

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.

Acta Cryst. (1954). 7, 450

Have Hauptman & Karle solved the phase problem? By W. COCHRAN and M. M. WOOLFSON, *Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England*

(Received 25 January 1954*)

In a recent publication Hauptman & Karle (1953) have given a proposed solution of the phase problem for centrosymmetrical structures. By expressing their equations in a simpler way we shall show that their method is no more useful than other direct methods already in use, such as those due to Harker & Kasper (1948) and Sayre (1952). We shall consider only the formulae given by Hauptman & Karle for the step-by-step determination of signs for the space group $P\bar{1}$. Their solution is of the following form:

The sign of $E_{\mathbf{h}}$ ($= F_{\mathbf{h}}/(\overline{F_{\mathbf{h}}^2})^{\frac{1}{2}}$) is the sign of

$$\Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4,$$

where

$$\Sigma_1 = \frac{\sum_j z_j^3}{4(\sum_j z_j^2)^{3/2}} \sum_{\mathbf{h}=2\mathbf{h}_\mu} (E_{\mathbf{h}_\mu}^2 - 1), \quad (4.03)$$

$$\Sigma_2 = \frac{\sum_j z_j^3}{2(\sum_j z_j^2)^{3/2}} \sum_{\mathbf{h}=\mathbf{h}_\mu+\mathbf{h}_\nu} E_{\mathbf{h}_\mu} E_{\mathbf{h}_\nu}, \quad (4.04)$$

$$\Sigma_3 = \frac{\sum_j z_j^4}{4(\sum_j z_j^2)^2} \sum_{\mathbf{h}=\mathbf{h}_\mu+2\mathbf{h}_\nu} E_{\mathbf{h}_\mu} (E_{\mathbf{h}_\nu}^2 - 1), \quad (4.05)$$

$$\Sigma_4 = \frac{\sum_j z_j^5}{8(\sum_j z_j^2)^{5/2}} \sum_{\mathbf{h}=2\mathbf{h}_\mu+2\mathbf{h}_\nu} (E_{\mathbf{h}_\mu}^2 - 1)(E_{\mathbf{h}_\nu}^2 - 1). \quad (4.06)$$

The equation numbers are those given by Hauptman & Karle. Equation (4.04) is, as Hauptman & Karle point out, a modified form of Sayre's equation, and the sign of $F_{\mathbf{h}}$ is the sign of Σ_2 , when enough terms are included in Σ_2 . However this follows inevitably only when all the atoms are equal; one of us (Cochran, 1953) has already pointed out that when the atoms are unequal it is possible to have

$$s\{F_{\mathbf{h}}\} \neq s\left\{\sum_{\mathbf{h}'} F_{\mathbf{h}'} F_{\mathbf{h}+\mathbf{h}'}\right\} = s(\Sigma_2),$$

since $E_{\mathbf{h}}$ is proportional to $F_{\mathbf{h}}$. Nevertheless we accept the fact that (4.04) will enable additional signs to be determined once a sufficient number of signs is known.

Equations (4.03) and (4.06) are the only ones which

* In a private communication, dated 23 November 1953, Prof. Pepinsky had informed us that he considered the method of Hauptman & Karle to be 'largely fallacious, and less useful than Harker sections', although no reasons were given for this conclusion.

can be used initially, as they give signs from a knowledge of the intensities alone.

We consider a centrosymmetrical structure containing N equal point atoms for which $Z = 1$.

Then

$$E_{\mathbf{h}} = F_{\mathbf{h}}/(\overline{F_{\mathbf{h}}^2})^{\frac{1}{2}} = F_{\mathbf{h}}/\sqrt{N}.$$

Equation (4.03) then amounts to the statement that $s(F_{2\mathbf{h}})$ is probably equal to $s(\Sigma_1)$, where

$$\Sigma_1 = \frac{1}{4\sqrt{N}} \left(\frac{F_{\mathbf{h}}^2}{N} - 1 \right),$$

or that

$$s(F_{2\mathbf{h}}) \approx s(F_{\mathbf{h}}^2 - N). \quad (1)$$

The symbol \approx is used to denote 'probably equals'.

Now let us consider the Patterson function of the same structure of N equal point atoms with coordinates $\pm \mathbf{r}_j$. There will be an origin peak of weight N , N single-weight peaks at $\pm 2\mathbf{r}_j$ and $\frac{1}{2}N^2 - N$ double-weight peaks at $\pm(\mathbf{r}_i - \mathbf{r}_j)$. The Fourier coefficients of the Patterson function are the values of $F_{\mathbf{h}}^2$, and from the above considerations we find

$$F_{\mathbf{h}}^2 = N + F_{2\mathbf{h}} + 4 \sum_{\substack{i=1 \\ i \neq j}}^{N/2} \sum_{j=1}^{N/2} \cos 2\pi \mathbf{h} \cdot (\mathbf{r}_i - \mathbf{r}_j),$$

there being $\frac{1}{4}(N^2 - 2N)$ terms in the summation.

This may be rewritten as

$$F_{2\mathbf{h}} = F_{\mathbf{h}}^2 - N - 4 \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} \cos 2\pi \mathbf{h} \cdot (\mathbf{r}_i - \mathbf{r}_j).$$

This shows that (1) is true to the extent that $F_{\mathbf{h}}^2 - N$ outweighs the summation terms. The correlation between $s(F_{2\mathbf{h}})$ and $s(F_{\mathbf{h}}^2 - N)$, looked at from this point of view, is seen to be weak, since $F_{2\mathbf{h}}$ is the difference of two much larger quantities, one of which is $F_{\mathbf{h}}^2 - N$. The statement (1) is in fact true to the extent that all peaks in the Patterson function can be ignored, except the origin peak and those which arise from atoms related by the centre of symmetry!

To estimate the probability that relation (1) is correct in a particular instance, we use equation (3.29) of Hauptman & Karle, which for space group $P\bar{1}$ gives the probability that $F_{2\mathbf{h}}$ is positive as

$$P_+(F_{2\mathbf{h}}) = \frac{1}{2} + \frac{1}{4N^2} |F_{2\mathbf{h}}| (F_{\mathbf{h}}^2 - N).$$

Since the distribution of the F 's is approximately Gaussian, values of $F_{2\mathbf{h}}$ or of $F_{\mathbf{h}}$ greater than $3\sqrt{N}$ are scarcely to be expected, and thus the maximum value of $P_+(F_{2\mathbf{h}})$

will be about $\frac{1}{2} + 6/\sqrt{N}$. (The fact that this probability may exceed unity is noted by Karle & Hauptman.) The point we wish to make is that as N increases, even the highest values of $P_+(F_{2h})$ tend to $\frac{1}{2}$. A similar conclusion can be reached from the 'Patterson function' point of view which we outlined above. It is easily shown that the situation for the determination of F_{2h} as negative is even less favourable.

We conclude that relation (1) is of value only for very simple crystal structures.

At first sight equation (4.06) looks more promising, since Σ_4 may contain a large number of terms. It might appear that although the indication of sign given by a single term of Σ_4 was weak, a large number of such indications would be statistically significant. We shall therefore consider the case where Σ_4 contains an infinite number of terms.

Then, in our notation,

$$\begin{aligned} s(\Sigma_4) &= s\left\{ \sum_{\mathbf{h}'} (F_{\mathbf{h}'}^2 - N)(F_{\mathbf{h}+\mathbf{h}'}^2 - N) \right\} \\ &= s\left\{ \sum_{\mathbf{h}'} (F_{\mathbf{h}'}^2 F_{\mathbf{h}+\mathbf{h}'}^2 - N^2) \right\}, \end{aligned}$$

since $\overline{F_{\mathbf{h}}^2} = N$.

Then, by (4.06),

$$s(F_{2h}) \approx s\left\{ \sum_{\mathbf{h}'} (F_{\mathbf{h}'}^2 F_{\mathbf{h}+\mathbf{h}'}^2 - N^2) \right\}. \quad (2)$$

It may be shown that $\sum_{\mathbf{h}'} F_{\mathbf{h}'}^2 F_{\mathbf{h}+\mathbf{h}'}^2$ is the \mathbf{h} th Fourier coefficient of P^2 , the squared Patterson function, while ΣN^2 is the contribution to this coefficient of the origin peak of P^2 . We conclude that (2) is then true to the extent that all peaks in P^2 may be ignored except that at the origin and those at $\pm 2\mathbf{r}_j$. In this way the correlation between $s(F_{2h})$ and Σ_4 is seen to be even weaker than that between the former and Σ_1 . The effect of squaring the Patterson function is to decrease, to one quarter, the relative weight of the peaks at $2\mathbf{r}_j$, compared with the peaks at $(\mathbf{r}_i - \mathbf{r}_j)$. This situation cannot be altered by increasing the number of terms in Σ_4 .

Acta Cryst. (1954). **7**, 451

The statistical approach of Hauptman & Karle to the phase problem. By V. VAND and R. PEPIN-SKY, *X-ray and Crystal Analysis Laboratory, Department of Physics, The Pennsylvania State University, State College, Pa., U.S.A.*

(Received 4 March 1954)

The statistical approach to X-ray structure analysis has been treated by Hauptman & Karle (1953), who derived a number of formulae for the probabilities that certain structure factors are positive. However, a peculiar situation arises from their analysis. Although their mathematical approach is valid, and the formulae for the determining quantities Σ_n are correct within the limitations of their derivation, the conclusions, that a practical solution of the phase problem has been attained for all centrosymmetrical structures—provided a sufficient number of structure factors is available—cannot be maintained.

We have constructed more accurate distribution functions using the *Tables* of Vand (1953), and have also formed a simpler mathematical derivation of the statistical formulae. The true distribution functions yield not

Equation (4.05) has not been considered in any detail but it is clear that it, too, embodies the inherent faults of (4.03) and (4.06).

It will also be seen that, even if a number of signs can be determined for the set of indices for which h , k and l are all even, the extension to the other sets of terms will be very tenuous. Only (4.04) remains for this purpose, if we do not take the discredited (4.05) into account, and we have seen that this is equivalent to applying the normal sign relationship $s(\mathbf{h}) = s(\mathbf{h}')s(\mathbf{h}+\mathbf{h}')$. But the only additional information will be the three arbitrary signs which define the choice of origin. The determination of the next sign for a coefficient for which h , k and l are not all even will depend on one sign relationship! There would then possibly be two sign relationships for the next sign determination, and so on. It is our experience that attempts to build a knowledge of signs on too narrow a base inevitably fail unless the crystal structure problem is a very simple one.

We thus conclude that the solution of the phase problem for the centrosymmetrical crystal given by Hauptman & Karle offers nothing new for solving complex structures. While our criticisms have been directed to the procedure suggested for space group $P\bar{1}$, the simple 'Patterson function' interpretation also shows the weakness of that suggested for the only other space group considered by Hauptman & Karle in detail, namely $P2_1/a$.

The example which they have worked out, naphthalene, is far too obliging in many respects. The high symmetry gives extremely large unitary structure factors and there is no doubt that this structure can be solved by a number of previously published direct methods.

References

- COCHRAN, W. (1953). *Acta Cryst.* **6**, 810.
 HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70.
 HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*. A.C.A. Monograph No. 3. Wilmington: The Letter Shop.
 SAYRE, D. (1952). *Acta Cryst.* **5**, 60.

only probabilities, but also the inequalities of Harker & Kasper. The Hauptman & Karle formulae are confirmed as approximations when the number of atoms is large and the magnitudes of structure factors involved are small.

However, when the formulae are applied to a structure of space group $P\bar{1}$, the following difficulties arise: Only the equations (4.03) and (4.06), as numbered by Hauptman & Karle, can be used in the initial step, since these alone contain squares of structure factors only. For N equal atoms these become

$$\begin{aligned} \Sigma_1 &= \frac{1}{4\sqrt{N}} (E_\mu^2 - 1) && \text{for } h = 2h, \\ \Sigma_4 &= \sum \frac{1}{8N\sqrt{N}} (E_\mu^2 - 1)(E_\nu^2 - 1) && \text{for } h = 2h_\mu \pm 2h_\nu, \end{aligned}$$