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**, 375

Probability distributions for atomic coordinates.

By J. KARLE and H. HAUPTMAN, *U. S. Naval Research Laboratory, Washington, D.C., U.S.A.*

(*Received* 16 *September* 1953)

Probability methods have proven to be useful in various aspects of the problem of crystal structure determination. Wilson (1949) was the first to obtain probability distributions for the magnitude of a structure factor and this work has been developed further (Howells, Phillips & Rogers, 1950; Wilson, 1950, 1951; Rogers & Wilson, 1953). More accurate distributions, valid for all space groups, were obtained later (Karle & Hauptman, 1953; Hauptman & Karle, 1953*a*). In the meantime a probable relationship among the phases of certain sets of structure factors had been found by Sayre (1952), Cochran (1952) and Zachariasen (1952). The probability approach has recently led to a solution of the phase problem for all centrosymmetric space groups (Hauptman & Karle, 1953*b*, 1954), and appears to be applicable to non-centrosymmetric space groups as well. Probability distributions for interatomic vectors have also been obtained (Hauptman & Karle, 1952). In view of the past successes of the probability methods, the question may well be asked whether probability distributions for atomic coordinates may not be obtained directly, rather than probability distributions of phases. The purpose of this note is to obtain these distributions for all the space groups.*

We treat first the centrosymmetric crystal. The crystal structure factor $F_{\mathbf{h}}$ is defined by

$$F_{\mathbf{h}} = \sum_{j=1}^{N/n} f_{j\mathbf{h}}\xi_{j\mathbf{h}} , \tag{1}$$

where N is the number of atoms in the unit cell, n is the symmetry number, $f_{j\mathbf{h}}$ is the atomic scattering factor, and $\xi_{j\mathbf{h}}$ is some trigonometric function of the atomic coordinates which depends on the space group, e.g. for $P\bar{1}$, $\xi_{j\mathbf{h}} = 2\cos 2\pi(hx_j+ky_j+lz_j)$.

In order to find the probability distribution of the atomic coordinates x_1, y_1, z_1 of the first atom with respect to a suitably chosen center of symmetry as origin, on the basis that a set of m structure factors $F_{\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$, have the known values $A_{\mathbf{h}_\mu}$, we make use of the relation

$$(\xi_{1\mathbf{h}_1}\ldots\xi_{1\mathbf{h}_m}, A_{\mathbf{h}_1}\ldots A_{\mathbf{h}_m}) = K(A_{\mathbf{h}_1}\ldots A_{\mathbf{h}_m}, \xi_{1\mathbf{h}_1}\ldots\xi_{1\mathbf{h}_m}), \tag{2}$$

where the left member of (2) is the probability that $\xi_{1\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$, lie between $\xi_{1\mathbf{h}_\mu}$ and $\xi_{1\mathbf{h}_\mu}+d\xi_{1\mathbf{h}_\mu}$ after it is known that the structure factors $F_{\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$, have the values $A_{\mathbf{h}_\mu}$; K is a normalizing factor independent of the atomic coordinates x_1, y_1, z_1 and the second factor in the right member of (2) is the probability that $F_{\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$, lie between $A_{\mathbf{h}_\mu}$ and $A_{\mathbf{h}_\mu}+dA_{\mathbf{h}_\mu}$ after the atomic coordinates x_1, y_1, z_1 are known (compare equation (12) of Hauptman & Karle (1952)).

Denote by $P_1(A_1, \ldots, A_m)dA_1\ldots dA_m$ the probability that $F_{\mathbf{h}_\mu}$ lie between A_μ and $A_\mu+dA_\mu$, $\mu = 1, \ldots, m$, after it is known that the coordinates of the first atom are x_1, y_1, z_1. Referring to equations (3·02) and (3·03) or equations (8) and (9) of Hauptman & Karle (1953*b* or 1954 respectively), we infer that

$$P_1(A_1, \ldots, A_m) = \frac{1}{(2\pi)^m}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left[-i\sum_{\mu=1}^{m}(A_\mu - f_{1\mathbf{h}_\mu}\xi_{1\mathbf{h}_\mu})w_\mu\right] \times \prod_{j=2}^{N/n} q(f_{j1}w_1, \ldots, f_{jm}w_m)dw_1\ldots dw_m , \tag{3}$$

where

$$q(f_{j1}w_1, \ldots, f_{jm}w_m) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(\xi_{j\mathbf{h}_1}, \ldots, \xi_{j\mathbf{h}_m}) \times \exp\left(i\sum_{\mu=1}^{m} f_{j\mathbf{h}_\mu}\xi_{j\mathbf{h}_\mu}w_\mu\right) d\xi_{j\mathbf{h}_1}\ldots d\xi_{j\mathbf{h}_m} \tag{4}$$

and $p(\xi_{j\mathbf{h}_1}, \ldots, \xi_{j\mathbf{h}_m})d\xi_{j\mathbf{h}_1}\ldots d\xi_{j\mathbf{h}_m}$ is the probability that $\xi_{j\mathbf{h}_\mu}$ lies between $\xi_{j\mathbf{h}_\mu}$ and $\xi_{j\mathbf{h}_\mu}+d\xi_{j\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$. From (3) and (4) we conclude that $P_1(A_1, \ldots, A_m)$ is proportional to

$$\exp\left(-\sum_{\mu=1}^{m}\frac{(A_{\mathbf{h}_\mu}-f_{1\mathbf{h}_\mu}\xi_{1\mathbf{h}_\mu})^2}{2m_2\sum_{j=2}^{N/n} f_{j\mathbf{h}_\mu}^2}\right), \tag{5}$$

where

$$m_2 = \int_0^1\int_0^1\int_0^1 \xi_{j\mathbf{h}_\mu}^2 dx_j dy_j dz_j , \tag{6}$$

and only the first term of a rapidly converging series is retained. From (2) we infer that (5) is also the probability distribution of the $\xi_{1\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$, after it is known that the values of $F_{\mathbf{h}_\mu}$ are $A_{\mathbf{h}_\mu}$, $\mu = 1, \ldots, m$. It may be concluded that the most probable values of the coordinates x_1, y_1, z_1 of those atoms characterized by the scattering factor f_1 coincide with the principal maxima of (5). Once the coordinates of $k-1$ atoms are known, use of this additional knowledge enables one to replace (5) by

$$\exp\left(-\sum_{\mu=1}^{m}\frac{\left(A_{\mathbf{h}_\mu}-\sum_{j=1}^{k-1} f_{j\mathbf{h}_\mu}\xi_{j\mathbf{h}_\mu}-f_{k\mathbf{h}_\mu}\xi_{k\mathbf{h}_\mu}\right)^2}{2m_2\sum_{j=k+1}^{N/n} f_{j\mathbf{h}_\mu}^2}\right). \tag{7}$$

* In this note we do not treat the case that atoms may occur in special positions. Should these methods prove useful, the extension to special positions may be readily carried out.

If the phases are unknown, and we assume that the sign of every $F_{\mathbf{h}}$ is just as likely to be plus as minus, (5) is replaced by

$$\prod_{\mu=1}^{m}\left\{\exp\left(-\frac{(|A_{\mathbf{h}_\mu}|-f_{1\mathbf{h}_\mu}\xi_{1\mathbf{h}_\mu})^2}{2m_2\sum_{j=2}^{N/n}f_{j\mathbf{h}_\mu}^2}\right)+\exp\left(-\frac{(|\overline{A_{\mathbf{h}_\mu}}|+f_{1\mathbf{h}_\mu}\xi_{1\mathbf{h}_\mu})^2}{2m_2\sum_{j=2}^{N/n}f_{j\mathbf{h}_\mu}^2}\right)\right\}, \quad (8)$$

a formula which, for space group $P\bar{1}$, may be compared with equation (78) of Hauptman & Karle (1952). Now, the principal maxima of (8) determine the most probable values of x_1, y_1, z_1. For a fixed form for ξ, there is an eightfold, fourfold, or twofold ambiguity in the choice of origin according as the crystal is of Category 1, 2 or 3 (Hauptman & Karle, 1953*b*). In any event, selecting any set of values for x_1, y_1, z_1 as given by (8), fixes the origin. Then the analogue of (8) derived from (7) with $k = 2$ is the probability distribution for the coordinates of any atom referred to a unique origin determined by the chosen values for x_1, y_1, z_1 and based upon the knowledge of a set of magnitudes.

The same methods apply to non-centrosymmetric crystals. The analogue of (7) is

$$\exp\left\{\sum_{\mu=1}^{m}\frac{-\left(A_{\mathbf{h}_\mu}-\sum_{j=1}^{k-1}f_{j\mathbf{h}_\mu}\xi_{j\mathbf{h}_\mu}-f_{k\mathbf{h}_\mu}\xi_{k\mathbf{h}_\mu}\right)^2-\left(B_{\mathbf{h}_\mu}-\sum_{j=1}^{k-1}f_{j\mathbf{h}_\mu}\eta_{j\mathbf{h}_\mu}-f_{k\mathbf{h}_\mu}\eta_{k\mathbf{h}_\mu}\right)^2}{2m_2\sum_{j=k+1}^{N/n}f_{j\mathbf{h}_\mu}^2}\right\}, \quad (9)$$

where the crystal structure factor is given by

$$F_{\mathbf{h}_\mu}=X_{\mathbf{h}_\mu}+iY_{\mathbf{h}_\mu}, \quad (10)$$

$$X_{\mathbf{h}_\mu}=\sum_{j=1}^{N/n}f_{j\mathbf{h}_\mu}\xi_{j\mathbf{h}_\mu}, \quad (11)$$

$$Y_{\mathbf{h}_\mu}=\sum_{j=1}^{N/n}f_{j\mathbf{h}_\mu}\eta_{j\mathbf{h}_\mu}, \quad (12)$$

$$m_2=\int_0^1\int_0^1\int_0^1\xi_{j\mathbf{h}_\mu}^2dx_jdy_jdz_j=\int_0^1\int_0^1\int_0^1\eta_{j\mathbf{h}_\mu}^2dx_jdy_jdz_j, \quad (13)$$

and the values of the real and imaginary parts $X_{\mathbf{h}_\mu}$ and $Y_{\mathbf{h}_\mu}$ of the structure factor $F_{\mathbf{h}_\mu}$ are known to be $A_{\mathbf{h}_\mu}$ and $B_{\mathbf{h}_\mu}$ respectively.

If the phases $\varphi_{\mathbf{h}_\mu}$ of $F_{\mathbf{h}_\mu}$ are unknown, we assume that all values of $\varphi_{\mathbf{h}_\mu}$ between 0 and 2π are equally probable. Transforming to polar coordinates,

$$A_{\mathbf{h}_\mu}=R_{\mathbf{h}_\mu}\cos\varphi_{\mathbf{h}_\mu}, \quad (14)$$

$$B_{\mathbf{h}_\mu}=R_{\mathbf{h}_\mu}\sin\varphi_{\mathbf{h}_\mu}, \quad (15)$$

where $R_{\mathbf{h}_\mu}$ is the known magnitude of $F_{\mathbf{h}_\mu}$, and integrating over $\varphi_{\mathbf{h}_\mu}$ we finally obtain the analogue of (8):

$$\exp\left(-\sum_{\mu=1}^{m}\frac{R_{\mathbf{h}_\mu}^2+f_{1\mathbf{h}_\mu}^2(\xi_{1\mathbf{h}_\mu}^2+\eta_{1\mathbf{h}_\mu}^2)}{2m_2\sum_{j=2}^{N/n}f_{j\mathbf{h}_\mu}^2}\right)\times\prod_{\mu=1}^{m}I_0\left(\frac{R_{\mathbf{h}_\mu}f_{1\mathbf{h}_\mu}\sqrt{(\xi_{1\mathbf{h}_\mu}^2+\eta_{1\mathbf{h}_\mu}^2)}}{m_2\sum_{j=2}^{N/n}f_{j\mathbf{h}_\mu}^2}\right), \quad (16)$$

where I_0 is the Bessel function of imaginary argument. For space group $P1$, (16) is independent of x_1, y_1, z_1, in agreement with the known fact that the origin may be arbitrarily specified, i.e. the coordinates of the first atom may be arbitrarily chosen. Once this is done, however, the case $k = 2$ of (9) may be obtained and then the analogue of (16) found, where now the origin is specified. Since, for non-centrosymmetric crystals, there are always two homometric structures, related by inversion through a point, which are solutions to the problem, there still remains a twofold ambiguity in the coordinates of the second atom. Once one of these sets of coordinates is chosen the ambiguity is resolved and the cases $k = 3, 4, \ldots$ of (9) determine all other atomic coordinates uniquely. Again, for space group $P2$,

$$\xi_{1\mathbf{h}_\mu}=2\cos 2\pi(h_\mu x_1+k_\mu y_1)\cos 2\pi l_\mu z_1, \quad (17)$$

$$\eta_{1\mathbf{h}_\mu}=2\cos 2\pi(h_\mu x_1+k_\mu y_1)\sin 2\pi l_\mu z_1, \quad (18)$$

so that (16) determines, except for a fourfold ambiguity, the values of x_1 and y_1, but is independent of z_1 (again in agreement with the nature of the arbitrariness in choosing the origin for this space group). If we choose any of the four permitted values for x_1, y_1, and specify z_1 arbitrarily, thus fixing the origin uniquely, we may obtain the case $k = 2$ of (9) and proceed to find the analogue of (16). The homometric ambiguities are resolved as in space group $P1$. Similar remarks apply to any non-centrosymmetric space group.

The usefulness of these formulas is as yet undetermined. In the first place the equations are sufficiently complex that practical application may be limited. In addition, the assumption of the uniform distribution of the phases used in deriving (8) and (16), and the retention of only the first term in the infinite series, for the probability distributions, may lead to inaccuracies in atomic positions. Nevertheless, owing to the importance of the possibility of the direct location of atomic coordinates, the approach outlined here merits further study.

References

Cochran, W. (1952). *Acta Cryst.* **5**, 65.

Hauptman, H. & Karle, J. (1952). *Acta Cryst.* **5**, 48.

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

Hauptman, H. & Karle, J. (1953*b*). A.C.A. Monograph No. 3. Ann Arbor: Edwards.

Hauptman, H. & Karle, J. (1954). *Acta Cryst.* **7**, 369.

Howells, E. R., Phillips, D. C. & Rogers, D. (1950). *Acta Cryst.* **3**, 210.

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

Rogers, D. & Wilson, A. J. C. (1953). *Acta Cryst.* **6**, 439.

Sayre, D. (1952). *Acta Cryst.* **5**, 60.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318.

Wilson, A. J. C. (1950). *Acta Cryst.* **3**, 297.

Wilson, A. J. C. (1951). *Research, Lond.* **4**, 141.

Zachariasen, W. H. (1952). *Acta Cryst.* **5**, 68.