Laboratory who computed the three-dimensional Fourier syntheses and structure factors on 'Deuce', the N.P.L. electronic computer.

References

BUERGER, M. J. (1946). *J. Appl. Phys.* **17**, 579.
COCHRAN, W. (1951). *Acta Cryst.* **4**, 408.
COCHRAN, W. (1952). *Acta Cryst.* **5**, 65.
CORSMIT, A. F., SCHUYFF, A. & FEIL, D. (1956). *Proc. K. Ned. Akad. Wet.* B, **59**, 470.
CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65.
DONOHUE, J. D. (1950). *J. Amer. Chem. Soc.* **72**, 2701.
DUNITZ, J. D. & ROLLETT, J. S. (1956). *Acta Cryst.* **9**, 327.

HARKER, D. (1948). *Amer. Min.* **33**, 764.
HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70.
OUGHTON, B. M. & HARRISON, P. M. (1957). *Acta Cryst.* **10**, 479.
PAULING, L. (1940). *Nature of the Chemical Bond*, p. 176. Ithaca: Cornell University Press.
SAYRE, D. M. (1952). *Acta Cryst.* **5**, 60.
SCHMIDT, G. M. J., HODGKIN, D. C. & OUGHTON, B. M. (1957). *Biochem. J.* **65**, 744.
STEINRAUF, L. K. & JENSEN, L. H. (1956). *Acta Cryst.* **9**, 539.
WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 152.
YAKEL, H. L., JR. & HUGHES, E. W. (1954). *Acta Cryst.* **7**, 291.
ZACHARIASEN, W. H. (1952). *Acta Cryst.* **5**, 68.

A Unified Program for Phase Determination, Type 1P

BY J. KARLE AND H. HAUPTMAN

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

The new probability approach, in which the crystal structure is fixed and the Miller indices range uniformly but not independently over the integers, has been developed to yield formulas with universal application to all the space groups. The formulation is general and includes with equivalent rigor both the unequal and equal atom cases. If a requirement of rational independence of atomic coordinates is fulfilled, the formulas have exact validity. On the other hand, the formulas should still be useful even if the condition of rational independence is only partially fulfilled. In the application to a particular space group, use is made of the space group symmetries to obtain relationships among the structure factors, by means of which the general formula is suitably specialized.

The present paper is concerned with the space groups comprising type 1P. A detailed procedure for phase determination in this type is described.

1. Introduction

The concept of the joint probability distribution of several structure factors, obtained by treating the atomic coordinates as random variables, was introduced in our Monograph I (Hauptman & Karle, 1953) for the purpose of determining phase directly from the observed X-ray intensities. Independently, Kitaigorodsky (1954), derived the joint probability distribution for three structure factors which constituted a study of the Sayre relation (1952) from the probability point of view. The results obtained from these joint probability distributions were confirmed by Bertaut (1955) and Klug (1958) who employed alternative manipulative techniques which were however mathematically equivalent to the formulation in Monograph I (1953).* It is to be emphasized that all these distribu-

tions were obtained on the basis that **h** is fixed and the atomic coordinates range uniformly and independently throughout the asymmetric unit. This is in marked contrast to our more recent work (1958) which treats the crystal structure as fixed and permits the indices to range uniformly, but not independently, throughout reciprocal space. In this way, considerably improved formulas have been obtained.

—— · —— · —— · —— ·

variances of individual contributors to phase determining formulas. However in a previous paper by us (1956), of which Klug is apparently unaware, measures of statistical significance based on the appropriate variances had already been derived and their application was discussed.

The method of Monograph I utilizes the unusually large *E* values and their interactions in several phase determining formulas to build an internally consistent set of signs. Thus it should be clear that more is required in a theoretical evaluation than a knowledge of the variance of single terms in individual formulas and the number and quality of the data. Quite fortunately, phase determining theories can be readily confronted by experiment and it is our opinion that the final judgment regarding the applicability of a method is most propitiously made in the laboratory.

---

* In his paper, Klug (1958) summarizes the theory and attempts to evaluate the status of probability methods based on atomic coordinates as random variables. Klug reiterates limitations stated by others in the past and offers, as new evidence, measures of statistical significance derived from the

The joint probability distributions, based on the atomic coordinates as random variables, lead to formulas having certain limitations. In general, they have probable, rather than exact, validity. In addition, although formulas for all the centrosymmetric space groups are readily obtained, they often assume different forms for the various space groups. Hence it is difficult to adapt the procedure for use with automatic computing machinery. A major defect, however, of these probability distributions is that, in general, they do not lead to phase determining formulas in the non-centrosymmetric space groups.

In contrast to the procedures described in Monograph I (1953), it would appear that an ideal program for phase determination should embody the following criteria: (a) A unified program valid not only for all the space groups, non-centrosymmetric as well as centrosymmetric, but also readily adapted to modern computing techniques. (b) A uniform treatment regardless of whether the crystal contains atoms of the same or different atomic numbers. (c) Explicit formulas having exact validity.

It was with the intention to fulfill more nearly the above criteria that we derived improved formulas for the case of equal atoms in space groups $P\bar{1}$ and $P1$ by algebraic methods (1957). The algebraic methods lead to essentially the same formulas for all the space groups. By means of an analysis related to the superposition of Patterson maps, Vaughan (1958) independently derived asymptotic formulas which are approximately equivalent to the algebraically exact ones for large unit cells.

The results obtained by algebraic methods were soon superseded by us (1958) by means of a new development of the concept of the joint probability distribution. This consisted in the application of the general formulation of Monograph I (1953) to obtain joint probability distributions based on the assumption that the crystal structure is fixed and the reciprocal vectors are randomly but not independently distributed variables. The earlier distributions used in Monograph I (1953), lead, in general, to formulas having probable validity, whereas the new probability approach leads to formulas having exact validity in the sense of extremely rapid convergence of an infinite series. In fact, the algebraic formulas are special cases of those obtained from the latest probability approach.

It is our aim to present the details of a routine procedure for phase determination in all the space groups, utilizing the latest formulas. In this connection, it is interesting to note to what extent these formulas fulfill the criteria listed above. Criteria (a) and (b) concerning a uniform program for treating all the space groups, regardless of the chemical content, are completely satisfied. So far as criterion (c) is concerned, it is also satisfied subject to the requirement that the atomic coordinates satisfy a condition of rational independence (Hauptman & Karle, 1957).* Naturally, in practical application, the accuracy of the calculations is affected by the number and quality of the data.

Whereas the same type of computation serves uniformly for all the space groups, the interpretation of the results is space group dependent. In making this interpretation, the theory of structure seminvariants (Monograph I, 1953 and Hauptman & Karle, 1959) will be seen to play a central role in that they identify the linear combinations of the phases which are directly determined by the observed intensities. The interpretation is completed by making use of relationships among the structure factors which are characteristic of the space group and the chosen functional form of the structure factor.

The formulas for phase determination to be proposed have been proven in detail by us (1958) only for space groups $P\bar{1}$ and $P1$ for the equal atom case. We have since extended the results to include all the space groups (non-centrosymmetric as well as centrosymmetric) and have generalized them to include the case of unequal atoms. Since the derivation of these extensions follows the same lines as that previously given by us (1958) and is extremely tedious, we omit the proofs for the present. It is found that the formulas always yield the phases and magnitudes for what may be called the 'squared structure', i.e. the structure isomorphous to the structure of interest with the weights of the atomic scattering factors proportional to the squares of the weights of those of the original structure. Thus, one obtains directly the squared structure from which the true structure may be inferred. Clearly, the squared structure and the true structure coincide for the case of equal atoms.

2. Notation

We define the 'quasi-normalized structure factor' $\varepsilon_{\mathbf{h}}$ by means of†

$$\varepsilon_{\mathbf{h}} = \frac{1}{\sigma_2^{1/2}} \sum_{j=1}^{N} f_{j\mathbf{h}} \exp\left(2\pi i \mathbf{h}.\mathbf{r}_j\right), \qquad (2\cdot1)$$

$$\varepsilon_{\mathbf{h}} = |\varepsilon_{\mathbf{h}}| \exp\left(i\varphi_{\mathbf{h}}\right), \qquad (2\cdot2)$$

where N is the number of atoms in the unit cell, $f_{j\mathbf{h}}$ is the atomic scattering factor of the jth atom, $\mathbf{r}_j$ is the position vector of the jth atom and

$$\sigma_n = \sum_{j=1}^{N} f_{j\mathbf{h}}^n . \qquad (2\cdot3)$$

The quasi-normalized structure factor, ε, is to be

distinguished from the normalized structure factor, E (Karle & Hauptman, 1956, equation (3·12)), although for general reflections the two coincide. The magnitudes of the $\varepsilon_{\mathbf{h}}$ are assumed to be known from experiment while their phases $\varphi_{\mathbf{h}}$ coincide with those of the corresponding structure factors $F_{\mathbf{h}}$. Whereas, the normalized structure factors, E, were well adapted to the earlier probability methods (Monograph I, 1953), it will be seen that the quasi-normalized structure factors ε are better adapted to the newer methods.

The quasi-normalized structure factor $\varepsilon_{\mathbf{h}}'$ for the squared structure (cf. § 1) is given by

$$\varepsilon_{\mathbf{h}}' = |\varepsilon_{\mathbf{h}}'| \exp(i\varphi_{\mathbf{h}}') = \frac{1}{\sigma_4^{1/2}} \sum_{j=1}^{N} f_{j\mathbf{h}}^2 \exp(2\pi i \mathbf{h}.\mathbf{r}_j), \quad (2\cdot4)$$

where $\varphi_{\mathbf{h}}'$ is the phase of the structure factor $\varepsilon_{\mathbf{h}}'$. For the fourth power structure, we have

$$\varepsilon_{\mathbf{h}}''' = |\varepsilon_{\mathbf{h}}'''| \exp(i\varphi_{\mathbf{h}}''') = \frac{1}{\sigma_8^{1/2}} \sum_{j=1}^{N} f_{j\mathbf{h}}^4 \exp(2\pi i \mathbf{h}.\mathbf{r}_j), \quad (2\cdot5)$$

where $\varphi_{\mathbf{h}}'''$ is the phase of the structure factor $\varepsilon_{\mathbf{h}}'''$.

We define now some quantities which occur in the phase determining formulas:

$$\mu_p = \langle |\varepsilon_{\mathbf{k}}|^p \rangle_{\mathbf{k}}, \quad (2\cdot6)$$

$$\lambda_{p\mathbf{h}} = |\varepsilon_{\mathbf{h}}|^p - \mu_p, \quad (2\cdot7)$$

$$M_t = \left\langle \frac{|\varepsilon_{\mathbf{k}}|^t - 1}{\log |\varepsilon_{\mathbf{k}}|} \right\rangle_{\mathbf{k}}, \quad (2\cdot8)$$

and

$$\Lambda_{t\mathbf{h}} = \frac{|\varepsilon_{\mathbf{h}}|^t - 1}{\log |\varepsilon_{\mathbf{h}}|} - M_t, \quad (2\cdot9)$$

where the expression $(|\varepsilon|^t - 1)/\log |\varepsilon|$ is to be replaced by t when $|\varepsilon| = 1$. Finally $C_n(t)$ is defined by

$$C_n(t) = \frac{1}{(2\pi)^{1/2}} \int_0^t x^n 2^{(x+1)/2} \Gamma\left(\frac{x+1}{2}\right) dx \quad (2\cdot10)$$

and is given in Table 1, where the entries, calculated by numerical integration, are in error by not more than two units in the third significant figure. The symbol Γ in (2·10) represents the Gamma function.

Table 1

The values of

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

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	0·847	0·408	0·268	0·200
2	1·72	1·74	2·36	3·59
3	2·98	4·92	10·55	24·93
4	5·17	12·73	38·50	125·3
5	9·61	32·99	131·3	552·1
6	19·6	88·6	443	2296

3. Phase determining formulas

3·1. Basic formulas

$$B_{2,0}: \quad \varepsilon_{\mathbf{h}}'^2 = 1 + \frac{4\pi\sigma_2^2}{2^{(p+q+2)/2} pq \Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\sigma_4} \langle \lambda_{p\mathbf{k}}\lambda_{q(\mathbf{h}+\mathbf{k})}\rangle_{\mathbf{k}} + R_{2,0}. \quad (3\cdot1\cdot1)$$

$$B_{3,0}: \quad \varepsilon_{\mathbf{h}_1}'\varepsilon_{\mathbf{h}_2}'\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'$$
$$= \frac{(2\pi)^{3/2}\sigma_2^3}{2^{(p+q+r+3)/2} pqr \Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{r+1}{2}\right)\sigma_4^{3/2}} \langle \lambda_{p\mathbf{k}}\lambda_{q(\mathbf{h}_1+\mathbf{k})}\lambda_{r(\mathbf{h}_1+\mathbf{h}_2+\mathbf{k})}\rangle_{\mathbf{k}}$$
$$-2\frac{\sigma_6}{\sigma_4^{3/2}} + \frac{\sigma_8^{1/2}}{\sigma_4}(\varepsilon_{\mathbf{h}_1}'\varepsilon_{\mathbf{h}_1}''' + \varepsilon_{\mathbf{h}_2}'\varepsilon_{\mathbf{h}_2}''' + \varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}''') + R_{3,0}. \quad (3\cdot1\cdot2)$$

3·2. Integrated formulas

$$I_{2,0}: \quad \varepsilon_{\mathbf{h}}'^2 = 1 + \frac{2\sigma_2^2}{C_1^2(t)\sigma_4}\langle \Lambda_{t\mathbf{k}}\Lambda_{t(\mathbf{h}+\mathbf{k})}\rangle_{\mathbf{k}} + R_{2,0}'. \quad (3\cdot2\cdot1)$$

$$I_{3,0}: \quad \varepsilon_{\mathbf{h}_1}'\varepsilon_{\mathbf{h}_2}'\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}' = \frac{\sigma_2^3}{C_1^3(t)\sigma_4^{3/2}}\langle \Lambda_{t\mathbf{k}}\Lambda_{t(\mathbf{h}_1+\mathbf{k})}\Lambda_{t(\mathbf{h}_1+\mathbf{h}_2+\mathbf{k})}\rangle_{\mathbf{k}}$$
$$-2\frac{\sigma_6}{\sigma_4^{3/2}} + \frac{\sigma_8^{1/2}}{\sigma_4}(\varepsilon_{\mathbf{h}_1}'\varepsilon_{\mathbf{h}_1}''' + \varepsilon_{\mathbf{h}_2}'\varepsilon_{\mathbf{h}_2}''' + \varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}''') + R_{3,0}'. \quad (3\cdot2\cdot2)$$

In these formulas, p, q, r and t are restricted to be positive. Ordinarily they are given values in the range 2–4. A further discussion of this matter is found in § 5.

The remainder terms are given in the appendix § 6. Equation (3·1·1) or (3·2·1) serves to determine the magnitudes of the structure factors $|\varepsilon_{\mathbf{h}}'|$ corresponding to the squared structure. By means of equation (3·1·2) or (3·2·2), the phases of these structure factors $\varphi_{\mathbf{h}}'$ may be determined. In the next section we describe in detail how these equations are to be used for the various space groups included in type 1P, the conventionally primitive space groups in the triclinic, monoclinic and orthorhombic systems (Table 1, p. 14, Monograph I, 1953).

It should be clear that although the emphasis in this paper is on a unified program for phase determination, several of the simpler phase determining formulas which have already appeared in the literature may serve to supplement the present method.

4. Phase determining procedure

It is assumed that the $|\varepsilon_{\mathbf{h}}|$ have been calculated from the observed intensities. From these, the $|\varepsilon_{\mathbf{h}}'|$ are obtained from (3·1·1) or (3·2·1). In fact the $|\varepsilon_{\mathbf{h}}'|$ so computed may be made to cover a range of reflections extending beyond that of the original set of observations. We are here concerned only with the larger $|\varepsilon_{\mathbf{h}}'|$ and it is the phases of these whose values are to be determined.

In the application of (3·1·2) or (3·2·2), the values

of some $|\varepsilon_\mathbf{h}'''|$ may be required. These may be obtained from (3·1·1) or (3·2·1) in which ε is replaced by ε' and ε' by ε'''.

Although the same phase determining formulas (3·1·2) or (3·2·2) are used for all the space groups, full use is made of the advantages accruing from the space group symmetries. This arises from the fact that special choices of the indices $\mathbf{h}_1$ and $\mathbf{h}_2$, peculiar to the space group, permit the calculation of special phases. The choices of indices are based upon relations among the structure factors which depend upon the particular space group and, in general, on the choice of functional form for the structure factor. Space groups of type $1P$, have, of course, only one functional form for the structure factor (Monograph I, 1953). In the phase determining procedures to follow, it will be seen that the first steps concern the application of (3·1·2) or (3·2·2) with choices of indices which take advantage of the space group symmetry. The final step is in the form of a general application which is the same for all the space groups.

The specification of the origin is carried out in conformance with the seminvariant theory developed in Monograph I (1953). Origin specification in all space groups of a given type is the same. Thus, the method of origin specification for space group $P\bar{1}$ serves as a model for the remaining ones of type $1P$.

4·1. Triclinic system

In the triclinic system we are concerned with space group $P\bar{1}$. The only special choice for $\mathbf{h}_1$ and $\mathbf{h}_2$ arises for the case that $\mathbf{h}_1 = \mathbf{h}_2$. Since equations (3·1·2) and (3·2·2) then give the values of $\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{2\mathbf{h}_1}'$, in practice only those values of $\mathbf{h}_1$ are chosen which give the largest values for $|\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{2\mathbf{h}_1}'|$. In this way the phases of some of the largest $\varepsilon_{2\mathbf{h}}'$ may be found. It is conceivable that in an unfavorable case no large values are obtained for $|\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{2\mathbf{h}_1}'|$. In this case we would omit this special calculation and go on to the next step.

For the purpose of specifying the origin, a linearly semi-independent triple of phases, $\varphi_{\mathbf{h}_i}'$, $i = 1, 2, 3$, having large corresponding $|\varepsilon_{\mathbf{h}_i}'|$ is chosen. The values of these $\varphi_{\mathbf{h}_i}'$ are then specified arbitrarily (i.e. 0 or π), thus fixing the origin. Systematic use of equation (3·1·2) or (3·2·2) then permits the determination of the phases $\varphi_{\mathbf{h}_i}'$ of the remaining $\varepsilon_{\mathbf{h}_i}'$ of interest, i.e. the larger ones. Of course, as the calculation proceeds, use is made of previously determined or specified phases. Examination shows that contained within this procedure is the possibility of a great many different ways for determining the same phase. This property may be expressed in terms of an identity involving the phases, as previously described (Hauptman & Karle, 1958).

4·2. Monoclinic system

There are two special choices of $\mathbf{h}_1$ and $\mathbf{h}_2$ characteristic of the space groups in the monoclinic system,

in addition to $\mathbf{h}_1 = \mathbf{h}_2$ which is valid for $P\bar{1}$. These are shown in the first two rows of Table 2. By means

Table 2

The coefficients of $\varepsilon_{\mathbf{h}_1}'^2\varepsilon_\mathbf{h}'$ given by the left sides of (3·1·2) or (3·2·2) for selected values of $\mathbf{h}_1$ and $\mathbf{h}_2$ (shown in the first two rows), and for each of the four primitive centrosymmetric space groups of the monoclinic system. From these, $\varepsilon_\mathbf{h}'$ may be found, where $\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2$

$\mathbf{h}_1$	hkl	hkl
$\mathbf{h}_2$	$\bar{h}k\bar{l}$	$h\bar{k}l$
$\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2$	$02k0$	$2h02l$
$P2/m$	$+1$	$+1$
$P2_1/m$	$(-1)^k$	$(-1)^k$
$P2/c$	$(-1)^l$	$(-1)^l$
$P2_1/c$	$(-1)^{k+l}$	$(-1)^{k+l}$

of the first of these, $\mathbf{h}_1 = (h, k, l)$ and $\mathbf{h}_2 = (\bar{h}, k, \bar{1})$, equation (3·1·2) or (3·2·2) yields the value of $\varepsilon_{hkl}'^2\varepsilon_{02k0}'$ multiplied by the numerical coefficient given in the second column of Table 2. In this way the value of the phase φ_{02k0}' is determined. Since h and l may be chosen arbitrarily, φ_{02k0}' may possibly be determined in many ways. Again we note that the computations are performed only for the larger values of $|\varepsilon_{\mathbf{h}_1}'^2\varepsilon_\mathbf{h}'|$.

The second relationship $\mathbf{h}_1 = (h, k, l)$ and $\mathbf{h}_2 = (h, \bar{k}, l)$, leads to the value of the phase φ_{2h02l}' by means of the numbers listed in the third column of Table 2 and (3·1·2) or (3·2·2).

The numbers listed in columns two and three of Table 2 arise from relationships among the structure factors which are characteristic of the particular space group. For example, in space group $P2/m$, $\varepsilon_{hkl}' = \varepsilon_{\bar{h}k\bar{l}}' = \varepsilon_{h\bar{k}l}'$, whereas in space group $P2_1/m$, $\varepsilon_{hkl}' = (-1)^k\varepsilon_{\bar{h}k\bar{l}}' = (-1)^k\varepsilon_{h\bar{k}l}'$. Thus the substitution of $\mathbf{h}_1$ and $\mathbf{h}_2$ of Table 2 into (3·1·2) or (3·2·2) yields, for the left sides, $\varepsilon_{\mathbf{h}_1}'^2\varepsilon_\mathbf{h}'$ times the coefficients listed in columns two and three.

We proceed by specifying the origin and determining the remaining phases of interest in the same way as for $P\bar{1}$.

4·3. Orthorhombic system

The special choices of $\mathbf{h}_1$ and $\mathbf{h}_2$ characteristic of the space groups in the orthorhombic system, in addition to $\mathbf{h}_1 = \mathbf{h}_2$ which is valid for $P\bar{1}$, are shown in the first two rows of Table 3. By means of these and (3·1·2) or (3·2·2), the values of many of the phases φ_{2h00}', φ_{02k0}', φ_{002l}', φ_{2h2k0}', φ_{2h02l}' and φ_{02k2l}' associated with large $|\varepsilon_\mathbf{h}'|$ may be found. These choices of $\mathbf{h}_1$ and $\mathbf{h}_2$ arise, as before, as a consequence of the special relationships among the structure factors in the various space groups, e.g. in P_{nnn}, $\varepsilon_{hkl}' = (-1)^{k+l}\varepsilon_{\bar{h}k\bar{l}}' = (-1)^{l+h}\varepsilon_{\bar{h}k\bar{l}}' = (-1)^{h+k}\varepsilon_{h\bar{k}\bar{l}}' = (-1)^{h+k}\varepsilon_{hk\bar{l}}' = (-1)^{l+h}\varepsilon_{hk\bar{l}}' = (-1)^{k+l}\varepsilon_{\bar{h}kl}'$. Again, the substitution of $\mathbf{h}_1$ and $\mathbf{h}_2$ of Table 3 into (3·1·2) or (3·2·2) yields, for the left sides, $\varepsilon_{\mathbf{h}_1}'^2\varepsilon_\mathbf{h}'$ times the coefficients listed in the appropriate columns of Table 3.

Table 3

The coefficients of $\varepsilon'^2_{\mathbf{h}_1}\varepsilon'_{\mathbf{h}}$ given by the left sides of (3·1·2) or (3·2·2) for selected values of $\mathbf{h}_1$ and $\mathbf{h}_2$ (shown in the first two rows), and for each of the sixteen primitive centrosymmetric space groups of the orthorhombic system. From these, $\varepsilon'_{\mathbf{h}}$ may be found, where $\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2$

$\mathbf{h}_1$	hkl	hkl	hkl	hkl	hkl	hkl
$\mathbf{h}_2$	$h\bar{k}\bar{l}$	$\bar{h}k\bar{l}$	$\bar{h}\bar{k}l$	$hk\bar{l}$	$h\bar{k}l$	$\bar{h}kl$
$\mathbf{h}=\mathbf{h}_1+\mathbf{h}_2$	$2h00$	$02k0$	$002l$	$2h2k0$	$2h02l$	$02k2l$
$Pmmm$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$
$Pnnn$	$(-1)^{k+l}$	$(-1)^{l+h}$	$(-1)^{h+k}$	$(-1)^{h+k}$	$(-1)^{l+h}$	$(-1)^{k+l}$
$Pccm$	$(-1)^l$	$(-1)^l$	$+1$	$+1$	$(-1)^l$	$(-1)^l$
$Pban$	$(-1)^k$	$(-1)^h$	$(-1)^{h+k}$	$(-1)^{h+k}$	$(-1)^h$	$(-1)^k$
$Pmma$	$(-1)^h$	$+1$	$(-1)^h$	$(-1)^h$	$+1$	$(-1)^h$
$Pnna$	$(-1)^{k+l}$	$(-1)^{h+k+l}$	$(-1)^h$	$(-1)^h$	$(-1)^{h+k+l}$	$(-1)^{k+l}$
$Pmna$	$+1$	$(-1)^{l+h}$	$(-1)^{l+h}$	$(-1)^{l+h}$	$(-1)^{l+h}$	$+1$
$Pcca$	$(-1)^{l+h}$	$(-1)^l$	$(-1)^h$	$(-1)^h$	$(-1)^l$	$(-1)^{l+h}$
$Pbam$	$(-1)^{h+k}$	$(-1)^{h+k}$	$+1$	$+1$	$(-1)^{h+k}$	$(-1)^{h+k}$
$Pccn$	$(-1)^{l+h}$	$(-1)^{k+l}$	$(-1)^{h+k}$	$(-1)^{h+k}$	$(-1)^{k+l}$	$(-1)^{l+h}$
$Pbcm$	$(-1)^k$	$(-1)^{k+l}$	$(-1)^l$	$(-1)^l$	$(-1)^{k+l}$	$(-1)^k$
$Pnnm$	$(-1)^{h+k+l}$	$(-1)^{h+k+l}$	$+1$	$+1$	$(-1)^{h+k+l}$	$(-1)^{h+k+l}$
$Pmmn$	$(-1)^h$	$(-1)^k$	$(-1)^{h+k}$	$(-1)^{h+k}$	$(-1)^k$	$(-1)^h$
$Pbcn$	$(-1)^{h+k}$	$(-1)^l$	$(-1)^{h+k+l}$	$(-1)^{h+k+l}$	$(-1)^l$	$(-1)^{h+k}$
$Pbca$	$(-1)^{h+k}$	$(-1)^{k+l}$	$(-1)^{l+h}$	$(-1)^{l+h}$	$(-1)^{k+l}$	$(-1)^{h+k}$
$Pnma$	$(-1)^{h+k+l}$	$(-1)^k$	$(-1)^{l+h}$	$(-1)^{l+h}$	$(-1)^k$	$(-1)^{h+k+l}$

The procedure is completed by specifying the origin and determining the remaining phases of interest in the same way as for $P\bar{1}$.

5. Concluding remarks

The computation of the averages in (3·1·1), (3·1·2), (3·2·1) and (3·2·2) on an IBM 704 has been programmed by Mr Peter O'Hara of the National Bureau of Standards. Thus, these computations are readily feasible with modern computing facilities.

Further experience in the application of the phase determining formulas will indicate what are the best values of p, q, r or t to use. In effect, there are two competing influences affecting the accuracy of the calculation. As p, q, r or t increase the variances of the averages increase and, in addition, the inaccuracies inherent in the experimental data are exaggerated. On the other hand, in compensation for this, the coefficients which multiply these averages are found to decrease with increasing p, q, r or t. On the basis of our limited experience, values of p, q, r or t in the range of 2 to 4 appear to be most suitable. We prefer to use the integrated formulas (3·2·1) and (3·2·2), especially since the correction terms listed in § 6 are generally quite small.

A few words are in order concerning space group extinctions. If an unlimited set of data were available, there would be no problem concerning the inclusion, in the computations of the averages, of those $\varepsilon_{\mathbf{h}}$ which are zero because of space group extinctions, since the average would not thereby be altered. However, with limited data, this question assumes some significance and is related to the fact that certain of the one and two dimensional reflections assume exaggerated importance when using the ε values instead of the E. In our opinion, coupled with the use of the ε in preference to the E, the $\varepsilon_{\mathbf{h}}$ which are zero because of space group extinction should be included in the calculation of the averages. Of course, accidental zeros are of necessity included in the computations.

As a result of the calculations, the values of the larger $|\varepsilon'_{\mathbf{h}}|$ and the corresponding phases $\varphi'_{\mathbf{h}}$ are found. In view of our discussion in a previous paper (Karle, Hauptman, Karle & Wing, 1958), of the use of ε maps, we recommend that the next step be the calculation of an ε map in which it is assumed that $\varphi_{\mathbf{h}} = \varphi'_{\mathbf{h}}$. For centrosymmetric structures and for large $|\varepsilon_{\mathbf{h}}|$ and $|\varepsilon'_{\mathbf{h}}|$, this latter assumption will rarely, if ever, be violated. An ε' map, giving the squared structure, may be helpful in interpreting the ε map since it facilitates the identification of different kinds of atoms. It should be noted in passing that a map computed with $|\varepsilon'_{\mathbf{h}}|^2$ may also prove to be a useful aid since it exaggerates the Patterson peaks arising from the heaviest atoms.

6. Appendix

The correction terms for the formulas listed in § 3 are given here. With the possible exception of the first term of (6·1), (6·2), (6·5) and (6·6), and when $\mathbf{h}_1 = \mathbf{h}_2$, the first terms of (6·3) and (6·7), these corrections are quite small. They all decrease as N increases. Nevertheless these terms are included here for the sake of completeness, and, in any given case, the investigator can evaluate their effect in practical application.

We define:

$$_1R_{2,0}^{(0)} = \frac{\sigma_8^{1/2}}{2\sigma_4}\varepsilon_{2\mathbf{h}}''' - \frac{\sigma_6}{\sigma_2\sigma_4}(p+q-4)$$

$$- \frac{\sigma_4}{32\sigma_2^2}(100-42p-42q+9pq+4p^2+4q^2)$$

$$+ \frac{\sigma_4}{8\sigma_2^2}(32-14p-14q+4pq+p^2+q^2)\varepsilon_{\mathbf{h}}'^2$$

$$+ \frac{\sigma_8^{1/2}}{\sigma_2\sigma_4^{1/2}}(p+q-4)\varepsilon_{\mathbf{h}}'\varepsilon_{\mathbf{h}}'''$$

$$- \frac{\sigma_4}{12\sigma_2^2}(p-2)(q-2)\varepsilon_{\mathbf{h}}'^4 + \frac{\sigma_4}{32\sigma_2^2}(p-2)(q-2) + \ldots, \quad (6\cdot1)$$

$$_2R_{2,0}^{(0)} = -\frac{\sigma_8^{1/2}}{\sigma_4}(\varepsilon_{2h00}'''+\varepsilon_{02k0}'''+\varepsilon_{002l}''')$$

$$- \frac{2\sigma_8^{1/2}}{\sigma_2\sigma_4^{1/2}}(p+q-4)\varepsilon_{\mathbf{h}}'\varepsilon_{\mathbf{h}}'''$$

$$- \frac{\sigma_4}{4\sigma_2^2}((p-2)(p-4)+(q-2)(q-4))\varepsilon_{\mathbf{h}}'^2$$

$$+ \frac{2\sigma_6}{\sigma_2\sigma_4}(p+q-4)$$

$$+ \frac{\sigma_4}{16\sigma_2^2}((p-2)(q-2)+2(p-2)(p-4)$$

$$+ 2(q-2)(q-4)) + \ldots, \quad (6\cdot2)$$

$$_1R_{3,0}^{(0)} = \frac{\sigma_8^{1/2}}{2\sigma_4}(\varepsilon_{\mathbf{h}_1}'\varepsilon_{\mathbf{h}_1+2\mathbf{h}_2}'''+\varepsilon_{\mathbf{h}_2}'\varepsilon_{2\mathbf{h}_1+\mathbf{h}_2}'''+\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'\varepsilon_{\mathbf{h}_1-\mathbf{h}_2}''')$$

$$- \frac{5\sigma_4^{1/2}}{16\sigma_2}(p+q+r-6)$$

$$+ \frac{\sigma_4^{1/2}}{4\sigma_2}((p+q-4)\varepsilon_{\mathbf{h}_1}'^2+(q+r-4)\varepsilon_{\mathbf{h}_2}'^2$$

$$+ (p+r-4)\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2)$$

$$- \frac{\sigma_4^{1/2}}{4\sigma_2}((q-2)\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{\mathbf{h}_2}'^2+(p-2)\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2$$

$$+ (r-2)\varepsilon_{\mathbf{h}_2}'^2\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2+\tfrac{1}{4}(p+q+r-6)) + \ldots, \quad (6\cdot3)$$

and

$$_2R_{3,0}^{(0)} = -\frac{\sigma_4^{1/2}}{8\sigma_2}((r-2)\varepsilon_{\mathbf{h}_1}'^2+(p-2)\varepsilon_{\mathbf{h}_2}'^2+(q-2)\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2)$$

$$- \frac{\sigma_8^{1/2}}{\sigma_4}\varepsilon_{\mathbf{h}_1}'(\varepsilon_{h_1+2h_2,\,k_1,\,l_1}'''+\varepsilon_{h_1,\,k_1+2k_2,\,l_1}'''+\varepsilon_{h_1,\,k_1,\,l_1+2l_2}''')$$

$$- \frac{\sigma_8^{1/2}}{\sigma_4}\varepsilon_{\mathbf{h}_2}'(\varepsilon_{2h_1+h_2,\,k_2,\,l_2}'''+\varepsilon_{h_2,\,2k_1+k_2,\,l_2}'''+\varepsilon_{h_2,\,k_2,\,2l_1+l_2}''')$$

$$- \frac{\sigma_8^{1/2}}{\sigma_4}\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'(\varepsilon_{h_1-h_2,\,k_1+k_2,\,l_1+l_2}'''$$

$$+ \varepsilon_{h_1+h_2,\,k_1-k_2,\,l_1+l_2}'''+\varepsilon_{h_1+h_2,\,k_1+k_2,\,l_1-l_2}''') + \ldots, \quad (6\cdot4)$$

Next we define (where $C_n(t)$) is replaced by C_n:

$$_1R_{2,0}^{(1)} = \frac{\sigma_8^{1/2}}{2\sigma_4}\varepsilon_{2\mathbf{h}}'''+\frac{2\sigma_6}{C_1\sigma_2\sigma_4}(2C_1-C_2)$$

$$- \frac{\sigma_4}{32C_1^2\sigma_2^2}(100C_1^2-84C_1C_2+9C_2^2+8C_1C_3)$$

$$+ \frac{\sigma_4}{4C_1^2\sigma_2^2}(16C_1^2-14C_1C_2+2C_2^2+C_1C_3)\varepsilon_{\mathbf{h}}'^2$$

$$- \frac{2\sigma_8^{1/2}}{C_1\sigma_2\sigma_4^{1/2}}(2C_1-C_2)\varepsilon_{\mathbf{h}}'\varepsilon_{\mathbf{h}}'''$$

$$- \frac{\sigma_4}{12C_1^2\sigma_2^2}(2C_1-C_2)^2\varepsilon_{\mathbf{h}}'^4+\frac{\sigma_4}{32C_1^2\sigma_2^2}(2C_1-C_2)^2$$

$$+ \ldots, \quad (6\cdot5)$$

$$_2R_{2,0}^{(1)} = -\frac{\sigma_8^{1/2}}{\sigma_4}(\varepsilon_{2h00}'''+\varepsilon_{02k0}'''+\varepsilon_{002l}''')$$

$$+ \frac{4\sigma_8^{1/2}}{C_1\sigma_2\sigma_4^{1/2}}(2C_1-C_2)\varepsilon_{\mathbf{h}}'\varepsilon_{\mathbf{h}}'''$$

$$- \frac{\sigma_4}{2C_1^2\sigma_2^2}(8C_1-6C_2+C_3)\varepsilon_{\mathbf{h}}'^2-\frac{4\sigma_6}{C_1\sigma_2\sigma_4}(2C_1-C_2)$$

$$+ \frac{\sigma_4}{16C_1^2\sigma_2^2}((2C_1-C_2)^2+4C_1(8C_1-6C_2+C_3))$$

$$+ \ldots, \quad (6\cdot6)$$

$$_1R_{3,0}^{(1)} = \frac{\sigma_8^{1/2}}{2\sigma_4}(\varepsilon_{\mathbf{h}_1}'\varepsilon_{\mathbf{h}_1+2\mathbf{h}_2}'''+\varepsilon_{\mathbf{h}_2}'\varepsilon_{2\mathbf{h}_1+\mathbf{h}_2}'''+\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'\varepsilon_{\mathbf{h}_1-\mathbf{h}_2}''')$$

$$+ \frac{15\sigma_4^{1/2}}{16C_1\sigma_2}(2C_1-C_2)$$

$$- \frac{\sigma_4^{1/2}}{2C_1\sigma_2}(2C_1-C_2)(\varepsilon_{\mathbf{h}_1}'^2+\varepsilon_{\mathbf{h}_2}'^2+\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2)$$

$$+ \frac{\sigma_4^{1/2}}{4C_1\sigma_2}(2C_1-C_2)(\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{\mathbf{h}_2}'^2+\varepsilon_{\mathbf{h}_1}'^2\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2$$

$$+ \varepsilon_{\mathbf{h}_2}'^2\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2+\tfrac{3}{4}) + \ldots, \quad (6\cdot7)$$

$$_2R_{3,0}^{(1)} = \frac{\sigma_4^{1/2}}{8C_1\sigma_2}(2C_1-C_2)(\varepsilon_{\mathbf{h}_1}'^2+\varepsilon_{\mathbf{h}_2}'^2+\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'^2)$$

$$- \frac{\sigma_8^{1/2}}{\sigma_4}\varepsilon_{\mathbf{h}_1}'(\varepsilon_{h_1+2h_2,\,k_1,\,l_1}'''+\varepsilon_{h_1,\,k_1+2k_2,\,l_1}'''+\varepsilon_{h_1,\,k_1,\,l_1+2l_2}''')$$

$$- \frac{\sigma_8^{1/2}}{\sigma_4}\varepsilon_{\mathbf{h}_2}'(\varepsilon_{2h_1+h_2,\,k_2,\,l_2}'''+\varepsilon_{h_2,\,2k_1+k_2,\,l_2}'''+\varepsilon_{h_2,\,k_2,\,2l_1+l_2}''')$$

$$- \frac{\sigma_8^{1/2}}{\sigma_4}\varepsilon_{\mathbf{h}_1+\mathbf{h}_2}'(\varepsilon_{h_1-h_2,\,k_1+k_2,\,l_1+l_2}'''$$

$$+ \varepsilon_{h_1+h_2,\,k_1-k_2,\,l_1+l_2}'''+\varepsilon_{h_1+h_2,\,k_1+k_2,\,l_1-l_2}''') + \ldots. \quad (6\cdot8)$$

For the conventionally primitive space groups of the triclinic and monoclinic systems,

$$R_{i,0}^{(j)} = {}_1R_{i,0}^{(j)}+\ldots; \quad i = 2,3; \quad j = 0,1. \quad (6\cdot9)$$

For the conventionally primitive space groups of the orthorhombic system,

$$R_{i,0}^{(j)} = {}_1R_{i,0}^{(j)}+{}_2R_{i,0}^{(j)}+\ldots; \quad i = 2,3; \quad j = 0,1. \quad (6\cdot10)$$

The remainder terms in the basic formulas are especially simple for the special case, $p = q = r = 2$. For this case, the formulas reduce to those obtainable by the algebraic methods proposed by us (1957).

References

BERTAUT, E. F. (1955). *Acta Cryst.* **8**, 537, 544.
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.
HAUPTMAN, H. & KARLE, J. (1958). *Acta Cryst.* **11**, 149.
HAUPTMAN, H. & KARLE, J. (1959). *Acta Cryst.* In press.
KARLE, I. L., HAUPTMAN, H., KARLE, J. & WING, A. B. (1958). *Acta Cryst.* **11**, 257.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.
KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515.
KARLE, J. & HAUPTMAN, H. (1958). *Acta Cryst.* **11**, 264.
KITAIGORODSKY, A. I. (1954). *Doklady Acad. Nauk U.S. S.R.* **94**, 225. *Trudy Inst. Crystallog.* **10**, 27.
KLUG, A. (1958). *Acta Cryst.* **11**, 515.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60.
VAUGHAN, P. A. (1958). *Acta Cryst.* **11**, 111.

Short Communications

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 500 words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible; and proofs will not generally be submitted to authors. Publication will be quicker if the contributions are without illustrations.

A new statistical test for distinguishing between centrosymmetric and non-centrosymmetric structures.

By G. N. RAMACHANDRAN and R. SRINIVASAN, *Department of Physics, University of Madras, Madras–25, India*

The problem of the distribution of intensity in the resultant obtained by compounding n vibrations of equal amplitude, but of random phase, was considered by Lord Rayleigh (1880) (see also Ramachandran, 1943). He considered two cases—one in which the phases are assumed to have only the values 0 or π, and the other, the more general case, in which the phases are arbitrary. It will be readily recognized that these two correspond to the problem of calculating the distribution of X-ray intensities for a centrosymmetric and a non-centrosymmetric crystal structure, in which all the atoms are alike. The formulae obtained by Lord Rayleigh may be stated in the following form:

$$_N P(z)dz = \exp(-z)dz \tag{1}$$

$$_C P(z)dz = \frac{1}{(2\pi z)^{\frac{1}{2}}} \exp(-z/2)dz \tag{2}$$

where $z = I/\langle I \rangle$ and $P(z)dz$ represents the probability that the fraction $I/\langle I \rangle$ occurs between z and $z+dz$. The subscripts N and C refer to non-centrosymmetric and centrosymmetric structures.

These formulae have also been derived by Wilson (1949) and on their basis Howells, Phillips & Rogers (1950) have suggested a test for distinguishing between centrosymmetric and non-centrosymmetric structures. However, they have used the fraction

$$N(z) = \int_0^z P(z)dz \tag{3}$$

for this purpose. The $N(z)$ curves, as is characteristic of all cumulative distribution functions, start at the origin and are more or less similar in their general shape for both types of structures. Consequently a critical distinc-

tion between the two is not always possible and, although success has been reported in various cases using this test, it has led to negative results (Whittaker, 1953; Eriks & McGillavry, 1954; Paton & MacDonald, 1957) and even false results in a few cases (e.g., see Robertson & Shearer, 1956; Robertson, Shearer, Sim & Watson, 1958).
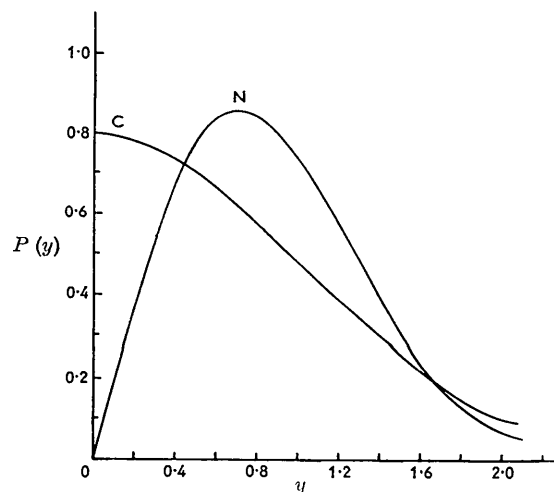
However, if we modify the original probability distri-

Fig. 1. Theoretical curves of $P(y)$ against y for centrosymmetric (C) and non-centrosymmetric (N) distributions.

bution functions such that the argument is $y = \sqrt{z}$ and not z, then the following formulae are obtained:

$$_N P(y)dy = 2y \exp(-y^2)dy \tag{4}$$