(that is, at $2\pi r_0\xi = 3$, where r_0 is the radius of the helix) and equation (7) is therefore a little too stringent when applied to this case.

An interesting feature of Fig. 3(*b*) is that distinct fringes appear in the regions in which Fig. 2(*a*) and Fig. 2(*b*) are identical, but not in all other regions. In particular, there are no fringes on the fourth layer line, since one of the diffraction patterns contributes nothing to this layer line. Thus the pattern in Fig. 3(*b*) tells us something about the variation of diffracted amplitude with ψ as well as giving the mean intensity as ψ is varied.

Reference

Hughes, W. & Taylor, C. A. (1953). *J. Sci. Instrum.* **30**, 105.

Acta Cryst. (1955). **8**, 29

Some Relations between Structure Factors

By R. K. Bullough and D. W. J. Cruickshank

School of Chemistry, The University, Leeds 2, England

(*Received* 30 *September* 1954 *and in revised form* 19 *October* 1954)

For a centrosymmetric structure of all like atoms, a series of relations between structure factors is derived. The relationship of these results to the Hauptman & Karle statistical formulae for structure-factor sign determination is indicated; but an important difference in the statistical weight of one formula leads to a relation, recently given also by Cochran, for the signs and magnitudes of certain structure factors.

Much attention is focused at the present time on the application of statistical methods to the derivation of relationships between structure factors. Some of these relationships, or their equivalents, can, however, be derived by direct trigonometrical manipulation when the structure consists of all like atoms. For small departures of the atomic scattering factors from their mean, the relationships remain statistically true for a structure of unlike atoms. In this note a set of relations similar in form to those derived by Hauptman & Karle (1953, 1954) will be derived trigonometrically.

In $P\bar{1}$ the structure factor $F_{\mathbf{h}}$ is given by

$$F_{\mathbf{h}} = 2\sum_{j=1}^{N/2} f_{j\mathbf{h}} \cos(2\pi\mathbf{h}.\mathbf{x}_j)\ ,$$

where $\mathbf{h}.\mathbf{x}_j = hx_j+ky_j+lz_j$.

A normalized structure factor $E_{\mathbf{h}}$ may then be defined by

$$E_{\mathbf{h}} = F_{\mathbf{h}}\Big/\Big(\sum_{j=1}^{N} f_{j\mathbf{h}}^2\Big)^{\frac{1}{2}}.$$

When all the atoms in the cell are of like kind, in $P\bar{1}$,

$$E_{\mathbf{h}} = \frac{2}{N^{\frac{1}{2}}}\sum_{j=1}^{N/2}\cos(2\pi\mathbf{h}.\mathbf{x}_j)\ .$$

Hence

$$E_{\mathbf{L}}^2 = \frac{4}{N}\sum_{i=1}^{N/2}\sum_{j=1}^{N/2}\cos(2\pi\mathbf{L}.\mathbf{x}_i)\cos(2\pi\mathbf{L}.\mathbf{x}_j)\ ,$$

and therefore

$$N^{\frac{1}{2}}(E_{\mathbf{L}}^2-1) = E_{2\mathbf{L}}+\frac{8}{N^{\frac{1}{2}}}S_{\mathbf{L}}\ , \qquad [A]$$

where

$$S_{\mathbf{L}} = \sum_{i>j}\cos(2\pi\mathbf{L}.\mathbf{x}_i)\cos(2\pi\mathbf{L}.\mathbf{x}_j)\ .$$

[A] has been given in terms of $F_{\mathbf{h}}/f_{\mathbf{h}}$ by Cochran & Woolfson (1954). $(E_{\mathbf{L}}^2-1)$ is a sharpened Patterson coefficient which [A] shows to be composed of contributions from vector peaks relating to vectors through the centre of symmetry via $E_{2\mathbf{L}}$, and to the remaining vectors via $S_{\mathbf{L}}$.

Hughes (1953) has derived Sayre's (1952) relation in terms of unitary structure factors. In terms of normalized structure factors this relation is

$$\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}\rangle_{\mathbf{h}_1+\mathbf{h}_2=\mathbf{L}} = \frac{1}{N^{\frac{1}{2}}}E_{\mathbf{L}}\ , \qquad [B]$$

where the left-hand side is the arithmetical mean of all products $E_{\mathbf{h}_1}E_{\mathbf{h}_2}$, formed subject to the condition $\mathbf{h}_1+\mathbf{h}_2 = \mathbf{L}$.

Hughes's method may be extended to derive relationships between intensities. By [A]

$$N(E_{\mathbf{h}_1}^2-1)(E_{\mathbf{h}_2}^2-1) = E_{2\mathbf{h}_1}E_{2\mathbf{h}_2}+\frac{8}{N^{\frac{1}{2}}}(E_{2\mathbf{h}_1}S_{\mathbf{h}_2}+E_{2\mathbf{h}_2}S_{\mathbf{h}_1})+\frac{64}{N}S_{\mathbf{h}_1}S_{\mathbf{h}_2}\ ,$$

and therefore

$$N\langle(E_{\mathbf{h}_1}^2-1)(E_{\mathbf{h}_2}^2-1)\rangle = \langle E_{2\mathbf{h}_1}E_{2\mathbf{h}_2}\rangle + \frac{8}{N^{\frac{1}{2}}}(\langle E_{2\mathbf{h}_1}S_{\mathbf{h}_2}\rangle+\langle E_{2\mathbf{h}_2}S_{\mathbf{h}_1}\rangle)+\frac{64}{N}\langle S_{\mathbf{h}_1}S_{\mathbf{h}_2}\rangle\ , \qquad (1)$$

where the means are formed subject to $2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}$.

$$2E_{2\mathbf{h}_1}S_{\mathbf{h}_2} = \frac{4}{N^{\frac{1}{2}}}\left\{\sum_{j=1}^{N/2}\cos(2\pi 2\mathbf{h}_1.\mathbf{x}_j)\right\}\left\{\sum_{i>j}\cos(2\pi\mathbf{h}_2.\mathbf{x}_i)\cos(2\pi\mathbf{h}_2.\mathbf{x}_j)\right\}$$
$$= \frac{1}{N^{\frac{1}{2}}}\sum_{i\neq j}\cos\{2\pi(\overline{2\mathbf{h}_1\pm\mathbf{h}_2}.\mathbf{x}_i\pm\mathbf{h}_2.\mathbf{x}_j)\}$$
$$+\frac{1}{N^{\frac{1}{2}}}\sum_{\substack{s\\ s\neq i,j}}\sum_{i>j}\cos\{2\pi(2\mathbf{h}_1.\mathbf{x}_s\pm\mathbf{h}_2.\overline{\mathbf{x}_i\pm\mathbf{x}_j})\}\,,$$

where all four $\pm$, $\pm$ combinations are to be included. Averaged over the infinite set $2\mathbf{h}_1+2\mathbf{h}_2 = 2\mathbf{L}$, all cosines vanish except such as are of the form

$$\cos(2\pi q\overline{\mathbf{h}_1+\mathbf{h}_2}.\mathbf{r})\,,$$

where q is a numerical constant. Accordingly,

$$\langle E_{2\mathbf{h}_1}S_{\mathbf{h}_2}\rangle = \langle E_{2\mathbf{h}_2}S_{\mathbf{h}_1}\rangle = 0\,. \tag{2}$$

$S_{\mathbf{h}_1}S_{\mathbf{h}_2}$ can be similarly expanded, but much tedium is avoided by noting that the only cosines of the form $\cos(2\pi q\overline{\mathbf{h}_1+\mathbf{h}_2}.\mathbf{r})$ arising from

$$\{\sum_{i>j}\cos(2\pi\mathbf{h}_1.\mathbf{x}_i)\cos(2\pi\mathbf{h}_1.\mathbf{x}_j)\}\times\{\sum_{i>j}\cos(2\pi\mathbf{h}_2.\mathbf{x}_i)\cos(2\pi\mathbf{h}_2.\mathbf{x}_j)\}$$

are contained in

$$\sum_{i>j}\cos(2\pi\mathbf{h}_1.\mathbf{x}_i)\cos(2\pi\mathbf{h}_2.\mathbf{x}_i)\times\cos(2\pi\mathbf{h}_1.\mathbf{x}_j)\cos(2\pi\mathbf{h}_2.\mathbf{x}_j)\,.$$

Thus, subject to $2\mathbf{h}_1+2\mathbf{h}_2 = 2\mathbf{L}$,

$$4\langle S_{\mathbf{h}_1}S_{\mathbf{h}_2}\rangle = \langle\sum_{i>j}\cos(2\pi\overline{\mathbf{h}_1+\mathbf{h}_2}.\mathbf{x}_i)\cos(2\pi\overline{\mathbf{h}_1+\mathbf{h}_2}.\mathbf{x}_j)\rangle = S_{\mathbf{L}}\,. \tag{3}$$

By (2), (3) and [B], relation (1) becomes

$$N^{3/2}\langle(E^2_{\mathbf{h}_1}-1)(E^2_{\mathbf{h}_2}-1)\rangle_{2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}} = E_{2\mathbf{L}}+\frac{16}{N^{\frac{1}{2}}}S_{\mathbf{L}}\,. \tag{D}$$

Similarly, by using [B] and a relation similar to (2),

$$N\langle(E^2_{\mathbf{h}_1}-1)(E_{\mathbf{h}_2})\rangle_{2\mathbf{h}_1+\mathbf{h}_2=\mathbf{L}} = E_{\mathbf{L}}\,. \tag{C}$$

Only the relations [B] and [C] yield information about structure factors not of the form $E_{2\mathbf{L}}$. Each of [A], [B], [C] and [D] can give information about the $E_{2\mathbf{L}}$.

Relations equivalent to [A], ..., [D] above can be developed for other space groups. In $P2_1/a$

$$E_{\mathbf{h}_1} = \frac{4}{N^{\frac{1}{2}}}\sum_{i=1}^{N/4}\cos 2\pi\left(h_1x_i+l_1z_i+\frac{h_1+k_1}{4}\right)\times\cos 2\pi\left(k_1y_i-\frac{h_1+k_1}{4}\right)\,,$$

and the form of the structure factor is such that it is also possible to average separately over the vectors (h_1, K, l_1) and (H, k_1, L), where $\mathbf{L} = (H, K, L)$ is a constant vector. For example, in this space group

$$N^{\frac{1}{2}}\langle(E^2_{\mathbf{h}_1}-1)(-1)^{H+k_1}\rangle_{\substack{2h_1=2H\\2l_1=2L}} = E_{2H,0,2L} \tag{4a}$$

and

$$N^{\frac{1}{2}}\langle(-1)^{h_1+K}(E^2_{\mathbf{h}_1}-1)\rangle_{2k_1=2K} = E_{0,2K,0}\,; \tag{4b}$$

whilst the equivalent of [D] is

$$N^{3/2}\langle(E^2_{\mathbf{h}_1}-1)(E^2_{\mathbf{h}_2}-1)\rangle_{2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}} = E_{2H,2K,2L}+(-1)^{H+K}E_{2H,0,2L}+(-1)^{H+K}E_{0,2K,0}+\frac{64}{N^{\frac{1}{2}}}T_{HKL}\,,$$

where T_{HKL} is an equivalent of $S_{\mathbf{L}}$ above. Equation (4a) should be compared with relation (4·41) of Hauptman & Karle (1953).

The relations [B], [C] and [D] are identities for an infinity of data. Any practical use of these relations involves an estimate of the reliability of a mean in [B], [C] or [D] calculated from a finite set of data. A single product $N^{\frac{1}{2}}E_{\mathbf{h}_1}E_{\mathbf{h}_2}$ can, for example, be expanded and becomes

$$N^{\frac{1}{2}}E_{\mathbf{h}_1}E_{\mathbf{h}_2} = E_{\mathbf{h}_1+\mathbf{h}_2}+\frac{2}{N^{\frac{1}{2}}}\sum_{i=1}^{N/2}\cos(2\pi\overline{\mathbf{h}_1-\mathbf{h}_2}.\mathbf{x}_i)+\frac{2}{N^{\frac{1}{2}}}\sum_{\substack{i=1\\i\neq j}}^{N/2}\sum_{j=1}^{N/2}\cos(2\pi\overline{\mathbf{h}_1.\mathbf{x}_i+\mathbf{h}_2.\mathbf{x}_j})+\frac{2}{N^{\frac{1}{2}}}\sum_{\substack{i=1\\i\neq j}}^{N/2}\sum_{j=1}^{N/2}\cos(2\pi\overline{\mathbf{h}_1.\mathbf{x}_i-\mathbf{h}_2.\mathbf{x}_j})\,. \tag{5}$$

Any correlation between cosines on the right of (5) makes no covariant contribution to the variance associated with a single product $N^{\frac{1}{2}}E_{\mathbf{h}_1}E_{\mathbf{h}_2}$. Each cosine not of the form $\cos 2\pi q\overline{\mathbf{h}_1+\mathbf{h}_2}.\mathbf{r}$ then contributes $\frac{1}{2}$ to this variance. The total variance of a single product $N^{\frac{1}{2}}E_{\mathbf{h}_1}E_{\mathbf{h}_2}$ is thus

$$1+2(N/2-1) = N-1\,.$$

If the ν_2 products $N^{\frac{1}{2}}E_{\mathbf{h}_1}E_{\mathbf{h}_2}$ are assumed to be independent estimates of their mean, the variance associated with that mean is

$$(N-1)/\nu_2\sim N/\nu_2$$

in the case of large N. Finally, using similar arguments, for large N the variance associated with an estimate of $E_{2\mathbf{L}}$ through [A] is $2N$; of $E_{2\mathbf{L}}$ through [B], N/ν_2; of $E_{2\mathbf{L}}$ through [C], $2N^2/\nu_3$; and of $E_{2\mathbf{L}}+(16/N^{\frac{1}{2}})S_{\mathbf{L}}$ through [D], $4N^3/\nu_4$.

The variance associated with $E_{2\mathbf{L}}$ in [A] is that calculated on the assumption that $\mathbf{L}$ varies; and it differs in this respect from the variances in [B], [C] and [D]. It is to be emphasized that the mean in [D] converges to $E_{2\mathbf{L}}+(16/N^{\frac{1}{2}})S_{\mathbf{L}}$; but if [D] is to be used to obtain an estimate of $E_{2\mathbf{L}}$ alone, then, as $\mathbf{L}$ varies, a further contribution $8N$, independent of ν_4, is to be included in the associated variance. If the ν_2, ν_3 and ν_4 products in [B], [C] and [D] do not contain the same data, [A], ..., [D] may each be used to give separate estimates of $E_{2\mathbf{L}}$. A weighted mean estimate of the sign of $E_{2\mathbf{L}}$ is then given by the sign of $\Xi_1+\Xi_2+\Xi_3+\Xi_4$, where

$$\Xi_1 = \frac{1}{2N}\cdot N^{\frac{1}{2}}(E_{\mathbf{L}}^2-1) = \frac{1}{2N^{\frac{1}{2}}}(E_{\mathbf{L}}^2-1)\ ;$$

$$\Xi_2 = \frac{\nu_2}{N}\cdot N^{\frac{1}{2}}\langle E_{\mathbf{h}_1}E_{\mathbf{h}_2}\rangle_{\mathbf{h}_1+\mathbf{h}_2=2\mathbf{L}} = \frac{1}{N^{\frac{1}{2}}}\sum_{\mathbf{h}_1+\mathbf{h}_2=2\mathbf{L}} E_{\mathbf{h}_1}E_{\mathbf{h}_2}\ ;$$

$$\Xi_3 = \frac{\nu_3}{2N^2}\cdot N\langle(E_{\mathbf{h}_1}^2-1)E_{\mathbf{h}_2}\rangle_{2\mathbf{h}_1+\mathbf{h}_2=2\mathbf{L}} = \frac{1}{2N}\sum_{2\mathbf{h}_1+\mathbf{h}_2=2\mathbf{L}}(E_{\mathbf{h}_1}^2-1)E_{\mathbf{h}_2}\ ;$$

$$\Xi_4 = \frac{1}{4N^3/\nu_4+8N}\cdot N^{3/2}\langle(E_{\mathbf{h}_1}^2-1)(E_{\mathbf{h}_2}^2-1)\rangle_{2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}}\ .$$

If $8N \ll \frac{4N^3}{\nu_4}$, Ξ_4 becomes

$$\Xi_4 = \frac{\nu_4}{4N^3}\cdot N^{3/2}\langle(E_{\mathbf{h}_1}^2-1)(E_{\mathbf{h}_2}^2-1)\rangle_{2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}}$$
$$= \frac{1}{4N^{3/2}}\sum_{2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}}(E_{\mathbf{h}_1}^2-1)(E_{\mathbf{h}_2}^2-1)\ .$$

The set $\Xi_i (i = 1, 2, 3)$, together with the second form of Ξ_4, should be compared with the Σ_i, evaluated for all like atoms, of Hauptman & Karle (1954).

A much better weighted mean estimate of $E_{2\mathbf{L}}$ can be obtained, however, by allowing for the correlation of the $S_{\mathbf{L}}$ terms in [A] and [D]. Omitting [B] and [C] for simplicity, and using the symbols [A] and [D] to denote just the left-hand sides of those equations, the best estimate of $E_{2\mathbf{L}}$ will be given by

$$\frac{[A]+k[D]}{1+k} = q \text{ (say)}\ , \tag{6}$$

where k is chosen to give minimum variance to q. Using the above variances of [A] and [D], ($2N$ and $(4N^3/\nu_4)+8N$), and their covariance $4N$, q is found to be the best estimate of $E_{2\mathbf{L}}$ when

$$k = -\tfrac{1}{2}\frac{1}{1+N^2/\nu_4}\ ; \tag{7}$$

the variance of q is then

$$\frac{4N^3/\nu_4+8N^5/\nu_4^2}{(1+2N^2/\nu_4)^2}\ . \tag{8}$$

As $\nu_4 \to 0$, $k \to 0$ and the variance $\to 2N$, so that the emphasis is correctly put on [A]. For large ν_4 the variance is $4N^3/\nu_4$, and as $\nu_4 \to \infty$, $k \to -\frac{1}{2}$ and the variance $\to 0$, giving the exact result

$$2N^{\frac{1}{2}}(E_{\mathbf{L}}^2-1)-N^{3/2}\langle(E_{\mathbf{h}_1}^2-1)(E_{\mathbf{h}_2}^2-1)\rangle_{2\mathbf{h}_1+2\mathbf{h}_2=2\mathbf{L}} = E_{2\mathbf{L}}\ . \tag{E}$$

This limiting result may, of course, be written down directly by eliminating $S_{\mathbf{L}}$ between [A] and [D]. In $P2_1/a$, a similar relation can be derived only by making use of additional relations of the type of (4).

Relation [E] has been derived from a different point of view by Cochran (1954), who has interpreted it in terms of the subtraction of a squared Patterson function from twice the normal Patterson function and has shown that it is valid only when the Patterson formed from the available data is completely resolved. In the present derivation the latter restriction corresponds to the failure of the averaging process in (2) and (3) for two close vectors and a finite set of data of low limiting reciprocal radius. Similarly, (7) and (8) are correct only when the ν_4 reflexions are a random selection from the whole of reciprocal space. However, it should be possible to obtain useful results when some of the Patterson peaks are unresolved, especially since any routine procedure based on (6) or [E] would also incorporate [B] or [C] on the lines suggested by Hauptman & Karle (1953, 1954). The number of unresolved Patterson peaks will increase with N, so that such a procedure will fall off in efficacy as the structures considered become more complex. This is also apparent from the fact that, on assuming the number ν of reflexions available to be approximately proportional to N, the variances of the estimates of $E_{2\mathbf{L}}$ through [A], [C], [D] and (6) all increase as powers of N. It is true that a process of sign determination starts with the largest $|E_{\mathbf{h}}|$'s and that the probable largest value of $|E_{\mathbf{h}}|$ increases with N since there are now more reflexions; but this largest value of $|E_{\mathbf{h}}|$ increases more slowly than a power of N, and therefore the increasing variances predominate.

For a given structure it is always possible to check whether formulae of the type (6) or [E] are likely to be useful in sign determination, since they have the merit of predicting both the sign and the magnitude of $E_{2\mathbf{L}}$ from a set of $|E_{\mathbf{L}}|$. Approximately one-eighth of the predicted $E_{2\mathbf{L}}$ can be compared with the measured $|E_{\mathbf{L}}|$ in magnitude and some *a posteriori* estimate of the reliability of a prediction given. Such *a posteriori* estimates will supplement the *a priori* estimates of the reliability obtained from (8) which will be in error when there are unresolved Patterson peaks. Good agreement between measured and estimated $|E_{2\mathbf{L}}|$'s is unlikely unless ν_4 is large.

References

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

COCHRAN, W. & WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 450.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* A. C. A. Monograph No. 3. Wilmington: The Letter Shop.

HAUPTMAN, H. & KARLE, J. (1954). *Acta Cryst.* **7**, 369.

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

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