for space group $P\bar{1}$, the ξ_ν are given by (3·7). For the other space groups the appropriate functional forms for ξ_ν are well known.

An alternative evaluation of the q function may be given by means of the Bessel function expansion of the exponential which appears in (3·3) (Watson, 1945, p. 22). In this way we may check the q function as obtained by means of the mixed moments (3·6).

3·1. *The distribution* $P(X_0, X_1, X_2)$ *of* $E_{\mathbf{k}}, E_{\mathbf{h}_1+\mathbf{k}}, E_{\mathbf{h}_2+\mathbf{k}}$

As a typical illustration of the procedure we derive next the joint probability distribution $P(X_0, X_1, X_2)$ of the three normalized structure factors $E_{\mathbf{k}}, E_{\mathbf{h}_1+\mathbf{k}}, E_{\mathbf{h}_2+\mathbf{k}}$, where $\mathbf{h}_1$ and $\mathbf{h}_2$ are fixed vectors and $\mathbf{k}$ ranges uniformly over all the vectors in reciprocal space.

First the 49 non-vanishing mixed moments $\mu_{\lambda_0\lambda_1\lambda_2}$, $2 \le \lambda_0+\lambda_1+\lambda_2 \le 6$, are computed. We find, for example,

$$\mu_{312} = 2^6\langle\cos^3 2\pi\mathbf{k}.\mathbf{r}_j \cos 2\pi(\mathbf{h}_1+\mathbf{k}).\mathbf{r}_j \times\cos^2 2\pi(\mathbf{h}_2+\mathbf{k}).\mathbf{r}_j\rangle_{\mathbf{k}} \,. \qquad (3·1·1)$$

$$\mu_{312} = 12\cos 2\pi\mathbf{h}_1.\mathbf{r}_j + 6\cos 2\pi(\mathbf{h}_1-2\mathbf{h}_2).\mathbf{r}_j + 2\cos 2\pi(\mathbf{h}_1+2\mathbf{h}_2).\mathbf{r}_j \,. \qquad (3·1·2)$$

Substituting the values of the 49 mixed moments into (3·4) we obtain $q(x_0, x_1, x_2)$, from which we find, after a tedious computation, $\log q(x_0, x_1, x_2)$ by means of the Maclaurin expansion

$$\log(1+x) = x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \ldots \,. \qquad (3·1·3$$

We then use

$$\sum_{j=1}^{N/2} \log q(x_0, x_1, x_2) = \log \prod_{j=1}^{N/2} q(x_0, x_1, x_2) \quad (3\cdot1\cdot4)$$

and

$$\prod_{j=1}^{N/2} q(x_0, x_1, x_2) = \exp\left(\log \prod_{j=1}^{N/2} q(x_0, x_1, x_2)\right) \quad (3\cdot1\cdot5)$$

to obtain the following expression for $\prod_{j=1}^{N/2} q(x_0, x_1, x_2)$:

$$\begin{aligned}
\prod_{j=1}^{N/2} & q(x_0, x_1, x_2) = \\
& \exp\left(-\tfrac{1}{2}x_0^2-\tfrac{1}{2}x_1^2-\tfrac{1}{2}x_2^2-\frac{E_{\mathbf{h}_1}}{N^{\frac{1}{2}}}x_0x_1-\frac{E_{\mathbf{h}_2}}{N^{\frac{1}{2}}}x_0x_2-\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{N^{\frac{1}{2}}}x_1x_2\right) \\
& \times\left\{1-\frac{1}{8N}(x_0^4+x_1^4+x_2^4)-\frac{1}{2N}(x_0^2x_1^2+x_0^2x_2^2+x_1^2x_2^2)\right. \\
& -\frac{E_{\mathbf{h}_1}}{2N^{3/2}}(x_0^3x_1+x_0x_1^3)-\frac{E_{\mathbf{h}_2}}{2N^{3/2}}(x_0^3x_2+x_0x_2^3)-\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{2N^{3/2}} \\
& \times(x_1^3x_2+x_1x_2^3)-\frac{E_{2\mathbf{h}_1}}{4N^{3/2}}x_0^2x_1^2-\frac{E_{2\mathbf{h}_2}}{4N^{3/2}}x_0^2x_2^2-\frac{E_{2\mathbf{h}_1-2\mathbf{h}_2}}{4N^{3/2}}x_1^2x_2^2 \\
& -\left(\frac{E_{\mathbf{h}_1+\mathbf{h}_2}}{2N^{3/2}}+\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{N^{3/2}}\right)x_0^2x_1x_2-\left(\frac{E_{2\mathbf{h}_1-\mathbf{h}_2}}{2N^{3/2}}+\frac{E_{\mathbf{h}_2}}{N^{3/2}}\right)x_0x_1^2x_2 \\
& -\left(\frac{E_{\mathbf{h}_1-2\mathbf{h}_2}}{2N^{3/2}}+\frac{E_{\mathbf{h}_1}}{N^{3/2}}\right)x_0x_1x_2^2-\frac{1}{18N^2}(x_0^6+x_1^6+x_2^6) \\
& -\frac{1}{2N^2}(x_0^4x_1^2+x_0^2x_1^4+x_0^4x_2^2+x_0^2x_2^4+x_1^4x_2^2+x_1^2x_2^4)-\frac{2}{N^2}x_0^2x_1^2x_2^2 \\
& +\frac{1}{128N^2}(x_0^8+x_1^8+x_2^8) \\
& +\frac{1}{16N^2}(x_0^6x_1^2+x_0^2x_1^6+x_0^6x_2^2+x_0^2x_2^6+x_1^6x_2^2+x_1^2x_2^6) \\
& +\frac{9}{64N^2}(x_0^4x_1^4+x_0^4x_2^4+x_1^4x_2^4) \\
& \left.+\frac{5}{16N^2}(x_0^4x_1^2x_2^2+x_0^2x_1^4x_2^2+x_0^2x_1^2x_2^4)+\dots\right\}. \qquad (3\cdot1\cdot6)
\end{aligned}$$

Finally, substituting from (3·1·6) into (3·2), we obtain, after an extremely lengthy analysis, the desired probability distribution:

$$\begin{aligned}
P(X_0, X_1, X_2) = & \frac{1}{(2\pi)^{3/2}D^{1/2}}\exp\left(-\tfrac{1}{2}\sum_{i,j=1}^{3}A_{ij}X_{i-1}X_{j-1}\right) \\
& \times\left\{1-\frac{1}{8N}[21-14(X_0^2+X_1^2+X_2^2)+(X_0^4+X_1^4+X_2^4\right. \\
& \qquad +4X_0^2X_1^2+4X_0^2X_2^2+4X_1^2X_2^2] \\
& -\frac{1}{4N^{3/2}}[(E_{2\mathbf{h}_1}+E_{2\mathbf{h}_2}+E_{2\mathbf{h}_1-2\mathbf{h}_2})-(E_{2\mathbf{h}_1}+E_{2\mathbf{h}_2})X_0^2 \\
& -(E_{2\mathbf{h}_1}+E_{2\mathbf{h}_1-2\mathbf{h}_2})X_1^2-(E_{2\mathbf{h}_2}+E_{2\mathbf{h}_1-2\mathbf{h}_2})X_2^2 \\
& +(20E_{\mathbf{h}_1}-2E_{\mathbf{h}_1-2\mathbf{h}_2})X_0X_1+(20E_{\mathbf{h}_2}-2E_{2\mathbf{h}_1-\mathbf{h}_2})X_0X_2 \\
& +(20E_{\mathbf{h}_1-\mathbf{h}_2}-2E_{\mathbf{h}_1+\mathbf{h}_2})X_1X_2-4E_{\mathbf{h}_1}(X_0^3X_1+X_0X_1^3) \\
& -4E_{\mathbf{h}_2}(X_0^3X_2+X_0X_2^3)-4E_{\mathbf{h}_1-\mathbf{h}_2}(X_1^3X_2+X_1X_2^3) \\
& +E_{2\mathbf{h}_1}X_0^2X_1^2+E_{2\mathbf{h}_2}X_0^2X_2^2+E_{2\mathbf{h}_1-2\mathbf{h}_2}X_1^2X_2^2 \\
& -(4E_{\mathbf{h}_1-\mathbf{h}_2}-2E_{\mathbf{h}_1+\mathbf{h}_2})X_0^2X_1X_2-(4E_{\mathbf{h}_2}-2E_{2\mathbf{h}_1-\mathbf{h}_2}) \\
& \times X_0X_1^2X_2-(4E_{\mathbf{h}_1}-2E_{\mathbf{h}_1-2\mathbf{h}_2})X_0X_1X_2^2] \\
& +\frac{1}{N^2}[\tfrac{153}{128}-\tfrac{1}{2}(E_{\mathbf{h}_1}^2+E_{\mathbf{h}_2}^2+E_{\mathbf{h}_1-\mathbf{h}_2}^2) \\
& \qquad +\tfrac{1}{2}(E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2}+E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2}+E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}) \\
& -(\tfrac{195}{32}-3(E_{\mathbf{h}_1}^2+E_{\mathbf{h}_2}^2)+\tfrac{1}{2}(E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2}+E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2} \\
& \qquad +E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}))X_0^2 \\
& -(\tfrac{195}{32}-3(E_{\mathbf{h}_1}^2+E_{\mathbf{h}_1-\mathbf{h}_2}^2)+\tfrac{1}{2}(E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad +E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2}+E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}))X_1^2 \\
& -(\tfrac{195}{32}-3(E_{\mathbf{h}_2}^2+E_{\mathbf{h}_1-\mathbf{h}_2}^2)+\tfrac{1}{2}(E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad +E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2}+E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}))X_2^2 \\
& +(\tfrac{9}{2}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}-\tfrac{3}{2}E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}-\tfrac{3}{2}E_{\mathbf{h}_1-\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2} \\
& \qquad -\tfrac{1}{2}E_{\mathbf{h}_1}(4E_{2\mathbf{h}_1}+E_{2\mathbf{h}_2}+E_{2\mathbf{h}_1-2\mathbf{h}_2}))X_0X_1 \\
& +(\tfrac{9}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_1-\mathbf{h}_2}-\tfrac{3}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_1+\mathbf{h}_2}-\tfrac{3}{2}E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad -\tfrac{1}{2}E_{\mathbf{h}_2}(E_{2\mathbf{h}_1}+4E_{2\mathbf{h}_2}+E_{2\mathbf{h}_1-2\mathbf{h}_2}))X_0X_2 \\
& +(\tfrac{9}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_2}-\tfrac{3}{2}E_{\mathbf{h}_1}E_{2\mathbf{h}_1-\mathbf{h}_2}-\tfrac{3}{2}E_{\mathbf{h}_2}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad -\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}(E_{2\mathbf{h}_1}+E_{2\mathbf{h}_2}+4E_{2\mathbf{h}_1-2\mathbf{h}_2}))X_1X_2 \\
& +(\tfrac{545}{192}-\tfrac{1}{2}(E_{\mathbf{h}_1}^2+E_{\mathbf{h}_2}^2))X_0^4+(\tfrac{545}{192}-\tfrac{1}{2}(E_{\mathbf{h}_1}^2+E_{\mathbf{h}_1-\mathbf{h}_2}^2))X_1^4 \\
& \qquad +(\tfrac{545}{192}-\tfrac{1}{2}(E_{\mathbf{h}_2}^2+E_{\mathbf{h}_1-\mathbf{h}_2}^2))X_2^4 \\
& +(\tfrac{1}{2}E_{\mathbf{h}_1}E_{2\mathbf{h}_1}+E_{\mathbf{h}_2}(\tfrac{1}{2}E_{\mathbf{h}_1+\mathbf{h}_2}-E_{\mathbf{h}_1-\mathbf{h}_2}))X_0^3X_1 \\
& +(\tfrac{1}{2}E_{\mathbf{h}_1}E_{2\mathbf{h}_1}+E_{\mathbf{h}_1-\mathbf{h}_2}(\tfrac{1}{2}E_{2\mathbf{h}_1-\mathbf{h}_2}-E_{\mathbf{h}_2}))X_0X_1^3 \\
& +(\tfrac{1}{2}E_{\mathbf{h}_2}E_{2\mathbf{h}_2}+E_{\mathbf{h}_1}(\tfrac{1}{2}E_{\mathbf{h}_1+\mathbf{h}_2}-E_{\mathbf{h}_1-\mathbf{h}_2}))X_0^3X_2 \\
& +(\tfrac{1}{2}E_{\mathbf{h}_2}E_{2\mathbf{h}_2}+E_{\mathbf{h}_1-\mathbf{h}_2}(\tfrac{1}{2}E_{\mathbf{h}_1-2\mathbf{h}_2}-E_{\mathbf{h}_1}))X_0X_2^3 \\
& +(\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}E_{2\mathbf{h}_1-2\mathbf{h}_2}+E_{\mathbf{h}_1}(\tfrac{1}{2}E_{2\mathbf{h}_1-\mathbf{h}_2}-E_{\mathbf{h}_2}))X_1^3X_2 \\
& +(\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}E_{2\mathbf{h}_1-2\mathbf{h}_2}+E_{\mathbf{h}_2}(\tfrac{1}{2}E_{\mathbf{h}_1-2\mathbf{h}_2}-E_{\mathbf{h}_1}))X_1X_2^3 \\
& +(\tfrac{59}{8}-\tfrac{5}{2}E_{\mathbf{h}_1}^2-\tfrac{1}{2}E_{\mathbf{h}_2}^2-\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}^2+\tfrac{1}{2}E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2} \\
& \qquad +\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2})X_0^2X_1^2 \\
& +(\tfrac{59}{8}-\tfrac{1}{2}E_{\mathbf{h}_1}^2-\tfrac{5}{2}E_{\mathbf{h}_2}^2-\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}^2+\tfrac{1}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad +\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2})X_0^2X_2^2 \\
& +(\tfrac{59}{8}-\tfrac{1}{2}E_{\mathbf{h}_1}^2-\tfrac{1}{2}E_{\mathbf{h}_2}^2-\tfrac{5}{2}E_{\mathbf{h}_1-\mathbf{h}_2}^2+\tfrac{1}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad +\tfrac{1}{2}E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2})X_1^2X_2^2 \\
& +(-\tfrac{5}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_2}+E_{\mathbf{h}_1}E_{2\mathbf{h}_1-\mathbf{h}_2}+E_{\mathbf{h}_2}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad +\tfrac{1}{2}E_{\mathbf{h}_1-\mathbf{h}_2}(E_{2\mathbf{h}_1}+E_{2\mathbf{h}_2}))X_0^2X_1X_2 \\
& +(-\tfrac{5}{2}E_{\mathbf{h}_1}E_{\mathbf{h}_1-\mathbf{h}_2}+E_{\mathbf{h}_1}E_{\mathbf{h}_1+\mathbf{h}_2}+E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1-2\mathbf{h}_2} \\
& \qquad +\tfrac{1}{2}E_{\mathbf{h}_2}(E_{2\mathbf{h}_1}+E_{2\mathbf{h}_1-2\mathbf{h}_2}))X_0X_1^2X_2 \\
& +(-\tfrac{5}{2}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}+E_{\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2}+E_{\mathbf{h}_1-\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2} \\
& \qquad +\tfrac{1}{2}E_{\mathbf{h}_1}(E_{2\mathbf{h}_2}+E_{2\mathbf{h}_1-2\mathbf{h}_2}))X_0X_1X_2^2
\end{aligned}$$

$$-\tfrac{83}{288}(X_0^6+X_1^6+X_2^6)-\tfrac{51}{32}(X_0^4X_1^2+X_0^2X_1^4+X_0^4X_2^2+X_0^2X_2^4+X_1^4X_2^2+X_1^2X_2^4)-\tfrac{29}{8}X_0^2X_1^2X_2^2$$
$$+\tfrac{1}{128}(X_0^8+X_1^8+X_2^8)$$
$$+\tfrac{1}{16}(X_0^6X_1^2+X_0^2X_1^6+X_0^6X_2^2+X_0^2X_2^6+X_1^6X_2^2+X_1^2X_2^6)$$
$$+\tfrac{9}{64}(X_0^4X_1^4+X_0^4X_2^4+X_1^4X_2^4)$$
$$+\tfrac{5}{16}(X_0^4X_1^2X_2^2+X_0^2X_1^4X_2^2+X_0^2X_1^2X_2^4)]+\dots, \quad (3\cdot1\cdot7)$$

where

$$D = 1-\frac{1}{N}(E_{\mathbf{h}_1}^2+E_{\mathbf{h}_2}^2+E_{\mathbf{h}_1-\mathbf{h}_2}^2)+\frac{2}{N^{3/2}}E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2},$$

$$A_{11} = \frac{1}{D}\left(1-\frac{1}{N}E_{\mathbf{h}_1-\mathbf{h}_2}^2\right),$$

$$A_{12} = A_{21} = \frac{1}{D}\left(-\frac{E_{\mathbf{h}_1}}{N^{1/2}}+\frac{E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}}{N}\right),$$

$$A_{13} = A_{31} = \frac{1}{D}\left(-\frac{E_{\mathbf{h}_2}}{N^{1/2}}+\frac{E_{\mathbf{h}_1}E_{\mathbf{h}_1-\mathbf{h}_2}}{N}\right),$$

$$A_{22} = \frac{1}{D}\left(1-\frac{1}{N}E_{\mathbf{h}_2}^2\right),$$

$$A_{23} = A_{32} = \frac{1}{D}\left(-\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{N^{1/2}}+\frac{E_{\mathbf{h}_1}E_{\mathbf{h}_2}}{N}\right),$$

$$A_{33} = \frac{1}{D}\left(1-\frac{1}{N}E_{\mathbf{h}_1}^2\right).$$

The remaining probability distributions are derived in a similar way and are listed without further proof.

3·2. *The probability distribution $P(X_0)$ of $E_\mathbf{k}$*

$$P(X_0) = \frac{1}{(2\pi)^{\frac{1}{2}}}\exp(-\tfrac{1}{2}X_0^2)\left\{1-\frac{1}{8N}(3-6X_0^2+X_0^4)-\frac{1}{1152N^2}(15+900X_0^2-930X_0^4+188X_0^6-9X_0^8)+\dots\right\}. \quad (3\cdot2\cdot1)$$

3·3. *The probability distribution $P(X_0, X_1)$ of the pair $E_\mathbf{k}, E_{\mathbf{h}+\mathbf{k}}$ where* **h** *is fixed*

$$P(X_0, X_1) = \frac{1}{2\pi(1-E_\mathbf{h}^2/N)^{\frac{1}{2}}}\times\exp\left(-\frac{1}{2(1-E_\mathbf{h}^2/N)}\left(X_0^2-2\frac{E_\mathbf{h}}{N^{\frac{1}{2}}}X_0X_1+X_1^2\right)\right)\left\{1\right.$$
$$-\frac{1}{N}[\tfrac{5}{4}-\tfrac{5}{4}(X_0^2+X_1^2)+\tfrac{1}{8}(X_0^4+4X_0^2X_1^2+X_1^4)]$$
$$-\frac{1}{N^{3/2}}\left[\frac{E_{2\mathbf{h}}}{4}(1-X_0^2-X_1^2+X_0^2X_1^2)+E_\mathbf{h}(4X_0X_1-X_0^3X_1-X_0X_1^3)\right]$$
$$-\frac{1}{N^2}\left[\left(\frac{E_\mathbf{h}^2}{2}-\tfrac{11}{96}\right)-(\tfrac{5}{2}E_\mathbf{h}^2-\tfrac{41}{16})(X_0^2+X_1^2)\right.$$
$$+(\tfrac{1}{2}E_\mathbf{h}^2-\tfrac{5}{3})(X_0^4+X_1^4)+(\tfrac{5}{2}E_\mathbf{h}^2-\tfrac{75}{16})X_0^2X_1^2$$
$$+\tfrac{65}{288}(X_0^6+X_1^6)+\tfrac{41}{32}(X_0^4X_1^2+X_0^2X_1^4)-\tfrac{1}{128}(X_0^8+X_1^8)$$
$$-\tfrac{1}{16}(X_0^6X_1^2+X_0^2X_1^6)-\tfrac{9}{64}X_0^4X_1^4$$
$$\left.\left.+E_\mathbf{h}E_{2\mathbf{h}}(2X_0X_1-\tfrac{1}{2}X_0^3X_1-\tfrac{1}{2}X_0X_1^3)\right]+\dots\right\}. \quad (3\cdot3\cdot1)$$

3·4. *The probability distribution $P'(X_0, X_1)$ of the pair $E_\mathbf{k}, E_{\mathbf{h}+2\mathbf{k}}$ where* **h** *is fixed*

$$P'(X_0, X_1) = \frac{1}{2\pi}\exp(-\tfrac{1}{2}X_0^2-\tfrac{1}{2}X_1^2)\left\{1-\frac{E_{2\mathbf{h}}}{2N}(X_1-X_0^2X_1)-\frac{1}{8N}(6-6X_0^2-6X_1^2+X_0^4+X_1^4)+\dots\right\}. \quad (3\cdot4\cdot1)$$

4. The average values

4·1. *The average value of $|E_\mathbf{k}|^p$*

In order to compute the expected or average value of $|E_\mathbf{k}|^p$ we employ

$$\langle|E_\mathbf{k}|^p\rangle_\mathbf{k} = \int_{-\infty}^{\infty}|X_0|^pP(X_0)dX_0, \quad (4\cdot1\cdot1)$$

where $P(X_0)$ is given by (3·2·1). In this way we derive (2·1·1).

4·2. *The average value of $|E_\mathbf{k}|^p|E_{\mathbf{h}+\mathbf{k}}|^q$*

Now we use

$$\langle|E_\mathbf{k}|^p|E_{\mathbf{h}+\mathbf{k}}|^q\rangle_\mathbf{k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|X_0|^p|X_1|^qP(X_0, X_1)dX_0dX_1, \quad (4\cdot2\cdot1)$$

where $P(X_0, X_1)$ is given by (3·3·1). Evaluation of the resulting double integrals finally yields the desired formula:

$$\langle|E_\mathbf{k}|^p|E_{\mathbf{h}+\mathbf{k}}|^q\rangle_\mathbf{k} = \frac{2^{(p+q+2)/2}}{2\pi}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\left\{1\right.$$
$$-\frac{1}{8N}[p(p-2)+q(q-2)-4pq(E_\mathbf{h}^2-1)]-\frac{E_{2\mathbf{h}}}{4N^{3/2}}pq$$
$$+\frac{1}{N^2}\left[\tfrac{1}{24}pq(p-2)(q-2)E_\mathbf{h}^4\right.$$
$$-\tfrac{1}{16}pq(p^2+4pq+q^2-6p-6q)E_\mathbf{h}^2$$
$$+\tfrac{1}{1152}p(p-2)(p-4)(9p+10)$$
$$+\tfrac{1}{1152}q(q-2)(q-4)(9q+10)$$
$$\left.\left.+\tfrac{1}{64}pq(4p^2+9pq+4q^2-10p-10q-28)\right]+\dots\right\}. \quad (4\cdot2\cdot2)$$

4·3. *The average value of $|E_\mathbf{k}|^p|E_{\mathbf{h}_1+\mathbf{k}}|^q|E_{\mathbf{h}_2+\mathbf{k}}|^r$*

Employing

$$\langle|E_\mathbf{k}|^p|E_{\mathbf{h}_1+\mathbf{k}}|^q|E_{\mathbf{h}_2+\mathbf{h}}|^r\rangle_\mathbf{k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|X_0|^p|X_1|^q|X_2|^rP(X_0, X_1, X_2)dX_0dX_1dX_2, \quad (4\cdot3\cdot1)$$

where $P(X_0, X_1, X_2)$ is given by (3·1·7), and evaluating the resulting triple integrals, we finally obtain after a very lengthy computation

$$\begin{aligned}\langle|E_{\mathbf{k}}|^p|E_{\mathbf{h}_1+\mathbf{k}}|^q|E_{\mathbf{h}_2+\mathbf{k}}|^r\rangle_{\mathbf{k}} &= \frac{2^{(p+q+r+3)/2}}{(2\pi)^{3/2}}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{r+1}{2}\right)\Big\{1\\ &-\frac{1}{8N}[p(p-2)+q(q-2)+r(r-2)-4pq(E^2_{\mathbf{h}_1}-1)\\ &\quad-4pr(E^2_{\mathbf{h}_2}-1)-4qr(E^2_{\mathbf{h}_1-\mathbf{h}_2}-1)]\\ &-\frac{1}{4N^{3/2}}[pqE_{2\mathbf{h}_1}+prE_{2\mathbf{h}_2}+qrE_{2\mathbf{h}_1-2\mathbf{h}_2}\\ &\quad-4pqrE_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}]\\ &+\frac{1}{N^2}\Big[\tfrac{1}{24}pq(p-2)(q-2)E^4_{\mathbf{h}_1}+\tfrac{1}{24}pr(p-2)(r-2)E^4_{\mathbf{h}_2}\\ &+\tfrac{1}{24}qr(q-2)(r-2)E^4_{\mathbf{h}_1-\mathbf{h}_2}+\tfrac{1}{4}pqr(p-2)E^2_{\mathbf{h}_1}E^2_{\mathbf{h}_2}\\ &+\tfrac{1}{4}pqr(q-2)E^2_{\mathbf{h}_1}E^2_{\mathbf{h}_1-\mathbf{h}_2}+\tfrac{1}{4}pqr(r-2)E^2_{\mathbf{h}_2}E^2_{\mathbf{h}_1-\mathbf{h}_2}\\ &-\tfrac{1}{16}pq(p^2+q^2+r^2+4pq+4pr+4qr-6p-6q-2r)E^2_{\mathbf{h}_1}\\ &-\tfrac{1}{16}pr(p^2+q^2+r^2+4pq+4pr+4qr-6p-2q-6r)E^2_{\mathbf{h}_2}\\ &-\tfrac{1}{16}qr(p^2+q^2+r^2+4pq+4pr+4qr\\ &\quad-2p-6q-6r)E^2_{\mathbf{h}_1-\mathbf{h}_2}\\ &+\tfrac{1}{1152}p(p-2)(p-4)(9p+10)\\ &+\tfrac{1}{1152}q(q-2)(q-4)(9q+10)\\ &+\tfrac{1}{1152}r(r-2)(r-4)(9r+10)\\ &+\tfrac{1}{64}pq(4p^2+9pq+4q^2-10p-10q-28)\\ &+\tfrac{1}{64}pr(4p^2+9pr+4r^2-10p-10r-28)\\ &+\tfrac{1}{64}qr(4q^2+9qr+4r^2-10q-10r-28)\\ &+\tfrac{1}{16}pqr(5p+5q+5r+2)\\ &-\tfrac{1}{2}pqr(E_{\mathbf{h}_1}E_{\mathbf{h}_1-2\mathbf{h}_2}+E_{\mathbf{h}_2}E_{2\mathbf{h}_1-\mathbf{h}_2}\\ &\quad+E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1+\mathbf{h}_2})\Big]+\ldots\Big\}. \end{aligned} \quad (4\cdot3\cdot2)$$

4·4. *The average value of* $E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}|E_{\mathbf{k}}|^p|E_{\mathbf{h}+\mathbf{k}}|^q$

In order to obtain this average we make use of

$$\langle E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}|E_{\mathbf{k}}|^p|E_{\mathbf{h}+\mathbf{k}}|^q\rangle_{\mathbf{k}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}X_0X_1|X_0|^p|X_1|^qP(X_0, X_1)\,dX_0dX_1\,, \quad (4\cdot4\cdot1)$$

where $P(X_0, X_1)$ is given by (3·3·1), and find

$$\begin{aligned}\langle E_{\mathbf{k}}E_{\mathbf{h}+\mathbf{k}}|E_{\mathbf{k}}|^p|E_{\mathbf{h}+\mathbf{k}}|^q\rangle_{\mathbf{k}} &= \frac{2^{(p+q+6)/2}}{2\pi}\Gamma\left(\frac{p+3}{2}\right)\Gamma\left(\frac{q+3}{2}\right)\frac{E_{\mathbf{h}}}{N^{\frac{1}{2}}}\Big\{1\\ &-\frac{1}{8N}[p(p-2)+q(q-2)-4(\tfrac{1}{3}E^2_{\mathbf{h}}-1)pq]+\ldots\Big\}.\end{aligned} \quad (4\cdot4\cdot2)$$

4·5. *The average value of*

$$|E_{\mathbf{k}}|^p|E_{\mathbf{h}_1+\mathbf{k}}|^q|E_{\mathbf{h}_2+\mathbf{k}}|^rE_{\mathbf{h}_1+\mathbf{k}}E_{\mathbf{h}_2+\mathbf{k}}$$

This average value is obtained from (3·1·7) by means of the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|X_0|^p|X_1|^q|X_2|^rX_1X_2 \times P(X_0, X_1, X_2)\,dX_0dX_1dX_2\,, \quad (4\cdot5\cdot1)$$

and is found to be

$$\begin{aligned}\langle|E_{\mathbf{k}}|^p|E_{\mathbf{h}_1+\mathbf{k}}|^q|E_{\mathbf{h}_2+\mathbf{k}}|^rE_{\mathbf{h}_1+\mathbf{k}}E_{\mathbf{h}_2+\mathbf{k}}\rangle_{\mathbf{k}} &= \frac{2^{(p+q+r+7)/2}}{(2\pi)^{3/2}}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+3}{2}\right)\Gamma\left(\frac{r+3}{2}\right)\Big\{\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{N^{\frac{1}{2}}}\\ &+\frac{pE_{\mathbf{h}_1}E_{\mathbf{h}_2}}{N}-\frac{pE_{\mathbf{h}_1+\mathbf{h}_2}}{2N^{3/2}}\\ &+\frac{E_{\mathbf{h}_1-\mathbf{h}_2}}{N^{3/2}}[\tfrac{1}{2}(pq+qr+p+q)E^2_{\mathbf{h}_1}+\tfrac{1}{2}(pr+qr+q+r)E^2_{\mathbf{h}_2}\\ &+\tfrac{1}{6}(4qr+3q+3r)E^2_{\mathbf{h}_1-\mathbf{h}_2}-\tfrac{1}{8}(p^2+q^2+r^2\\ &+4pq+4pr+4qr+6p+2q+2r)]+\ldots\Big\}.\end{aligned} \quad (4\cdot5\cdot2)$$

4·6. *The average value of* $|E_{\mathbf{k}}|^p|E_{\mathbf{h}+2\mathbf{k}}|^qE_{\mathbf{h}+2\mathbf{k}}$

We obtain this average in the usual way by means of $P'(X_0, X_1)$, as given by (3·4·1), and find

$$\langle|E_{\mathbf{k}}|^p|E_{\mathbf{h}+2\mathbf{k}}|^qE_{\mathbf{h}+2\mathbf{k}}\rangle_{\mathbf{k}} = \frac{E_{\mathbf{h}}}{4\pi N}2^{(p+q+4)/2}p\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+3}{2}\right)+\ldots. \quad (4\cdot6\cdot1)$$

5. Proof of the basic formulas

Equation (2·1·2) is proved by means of

$$\langle(|E_{\mathbf{k}}|^p-\overline{|E|^p})(|E_{\mathbf{h}+\mathbf{k}}|^q-\overline{|E|^q})\rangle_{\mathbf{k}} = \langle|E_{\mathbf{k}}|^p|E_{\mathbf{h}+\mathbf{k}}|^q\rangle_{\mathbf{k}}-\langle|E_{\mathbf{k}}|^p\rangle\langle|E_{\mathbf{k}}|^q\rangle\,, \quad (5\cdot1)$$

where $\overline{|E|^p}$ means $\langle|E_{\mathbf{k}}|^p\rangle_{\mathbf{k}}$. Employing (4·2·2) and (2·1·1) to compute the right side of (5·1), and solving for $E_{2\mathbf{h}}$, we obtain (2·1·2). Equation (2·1·3) is obtained in the same way by making use of (4·3·2), (4·2·2), and (2·1·1). Again, (2·1·4) is an immediate consequence of (4·6·1) and (2·1·1). The theoretical value for the denominator of (2·1·4) has been replaced (using (2·1·1)) by the average there shown in order to permit averaging over the restricted values of $\mathbf{k}$ for which the signs of the $E_{\mathbf{h}+2\mathbf{k}}$ are known. Since these $|E_{\mathbf{h}+2\mathbf{k}}|$ are generally large, and always greater than zero, the possibility exists for using negative values of q (greater than -1) in (2·1·4). In a similar way, (2·1·5) follows from (4·5·2) and (2·1·1). Finally, (2·1·6) is an immediate consequence of (4·4·2) and (2·1·1).

Equation (2·1·3*a*) is obtained from (2·1·3) by means of the substitution $\mathbf{h} = \mathbf{h}_1 = -\mathbf{h}_2$, while (2·1·4*a*) follows from (2·1·4) by replacing $\mathbf{h}$ by $2\mathbf{h}$ and $\mathbf{k}$ by $-\mathbf{h}-\mathbf{k}$.

6. Proof of the integrated formulas

The basic formulas are valid for every set of non-negative values for p, q, r. It is to be expected that the use of several different sets of values for p, q, r would in general be better than the use of a single set of values. However, using many different sets of values for p, q, r would result in a formidable computing problem. Therefore, a great advantage accrues by integrating the basic formulas for values of p, q, r from 0 to t. The reason for this is that each integrated formula is equivalent to computing the related basic formula over all possible combinations of values of p, q, r in the range from 0 to t. It is advantageous to use as large a value of t as the accuracy of the experimental data warrants.

In order to prove the typical integrated formula (2·2·2) we multiply (2·1·2) by

$$2^{(p+q+2)/2}pq\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right),$$

integrate between limits 0 to t for p and q (using $\int a^x dx = a^x/\log a$), and solve for $E_{2\mathbf{h}}$. The remaining integrated formulas are derived in the same way, where reference should be made to Table 1.

7. An important identity

In space group $P1$ a simple identity among the structure invariants (equation (4·13), Karle & Hauptman, 1957) had important application in phase determination. Here, too, the corresponding identity is found to have important practical application.

We consider the four structure invariants

$$\left.\begin{aligned} b_1 &= \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_1-\mathbf{h}_2}, \\ b_2 &= \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_3} + \varphi_{\mathbf{h}_1-\mathbf{h}_3}, \\ b_3 &= \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_3} + \varphi_{\mathbf{h}_2-\mathbf{h}_3}, \\ b_4 &= \varphi_{\mathbf{h}_1-\mathbf{h}_2} + \varphi_{\mathbf{h}_1-\mathbf{h}_3} + \varphi_{\mathbf{h}_2-\mathbf{h}_3}, \end{aligned}\right\} \quad (7{\cdot}1)$$

where, for example, $\varphi_{\mathbf{h}_1}$ is the phase (either 0 or π) of the normalized structure factor $E_{\mathbf{h}_1}$. It is then easily verified that

$$b_1+b_2+b_3+b_4 \equiv 0 \quad (7{\cdot}2)$$

provided that the sum is reduced modulo 2π. In terms of the normalized structure factors the relation (7·2) becomes (replacing $\mathbf{h}_3$ by $\mathbf{k}'$)

$$E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2} \equiv \frac{(E_{\mathbf{h}_1}E_{\mathbf{k}'}E_{\mathbf{h}_1-\mathbf{k}'})(E_{\mathbf{h}_2}E_{\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'})(E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'})}{E^2_{\mathbf{k}'}E^2_{\mathbf{h}_1-\mathbf{k}'}E^2_{\mathbf{h}_2-\mathbf{k}'}}. \quad (7{\cdot}3)$$

Summing (7·3) over all vectors $\mathbf{k}'$, we obtain the important identity

$$E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2} \equiv \frac{\sum_{\mathbf{k}'} (E_{\mathbf{h}_1}E_{\mathbf{k}'}E_{\mathbf{h}_1-\mathbf{k}'})(E_{\mathbf{h}_2}E_{\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'})(E_{\mathbf{h}_1-\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{k}'}E_{\mathbf{h}_2-\mathbf{k}'})}{\sum_{\mathbf{k}'} E^2_{\mathbf{k}'}E^2_{\mathbf{h}_1-\mathbf{k}'}E^2_{\mathbf{h}_2-\mathbf{k}'}}, \quad (7{\cdot}4)$$

where it is to be emphasized that the three factors (in parentheses) in the numerator of (7·4) are independently computed by means of (2·2·3). Thus (7·4) is to be regarded as a valuable supplement to (2·2·3).

In the case that the structure consist of N identical point atoms and that the data are very extensive and accurate, (2·2·3) and (for each $\mathbf{k}'$) (7·3) will yield reliable values for $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}$. In practice, when the atoms are not all identical and the data are limited and subject to experimental errors, it is to be expected that (2·2·3) and (7·3) will yield values which are distributed about $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}$. In this case the fundamental identity (7·4) will give an improved value for $E_{\mathbf{h}_1}E_{\mathbf{h}_2}E_{\mathbf{h}_1-\mathbf{h}_2}$.

8. Procedure for phase determination

The integrated formulas (2·2·1)–(2·2·6) and (7·4) are an improvement over the corresponding basic formulas (2·1·1)–(2·1·6). They provide the basis for a procedure for determining the value of every phase when only the magnitudes of the structure factors are known. However, if desired, the basic formulas may be used instead of the integrated formulas. It is readily seen that the basic formulas include as special cases all previously known formulas for phase determination. Minor modifications required when the structure contains unequal atoms are suggested below.

Initially only formulas (2·2·2) and (2·2·3a) (or 7·4 with $\mathbf{h}_2 = -\mathbf{h}_1$) are available for determining phases, since these are the only ones which require a knowledge of the magnitudes, and not the signs, of the structure factors. Evidently (2·2·3a) and (7·4) are the most powerful formulas available for determining the signs of the $E_{2\mathbf{h}}$, especially if $E^2_{\mathbf{h}}$ is large. In (2·2·3a), as in the other formulas, t may be an arbitrarily chosen positive number. However, the larger the value of t the more reliable the resulting formula will be, assuming a sufficiently large number of perfectly accurate data. Therefore the value of t to be chosen in practice is the largest value compatible with the accuracy of the experimental data. It is suggested that with an average error of about 20% in the observed intensities, values of t equal to 3 or 4 may be suitable, whereas with an average error of 10% or less, $t = 5$ or 6 may be preferred. The value to be gained from more accurate data is clearly indicated by these formulas.

In (2·2·2) and (2·2·3a) as well as in the remaining phase-determining formulas, the averages M, $|E|^p$, etc. are those computed from the available experimental data rather than the theoretical averages given by (2·2·1) and (2·1·1). Small differences between the theoretical and computed averages, owing to the limited number of data, experimental errors, etc. are to be expected. It is anticipated that the use of the experimentally determined averages instead of the theoretical ones will compensate somewhat for the limitations inherent in the experimental data. As soon as the signs of several of the $E_{2\mathbf{h}}$ have been determined, (2·2·4*a*) and (2·2·6) may be used to determine the signs of others.

As is well known, the signs of three large structure factors, constituting a linearly independent triple modulo 2, are to be specified arbitrarily in order to fix the origin uniquely. Denoting by $E_{\mathbf{h}_1}$ any one of these three, (2·2·3), (7·4), and (2·2·5) are available for determining the sign of any structure factor $E_{\mathbf{h}_2}$ which is linearly dependent modulo 2 on $E_{\mathbf{h}_1}$. These formulas require a knowledge of the signs of many of the $E_{2\mathbf{h}}$, which signs will have already been obtained as indicated in the previous paragraph. In using (2·2·5) the averages are to be taken over all vectors $\mathbf{k}$ congruent modulo 2 to $\mathbf{h}_1$ or $\mathbf{h}_2$.

Once the signs of a sufficiently large number of structure factors, linearly dependent modulo 2 on one of the three large origin-determining structure factors, have thus been found, (2·2·6) will be useful in determining the signs of the remaining structure factors.

In the case that the structure consist of N atoms per unit cell not all identical, then N in (2·1·2)–(2·1·6) and (2·2·2) to (2·2·6) is to be replaced by σ_2^3/σ_3^2, where

$$\sigma_k = \sum_{j=1}^{N} Z_j^k \qquad (8\cdot1)$$

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. Equations (2·1·2)–(2·1·6) and (2·2·2)–(2·2·6) then no longer have exact, but merely probable, validity. However, this matter has already been discussed elsewhere (Karle & Hauptman, 1956) and we may conclude that, in general, for centrosymmetric structures this limitation will have an adverse effect on only some of the smaller E's. Thus the introduction of (8·1) together with the identity (7·4) makes unimportant the effect of unequal atoms in phase determination.

It is to be noted that (2·2·3) and the special case (2·2·3*a*) play a central role in the procedure for phase determination in that they are sufficient to determine the phases of all the structure factors. The remaining formulas constitute a valuable adjunct to these. In practice the identity (7·4) will be a useful supplement, especially in the case that the structure consists of unequal atoms.

9. Concluding remarks

In this paper joint probability distributions of one, two, and three structure factors have been obtained on the basis that the crystal structure is fixed and the indices range over vectors in reciprocal space. They have led to main formulas (2·1·3) and (2·2·3) as well as several auxiliary ones. These, together with the identity (7·4), yield an improved procedure for phase determination.

Joint probability distributions of four or more structure factors may be found by the methods described in this paper. They should lead to new phase-determining formulas whose usefulness is yet to be decided.

The mathematical analyses in this paper are extremely long and tedious and have been given here only in barest outline. We hope to present these details in a new edition of our Monograph I.

References

Bullough, R. K. & Cruickshank, D. W. J. (1955). *Acta Cryst.* **8**, 29.

Cochran, W. (1954). *Acta Cryst.* **7**, 581.

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

Hauptman, H. & Karle, J. (1957). *Acta Cryst.* **10**, 267.

Hughes, E. W. (1953). *Acta Cryst.* **6**, 871.

Karle, J .& Hauptman, H. (1953). *Acta Cryst.* **6**, 131.

Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635.

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

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