Acta Cryst. (1957). **10**, 489

A Method for the Systematic Application of Sign Relations

By D. F. Grant, R. G. Howells and D. Rogers

Viriamu Jones Laboratory, University College, Cardiff, Wales

(*Received* 27 *April* 1956 *and in revised form* 11 *December* 1956)

A method is described for the systematic application of the sign relation, $S(\mathbf{h}')S(\mathbf{h}+\mathbf{h}') \approx S(\mathbf{h})$, to projections with symmetry *pgg*, *pmg*, and *p4g*. The procedure allows large amounts of data to be handled compactly and does not require a computer. It involves the recognition and correlation of sign 'coincidences', which are defined by using a variant, $S(\mathbf{A})S(\mathbf{B})S(\mathbf{C}) \approx +1$, of the sign relation. As a consequence of the operation of symmetry elements, the product $S(\mathbf{A})S(\mathbf{B})$ can enter into such a relation with several different third terms, $S(\mathbf{C}_1)$, $S(\mathbf{C}_2)$, etc., the signs of which are said to 'coincide'. From these coincidences it is possible to deduce tentatively the signs of a considerable number of terms. An iterative process is used to adjust and extend the set of tentative signs so as to ensure optimal agreement with the original data.

The solving of two structures, harmine ($C_{13}H_{12}N_2O$) and hydroxydihydroeremophilone ($C_{15}H_{24}O_2$), by this method is described, and details are given of the results produced by EDSAC when the harmine data were submitted to the Cochran–Douglas procedure. The ability of the new method to reach a single optimum solution (in contrast to EDSAC) is discussed.

1. Introduction

It was shown by Sayre (1952) in the account of his 'Squaring method' that the structure factors of a crystal composed of identical resolved atoms satisfy the relations

$$\sum_{\mathbf{h}'} F(\mathbf{h}')F(\mathbf{h}+\mathbf{h}') = v\varphi(\mathbf{h})F(\mathbf{h}) \,, \qquad (1)$$

where $\varphi(\mathbf{h})$ is a function to account for the change in atomic shape in the squared structure and v is the volume of the unit cell. By considering all the *products* $F(\mathbf{h}')F(\mathbf{h}+\mathbf{h}')$ for every structure factor or *term*, $F(\mathbf{h})$, it is theoretically possible to choose a set of signs (term signs) which will make every equation (1) true. In practice, even for a projection this task is enormous. Moreover, it would appear that unless quite a number of the contributing products $F(\mathbf{h}')F(\mathbf{h}+\mathbf{h}')$ are sufficiently large to dominate the sign of the corresponding term, $F(\mathbf{h})$, it will be virtually impossible to arrive at a self-consistent set of term signs, even for a small set. The method has the merit, however, of using all the information available.

Cochran (1952) and others (Zachariasen, 1952; Hughes, 1953; Hauptman & Karle, 1953; Cochran & Woolfson, 1955) have shown that the signs of three unitary structure factors $U(\mathbf{h})$, $U(\mathbf{h}')$, $U(\mathbf{h}+\mathbf{h}')$ are probably related by the equation,

$$S(\mathbf{h}')S(\mathbf{h}+\mathbf{h}') \approx S(\mathbf{h}) \,, \qquad (2)$$

and, for the conditions in which (1) is valid, the probability of (2), in Cochran & Woolfson's notation, is,

$$p = \tfrac{1}{2}+\tfrac{1}{2}\tanh\,[N|U(\mathbf{h})U(\mathbf{h}')U(\mathbf{h}+\mathbf{h}')|] \,. \qquad (3)$$

Relation (2) was used by some of these authors to determine the signs of additional terms once the signs of some of the largest unitary structure factors were known, usually from inequalities. They made no use of the magnitudes of the U's, except in so far as the application of (2) was usually limited to those terms for which $|U| \gtrsim 0{\cdot}3$. Recently, however, Cochran & Douglas (1955) and Vand & Pepinsky (1956) have described methods incorporating the individual probabilities—at least approximately. They used computers to determine the sets of term signs for which $\chi = \Sigma Y(\mathbf{h}, \mathbf{h}')P(\mathbf{h}, \mathbf{h}')$ is large and positive. $Y(\mathbf{h}, \mathbf{h}')$ is defined as $S(\mathbf{h})S(\mathbf{h}')S(\mathbf{h}+\mathbf{h}')$, and $P(\mathbf{h}, \mathbf{h}')$ is defined as $|U(\mathbf{h})U(\mathbf{h}')U(\mathbf{h}+\mathbf{h}')|$. The correct set gives a value of χ, which, although probably among the largest of these sums, may not be the greatest. The correct set is not found uniquely, but must be sought from a number of possible sets.

The method described below was developed to adhere as closely as possible to Sayre's original proposal, but to reduce the labour by concentrating only on the more important terms. It has been arranged to be almost mechanical and largely self-checking, and is offered for use where there is no computer, for, although a 'manual' method, it has proved to be convenient and speedy. Moreover, for harmine, where a direct comparison has been made with the Cochran–Douglas method, it was found capable of handling more data than the computer, and gave a single solution sufficiently near the truth for the corresponding structure to be solved.

An important basic concept of this paper is that of sign 'coincidences', which are defined by using a variant of (2):

$$S(\mathbf{A})S(\mathbf{B})S(\mathbf{C}) \approx +1 \,. \qquad (2a)$$

The product $S(\mathbf{A})S(\mathbf{B})$ can enter into such a relation with several* different third terms, $S(\mathbf{C}_1)$, $S(\mathbf{C}_2)$, etc. and the signs of these are said to 'coincide'; i.e.

$$[S(\mathbf{C}_1) \approx S(\mathbf{C}_2) \ldots]_{\mathbf{A},\mathbf{B}} \,. \tag{4}$$

§ 2 describes the preliminary process of collecting and sorting the many products,

$$S(\mathbf{h}')S(\mathbf{h}+\mathbf{h}')|U(\mathbf{h}')U(\mathbf{h}+\mathbf{h}')| \,,$$

obtained. Comparatively few of these, however, lead to coincidences. § 3 describes a routine for recognizing and correlating the coincidences. § 4 describes an iterative procedure which, in principle at least, can be considered as equivalent to a converse process, represented by

$$S(\mathbf{C}) \approx S[\sum S(\mathbf{A})S(\mathbf{B})|U(\mathbf{A})U(\mathbf{B})|]_{\mathbf{C}} \,, \tag{5}$$

which is closely related to the quantity χ evaluated by Cochran & Douglas, and by Vand & Pepinsky, and is, in fact, used by Cochran & Douglas as a 'further test' (see their equation (3)). It serves to correct and extend the list of term signs deduced from the coincidences so as to secure the best agreement with the larger body of sign relations.

Although less information is used than in Sayre's method, the labour is now within reasonable bounds. Considerably more information is, however, brought to bear on the problem than can be used in Cochran's sign relation (2), and more than has so far been reported to have been fed into a computer, as in the Cochran–Douglas and Vand–Pepinsky methods. As shown below, it does not in favourable circumstances require a prior knowledge of any signs.

2. The method of determining the products

To describe the method, its application to the determination of signs in a principal zone (0*kl*) of the space group $P2_12_12_1$ will be considered, i.e. in the centrosymmetric plane group *pgg*. (The method applies equally to the plane groups *pmg* and *p*4*g*.) The crystal structures of both harmine (Howells, 1954) and hydroxydihydroeremophilone (Grant, 1957) have been determined from the solution of such a projection, and a brief outline of the method was presented by Howells to the Paris Congress, 1954, when describing the solving of harmine.

2·1. Unitary structure factors are determined by the use of the expression

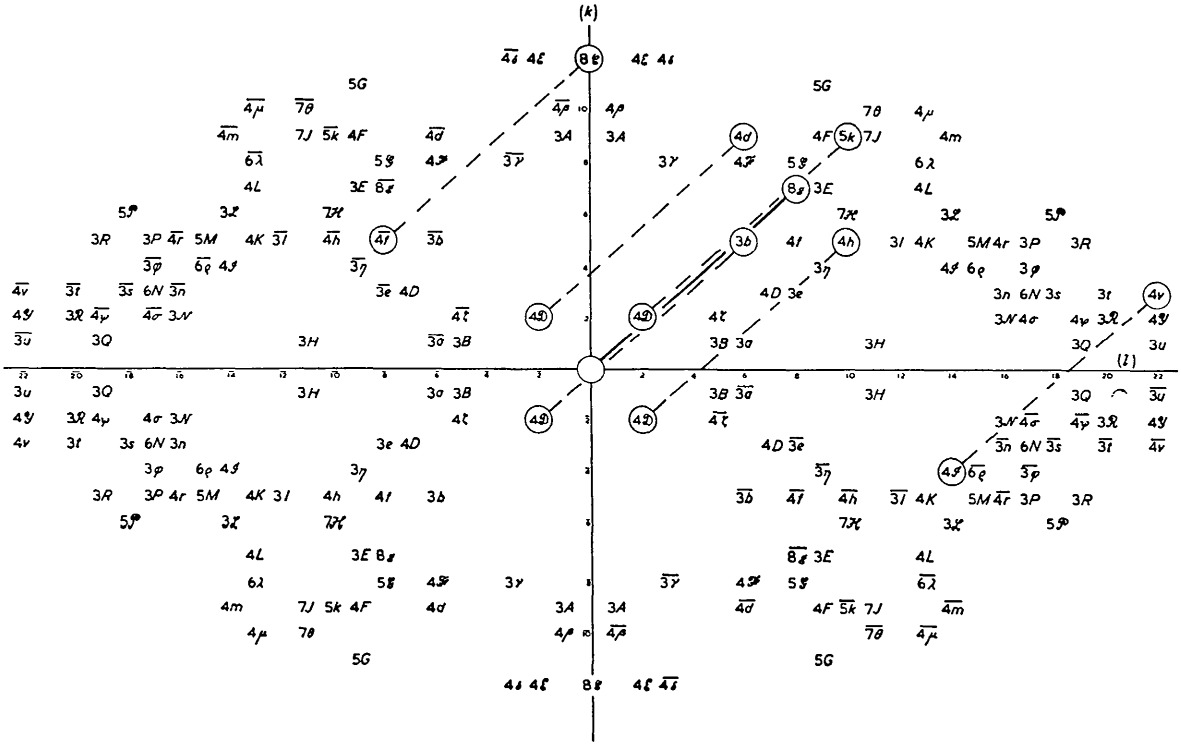

Fig. 1. The square lattice for the (100) projection of hydroxydihydroeremophilone. The letters denote $S(\mathbf{h})$ and the numbers $20|U(\mathbf{h})|$ rounded off to the nearest integer. The vector corresponding to the (078) reflexion, of sign *g*, is shown in various positions.

---

[*] From Friedel's law it follows that

$$S(\mathbf{h})S(\mathbf{h}') = S(\mathbf{h})S(\bar{\mathbf{h}}') \approx S(\mathbf{h}+\mathbf{h}') \quad \text{and} \\ \approx S(\mathbf{h}-\mathbf{h}') \,,$$

so that at least the two terms $S(\mathbf{h}+\mathbf{h}')$ and $S(\mathbf{h}-\mathbf{h}')$ coincide. When the Laue symmetry includes symmetry elements additional to the centre of symmetry, it is possible for more than these two terms to coincide (see § 3·1 and Fig. 1).

$$|U(\mathbf{h})| = \left[I(\mathbf{h})\frac{\Sigma f^2}{\langle I\rangle}\right]^{\frac{1}{2}}\frac{1}{\Sigma f},$$

where $I(\mathbf{h})$ is the relative intensity of the $0kl$ reflexion, $\langle I\rangle$ is the average intensity of the reflexions having nearly the same $\sin\theta$ value, and f is the atomic scattering factor.

2·2. Those terms for which $|U(\mathbf{h})| \gtrsim 0{\cdot}15$ are selected and arranged in a square lattice (Fig. 1), in which the letters denote $S(\mathbf{h})$, and the numbers indicate the relative magnitudes of $|U(\mathbf{h})|$. For reasons which will emerge in § 3, it is desirable to preserve a distinction between the four groups of term signs characterized by:

k even, l even;
k odd, l even;
k even, l odd;
k odd, l odd.

For convenience they will be referred to briefly as groups (e, e), (o, e), (e, o) and (o, o) respectively, and, to facilitate recognition, a different fount is used for the letters representing the signs of each group. The negative signs in Fig. 1 occur because in plane groups *pgg* and *p4g*.

$$\left.\begin{aligned} &S(0kl) = S(0\bar{k}l) = S(0k\bar{l}) = S(0\bar{k}\bar{l}) \\ &\qquad\qquad\qquad\qquad\text{for } (k+l) \text{ even} \\ &\text{and} \\ &S(0kl) = -S(0\bar{k}l) = -S(0\bar{k}l) = S(0\bar{k}\bar{l}) \\ &\qquad\qquad\qquad\qquad\text{for } (k+l) \text{ odd}\,. \end{aligned}\right\} \tag{6}$$

Similar relations hold for *pmg* when k is even or odd.

2·3. The products $S(\mathbf{h}')S(\mathbf{h}+\mathbf{h}')|U(\mathbf{h}')U(\mathbf{h}+\mathbf{h}')|$ corresponding to a given term $S(\mathbf{h})|U(\mathbf{h})|$ are quickly found in the following way. A sheet of perspex is placed over the lattice, the vector $(\mathbf{h})$ is marked on it, as for (078) in Fig. 1, and the sheet is then 'ranged' over the lattice. Every position of the marked vector in which its ends simultaneously touch two occupied lattice points is found. Each pair of occupied lattice points connected thus by the vector $(\mathbf{h})$ corresponds to a pair of values of $(\mathbf{h}')$ and $(\mathbf{h}+\mathbf{h}')$. The products are used in two ways. (i) They are listed in separate tables for each term. Table 1 (column (1)) gives the list of products for the (025) term. These tables are used later (§ 4·1). (ii) They are entered in the *sorting array*, part of which is shown in Fig. 2(*a*). An example of such an entry is given in § 2·4. It assists in the filling in and interpretation of the sorting array if the term signs are set out in the groups (e, e), (o, e), (e, o), (o, o) along the edges, as in Fig. 2(*b*). The squares contained within the large blocks of Fig. 2(*b*) will then contain terms of one group only, as shown in parentheses.

Table 1

In column (1) are listed in order of magnitude the products corresponding to the (025) term of hydroxydihydroeremophilone, the term sign of which is denoted by the letter ζ in Figs. 1 and 2; it is one corrected by the iterative process. Column (2) contains the signs of the products determined from the tentative list (§§ 3·2, 3·3). Columns (3)–(8) give the product signs at each stage of the iteration. A question mark indicates that the sign of one or both members of the product was uncertain at that stage. At the foot of each column is the sign deduced for ζ from consideration of all the product signs listed in that column; this sign was used in the following cycle.

(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)
42	$\overline{\varrho\mathscr{H}}$	+	−	+	?	−	−	−
32	$\overline{Kg}$	−	−	−	−	−	−	−
30	$\overline{\lambda\mathscr{P}}$	−	−	−	−	−	−	−
28	$\theta\mathscr{F}$	+	+	+	+	+	+	+
20	$\overline{Gm}$	+	+	?	+	+	+	+
20	$\mu\mathscr{G}$	?	?	+	?	?	−	−
18	$\overline{Nu}$	−	−	−	?	−	−	−
18	$\overline{\varrho\mathscr{R}}$	+	+	−	+	−	−	−
18	$\overline{Nl}$	−	?	+	?	+	−	−
16	$\overline{\psi\mathscr{I}}$	+	+	?	+	+	+	+
16	Lf	−	−	−	−	−	−	−
16	$\overline{\beta\mathscr{F}}$	+	−	+	−	?	−	−
15	$\overline{Mt}$	+	?	−	+	+	?	−
12	Dl	−	?	+	?	−	−	−
12	Ke	+	+	+	+	+	+	+
12	$\overline{Ks}$	?	+	−	−	?	−	−
12	$\overline{Pv}$	?	?	+	?	+	+	−
12	Em	−	+	+	+	+	−	−
12	$\varphi\mathscr{T}$	+	−	−	−	−	−	−
9	$\overline{Ha}$	−	−	?	−	−	−	−
9	Hn	−	+	−	−	−	−	−
9	$\eta\mathscr{L}$	?	−	+	−	−	−	−
		+	−	+	−	−	−	−

The sign of the (025) term in the last Fourier, after refinement, was −.

2·4.* The product $(\overline{4f})(8\mathscr{C})$, one of those corresponding to g (078) (Fig. 1), is considered as an example. This is (i) entered in the table of products corresponding to g as $32\overline{f\mathscr{C}}$, and (ii) entered as $\bar{g}$ in square $(4f, 8\mathscr{C})$ of the sorting array; this is one of two entries ringed in Fig. 2(*a*). The process of § 2·3 is repeated for all vectors $(\mathbf{h})$ appearing in the lattice of Fig. 1.

After the sorting array has been filled in, and before the coincidences are used (§ 3·1), it is most important to check as follows. All entries occur in threes, e.g. the sign relation

$$\bar{a}H\zeta \approx +1$$

appears three times:

$$\text{(i) as } \bar{a} \text{ in } (3H, 4\zeta), \quad \text{(ii) as } \bar{H} \text{ in } (4\zeta, 3a),$$
$$\text{(iii) as } \bar{\zeta} \text{ in } (3H, 3a)\,.$$

It is necessary to check that for every sign relation

---

* The italic, script and Greek characters in the text are equivalent to the corresponding characters in Figs. 1–3 although somewhat different in type face.

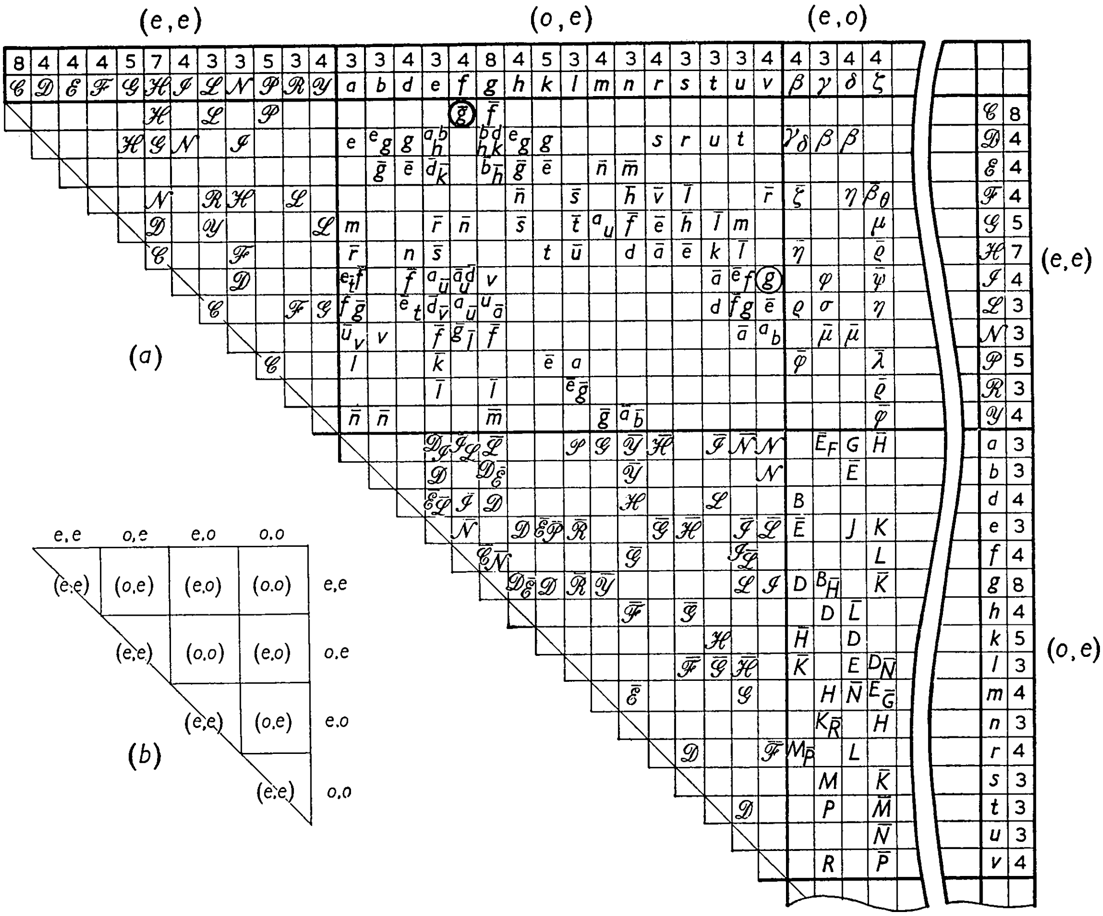

Fig. 2. (*a*) A portion of the sorting array: it is the top lefthand part of (*b*). The term sign symbols are set out along the edges together with $20|U|$ rounded off to the nearest integer. Of the products corresponding to g, (§ 2·4), which are indicated in Fig. 1, two are ringed here. They are:

(i) $(\overline{4f}).(8\mathscr{C})$, which is entered as $\bar{g}$ in square $(4f, 8\mathscr{C})$, and
(ii) $(4v).(4\mathscr{I})$, which is entered as g in square $(4v, 4\mathscr{I})$.

(*b*) A schematic outline of the main features of the sorting array. This lay-out facilitates both the filing and checking of the entries.

all three are present and correctly filed. This helps to eliminate errors and to ensure that no coincidences are missed.

3. The recognition and correlation of coincidences

3·1. The sorting array of products (Fig. 2(*a*)) is first searched for two or more entries occurring in the same square, i.e. for 'coincidences' between term signs. For example in square $(4f, 4\mathscr{I})$ there are $\bar{a}$, $\bar{d}$ and u; hence it is inferred that a and d are probably of the same sign, but are probably of opposite sign to u. This combination of signs arises, of course, as a result of (6), and for *pgg*, *pmg* and *p4g* a maximum of four can occur in one square of the sorting array.

Such inferences will fall naturally into the four distinct groups (e, e), (o, e), (e, o) and (o, o) (see § 2·3 and Fig. 2(*b*)). It is convenient to tabulate the inferences in small *coincidence arrays*, one for each group; two are shown in Fig. 3. For example, from square $(3a, 3\mathscr{L})$ of Fig. 2(*a*), $f \approx \bar{g}$. The appropriate weight for

$$a\mathscr{L}\bar{g} \approx +1 \text{ is } 3\times3\times8 = 72\,,$$

and that for

$$a\mathscr{L}f \approx +1 \text{ is } 3\times3\times4 = 36\,.$$

The *conjoint weight* for the coincidence is $72\times36 = 2592$; the corresponding entry in squares $(4f, 8g)$ and $(8g, 4f)$ of the (o, e) coincidence array (Fig. 3) is $\bar{9}$, which should be interpreted as meaning that $f \approx \bar{g}$ from coincident relations of conjoint weight $8\times4\times(9)^2 = 2592$. The second entry of $\bar{9}$ in these squares $(4f, 8g)$ and $(8g, 4f)$ of the (o, e) coincidence array is due to the coincidence recorded in square $(3u, 3\mathscr{L})$ of the sorting array (Fig. 2(*a*)).

There is no restriction on the number of entries which can occur in any one square of the coincidence

		$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{L}$	$\mathscr{N}$	$\mathscr{P}$	$\mathscr{R}$	$\mathscr{Y}$
		8	4	4	4	5	7	4	3	3	5	3	4
$\mathscr{C}$	8	+49 +25 +9								32			
$\mathscr{D}$	4			$\overline{32}$ $\overline{24}$ 20				9				21 35	
$\mathscr{E}$	4		$\overline{32}$ $\overline{24}$ 20						12		15		
$\mathscr{F}$	4											16	
$\mathscr{G}$	5										12		20
$\mathscr{H}$	7											24	
$\mathscr{I}$	4		9						9 $\overline{12}$ $\overline{12}$				
$\mathscr{L}$	3			12				9 $\overline{12}$ $\overline{12}$	+16				
$\mathscr{N}$	3	32											
$\mathscr{P}$	5			15		12							
$\mathscr{R}$	3		21 35		16		24						
$\mathscr{Y}$	4					20							+9
		+	−	+	−	+	−	−	+	+	+	−	+

		a	b	d	e	f	g	h	k	l	m	n	r	s	t	u	v
		3	3	4	3	4	8	4	5	3	4	3	4	3	3	3	4
a	3		12 12 12	16 $\overline{15}$				12				$\overline{12}$				$\overline{12}$ $\overline{12}$ 20 $\overline{16}$ $\overline{24}$	
b	3	12 12 12		32 $\overline{16}$				32 12 32	32	$\overline{12}$							
d	4	16 $\overline{15}$	32 $\overline{16}$					32	32 12							$\overline{16}$	9
e	3					$\overline{12}$ $\overline{12}$	16 9 $\overline{16}$ 12 12							$\overline{16}$	$\overline{12}$ $\overline{12}$		
f	4				$\overline{12}$ $\overline{12}$		$\overline{9}$ $\overline{9}$	9							12		
g	8				16 9 $\overline{16}$ 12 12	$\overline{9}$ $\overline{9}$					$\overline{9}$			16		24	
h	4	12	32 12 32	32		9			32 $\overline{12}$				$\overline{16}$		18		
k	5		32	32 12				32 $\overline{12}$									
l	3		$\overline{12}$													24	9
m	4						$\overline{9}$										
n	3	$\overline{12}$															$\overline{9}$
r	4							$\overline{16}$									
s	3				$\overline{16}$		16										
t	3				$\overline{12}$ $\overline{12}$	12		18									
u	3	$\overline{12}$ $\overline{12}$ 20 $\overline{16}$ $\overline{24}$		$\overline{16}$			24			24							12 $\overline{9}$
v	4			9						9		$\overline{9}$				12 $\overline{9}$	
		−	−	−	+	−	⊕	−	−	+	−	+	+	+	−	+	−

Fig. 3. The coincidence arrays for groups (e, e) and (o, e). An entry of, for example, $\overline{9}$ in square $(4f, 8g)$ indicates that $f \approx \bar{g}$ with conjoint weight of $4 \times 8 \times (9)^2$. A triple coincidence such as $e \approx \bar{f} \approx \bar{t}$ in $(3a, 4\mathscr{I})$ (Fig. 2(a)) gives three distinct entries in the (o, e) array: $\overline{12}$ in $(3e, 3t)$ and $(3t, 3e)$; $\overline{12}$ in $(3e, 4f)$ and $(4f, 3e)$; 12 in $(4f, 3t)$ and $(3t, 4f)$. The fourfold coincidence in $(8g, 4\mathscr{D})$ gives six distinct entries.

In these two arrays there were enough coincidences to permit all term signs to be tentatively determined, but this does not always happen. The *resultant* conjoint weights in the (e, e) array form a completely consistent set, whereas there are a few minor inconsistencies which cannot be eliminated from the (o, e) array, and this is more usual.

array, but sometimes, e.g. in square $(4\mathscr{D}, 4\mathscr{E})$ of the (e, e) coincidence array, they do not all agree. Hence, the *resultant* conjoint weight of the coincidence is $4 \times 4 \times \{20^2 - 32^2 - 24^2\} = -19{,}200$, i.e. $\mathscr{D} \approx \mathscr{E}$ with this resultant conjoint weight. (For a more exact discussion see the Appendix.)

3·2. The term signs in group (e, e) are independent of the origin chosen. Some may be found absolutely from two types of relation:

$$S(0,2k,2l) \approx S(0kl)S(0\bar{k}\bar{l}) = +1 \qquad (7a)$$

and

$$S(0,0,2l) \approx S(0kl)S(0k\bar{l}) = (-1)^{k+l}, \qquad (7b)$$

the entries for which appear on the diagonal of the sorting array (Fig. 2(a)). (Others such as

$$S(0,2k,2l)S(0kl) \approx S(0kl)\,,$$

which is a variant of (7a), give entries elsewhere in the array but they are not independent and provide no new information.) No coincidence entries can appear on the diagonals of coincidence arrays, and so the diagonal of the (e, e) coincidence array is a convenient place for entering indications of absolute sign given by either (7a) or (7b). There is none of the latter in Fig. 3; the former are ringed to avoid confusing them with the coincidences.

From these terms of known sign, the process of allotting tentative signs to the terms in the (e, e) coincidence array is started. Thus, in square $(8\mathscr{C}, 8\mathscr{C})$ on the diagonal of this array there are three indications that $\mathscr{C}$ is positive and a + sign is therefore entered at the foot of column $(8\mathscr{C})$. From the top row, square $(3\mathscr{N}, 8\mathscr{C})$, it appears that $\mathscr{N}$ probably has the same sign as $\mathscr{C}$. A + sign is therefore inserted at the foot of column $(3\mathscr{N})$ also. Again, in square $(3\mathscr{L}, 3\mathscr{L})$ there is an indication that $\mathscr{L}$ is positive; this is recorded by a + sign at the foot of column $(3\mathscr{L})$. From the row $3\mathscr{L}$ it appears that $\mathscr{E}$ has the same sign as $\mathscr{L}$, and that $\mathscr{I}$ is of opposite sign to $\mathscr{L}$. If this process is continued throughout the coincidence array, it will be found that tentative signs may be allotted to all the terms in the (e, e) group*. All the entries in a column are taken into account in fixing the sign, which is finally placed at the bottom of that column. (It may be necessary to go through the array more than once, and to adjust the signs at the foot of the table until the closest possible approach to a self-consistent set of signs is reached.)

3·3. In the plane group *pgg* there are four possible centrosymmetric origins. If two term signs be chosen arbitrarily, one from each of two of the remaining

---

* Woolfson (1954) has described a different procedure for getting absolute signs for terms in this group, but, as it too is based on the available triple sign products, it yields no independent information.

three groups (e, o), (o, e) and (o, o), the origin is thereby selected. The coincidence arrays for the three groups are examined in order to select the terms which are to be given arbitrary signs; two terms, which enter into a large number of coincidences, should be chosen so that as many as possible of the term signs in their respective groups may be deduced tentatively from the coincidence arrays. Thus the solution of the (o, e) coincidence array (Fig. 3) starts from g, which is arbitrarily taken as positive. The procedure for allotting tentative signs to the rest of the (o, e) terms is the same as that for the (e, e) array (§ 3·2), except that there is only one sign $(g = +)$ from which to start. One of the remaining two groups, say (e, o), may be treated similarly, leaving the last group (o, o) in which no term signs are known at this stage. If the lower limit of $|U|$ has been suitably chosen (see § 7), tentative signs can be allotted to almost all the terms in the three groups (e, e), (e, o) and (o, e). There is no need to do anything about the (o, o) group of terms at this point; the iterative process of § 4·1 will yield signs for the terms in this group.

4. The iterative process

A tentative set of signs is now available for a substantial proportion of the terms in the lattice, and this set forms the basis for the iterative procedure.

4·1. Attention is now directed to the lists of products, $S(\mathbf{h}')S(\mathbf{h}+\mathbf{h}')|U(\mathbf{h}')U(\mathbf{h}+\mathbf{h}')|$, referred to in § 2·3(i). Table 1 gives the list for the (025) term, sign ζ, of hydroxydihydroeremophilone. Column (2), which contains the signs of the products, is filled in wherever possible by using the tentative set of signs (§ 3·2 and § 3·3). The entries in this column are then used in conjunction with the weights in column (1) to deduce the probable sign ζ of this term (cf. equation (5)). The sign so found is entreed at the foot of column (2). All the terms are treated in this way in turn; those to which signs have already been tentatively allotted are reassessed. As a result, a new, and usually longer, list of term signs is available with which the whole process is repeated. The results of each subsequent stage are shown in columns (3), (4), (5), ... and at the foot of each column is shown the revised assessment of the term sign to which the products correspond. The iterative process is continued until a minimum number of term signs remains doubtful; some five or six cycles will usually suffice. The sign finally allotted to each term is that at the foot of the last column of the table which relates to that term.

It should be pointed out that there are variants which may speed up the procedure and which may improve the rate of convergence, but the one described has been chosen for convenience of presentation.

4·2. It is necessary to make a check on the conclusions so far reached. If a different pair of arbitrary terms (§ 3·3) is taken, the whole sequence can be repeated, and if the same, or an equivalent, set of signs is converged upon, it is reasonable to suppose that no false deductions have been made. Whether the set obtained is the correct one or not can be determined only by the subsequent interpretation and refinement of the electron-density map so obtained.

5. The extension of the method to terms of smaller $|U|$

The signs of some terms of smaller $|U|$ may be obtained by using, on the original lattice, the vector appropriate to each such small term. A large number of products must be taken into account before these term signs can be assessed with any confidence; even so, they will be less reliable than the signs of the terms of the original lattice (Fig. 1).

6. Applications of the method

6·1. Harmine, $C_{13}H_{12}N_2O$, orthorhombic, space group $P2_12_12_1$,

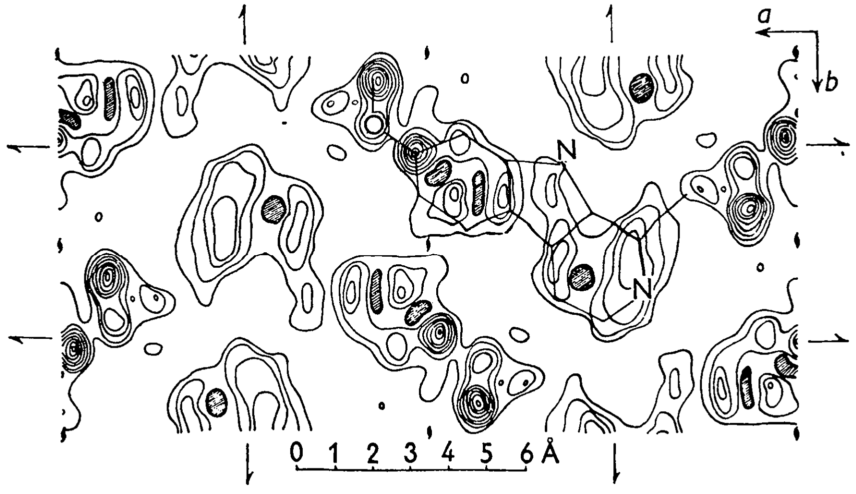

Fig. 4. An electron-density map of the (001) projection of harmine based on 41 terms with signs found as described in § 6·1. The skeleton indicates the interpretation in terms of the known chemistry.

$$a = 19{\cdot}3\ \text{Å},\quad b = 9{\cdot}8\ \text{Å},\quad c = 5{\cdot}8\ \text{Å}\,.$$

Of the 183 independent observed $F(hk0)$'s, 41 had $|U| \geq 0{\cdot}15$, the largest being $0{\cdot}38$; consequently there was no possibility of a solution by inequalities. The method was applied and gave 120 distinct triple sign relations and 54 coincidences. The whole set of 41 signs was obtained. An electron-density map, prepared from these 41 terms (Fig. 4), was interpreted in terms of the known chemistry of the molecule. Subsequent refinement has given Fig. 5, and has shown that 27 of the 41 terms were correct; many of the remaining 14 were high-order terms with relatively low values of $|F|$, so that their influence on the main features of the map of Fig. 4 is slight. A further 31 terms were investigated with $|U| < 0{\cdot}15$. Of these, 9 could not be allotted a sign, and, of the remaining 22 signs, 16 were subsequently found to be correct.

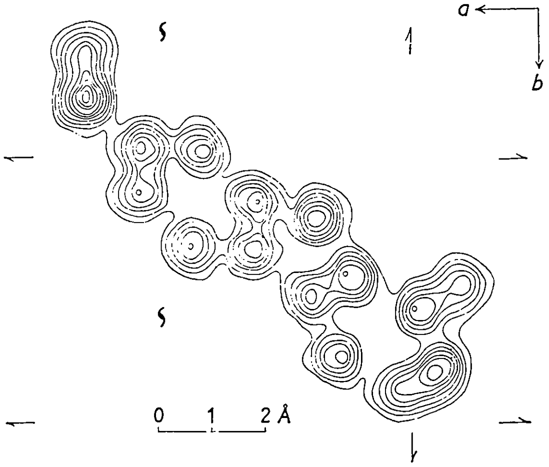

Fig. 5. The most recent electron-density map of the (001) projection of harmine. ($R = 0{\cdot}22$ taken over all observed reflexions.)

6·2. Hydroxydihydroeremophilone, $C_{15}H_{24}O_2$, orthorhombic, space group $P2_12_12_1$, $a = 7{\cdot}5$ Å, $b = 10{\cdot}0$ Å, $c = 19{\cdot}7$ Å.

Of the 148 observed $F(0kl)$'s, 55 had $|U| \geq 0{\cdot}15$, the largest again being $0{\cdot}38$ (only 7 exceeded $0{\cdot}3$) so inequalities were, therefore, unusable. The method was applied and gave 230 distinct triple sign relations leading to 105 distinct coincidences. Consideration of the coincidences made it possible to give signs tentatively to 36 out of the 38 terms in the (e, e), (o, e) and (e, o) groups; after six cycles of the iterative process a set of 54 signs (including the two arbitrarily fixed signs) was obtained. The method was extended to a further 38 terms for which $0{\cdot}1 < |U| < 0{\cdot}15$, and it was possible to give signs to 24 of them. Two electron-density maps were calculated, one containing the original 54 terms and the other all the 78 terms, and both were interpreted in the same way (see Figs. 2(*a*) and 2(*b*), Grant, 1957). Subsequent refinement has shown that 51 of the original set and 22 of the second set had correct signs.

The method was also applied to the (010) projection of the same crystal. Of the 90 independent observed $F(h0l)$'s only 28 had $|U| \geq 0{\cdot}15$. The method when applied gave only 42 sign relations. Signs were found for all 28 terms, but the electron-density map so obtained (Fig. 2(*c*), Grant, 1957) could not be interpreted and the projection was eventually solved in another way. Subsequent refinement has shown that 9 of the 28 terms had incorrect signs; the correct set is another self-consistent solution of the products used.

7. Discussion

The solution under favourable conditions of two crystal structures of moderate size (17 atoms per molecule) by the method described would appear to have extended the range of Sayre's or Cochran's methods beyond the 10–11 light atoms considered by Cochran to be the useful limit. Admittedly, in the case of hydroxydihydroeremophilone, the projection solved contained much overlap and the effective number of scattering points was less than the 17 atoms. In harmine, however, there is no overlap in the projection solved. It is emphasized that in both the zones successfully studied no previous knowledge of the signs of any terms was available or was required: the solutions were obtained *ab initio*.

The set of unitary structure factors for harmine was sent to Dr W. Cochran, who kindly tried the digital-computer method on the data. The machine (EDSAC) could not accept all the data we used, so 24 terms of largest $|U|$ (2 of which were given arbitrary signs) were selected and 36 sign relations between these were considered. The relations were arranged in order of decreasing probability, and the machine was asked to give the sets of signs for which no more than 3 of the first 22 relations, and not more than 6 out of the total of 36, would be incorrect. 158 such sets of signs were produced by the machine. One of them, that with the maximum value of χ (see § 1), agreed with that obtained by Howells, but is now known to have 5 incorrect signs. There were also 5 other sets, each with 5 incorrect signs, and a further 4 sets, each with 4 incorrect signs (all with smaller χ), but none with less than 4 incorrect signs. There was, however, no simple way of selecting which of the 158 sets was closest to the correct one.

The present method has the following points in its favour:

(*a*) The principle of coincidences has proved useful in getting a start on a problem. The number of coincidences increases rapidly as the cut-off limit of the $|U|$'s is lowered, and their conjoint probabilities can be remarkably large (see Appendix).

(*b*) The iterative process compels the solution to conform to the very large number of sign relations; this number increases roughly as the square of the number of terms used.

(*c*) This wealth of sign relations and coincidences causes the problem to be much more elaborately overdetermined than has been possible in other methods, and this is believed to be the principal reason for reaching single solutions in two of the applications recorded here. It appears from experience gained from harmine, hydroxydihydroeremophilone and other problems that, for success, the ratio of the number of sign relations to terms should be at least 2 or 3 to 1, and the lower limit of $|U|$ should be taken accordingly. Admittedly there is a rapid increase in the volume of work as the number of terms increases, but it leads usually to greater certainty in the sign discrimination and also often improves the rate of convergence of the iterative process. A problem such as the main projection of hydroxydihydroeremophilone can, after a little practice, be carried through all the stages from § 2·2 onward in three or four days.

It is possible that the machine results for harmine may have been vitiated by random and systematic errors in the $|U|$'s, which are usually more serious for the high-order terms. The top 24 chosen for EDSAC might well have been different had accurate values of $|U|$ been known, and it is not known if they would have shortened the list of possible solutions or given one nearer the truth. But it would appear that the manual method copes more successfully with such errors because it uses a larger amount of information.

A practical problem remains in the interpretation of the electron-density maps, which are inevitably rather 'garbled' versions, and, as in most structure determinations, this may well be the least easy step in the entire procedure. Intermolecular overlap, which complicates the interpretation rather more than intramolecular overlap, may indeed make recognition impossible.

The method may fail in a particular problem for a variety of reasons, but, until it has been used more extensively, there will not be sufficient evidence to formulate the conditions under which one might expect an application of the method to be successful.

It should be emphasized that *any* centrosymmetric plane group will yield (i) absolute signs for terms in group (e, e) (see § 3·2) and (ii) coincidence arrays. It will, however, be impossible to get any negative signs or negative coincidences except in plane groups *pgg*, *pmg* and *p4g*. In the absence of negative coincidences, the inevitable solution would be the improbable and unacceptable result that all terms have the same sign.

APPENDIX

Consider the two coincident triple sign relations:

$$S(\mathbf{A})S(\mathbf{B}) \approx S(\mathbf{C}_1)\,,$$
$$S(\mathbf{A})S(\mathbf{B}) \approx S(\mathbf{C}_2)\,,$$

for which the probabilities (3) are respectively

$$\alpha = \tfrac{1}{2}[1+\tanh\theta_1]$$

and

$$\beta = \tfrac{1}{2}[1+\tanh\theta_2]\,,$$

where

$$\theta_1 = |NU(\mathbf{A})U(\mathbf{B})U(\mathbf{C}_1)|$$

and

$$\theta_2 = |NU(\mathbf{A})U(\mathbf{B})U(\mathbf{C}_2)|\,,$$

where N is the number of atoms in the unit cell. The probability that $S(\mathbf{C}_1) \approx S(\mathbf{C}_2)$ is

$$p(=) = \alpha\beta+(1-\alpha)(1-\beta) = 1-\alpha-\beta+2\alpha\beta\,,$$
$$= \tfrac{1}{2}[1+\tanh\theta_1 . \tanh\theta_2]\,, \quad (8)$$

i.e. $\tfrac{1}{2} \leq p\,(=) \leq 1$.

For example, from square $(8g, 4\mathscr{D})$ in Fig. 2(*a*), $b \approx h$.

For

$$b \approx g\mathscr{D},\quad \theta_1 = 50\times3\times8\times4/(20)^3 = 0{\cdot}6,\quad \alpha = 0{\cdot}77\,,$$

and for

$$h \approx g\mathscr{D},\quad \theta_2 = 50\times4\times8\times4/(20)^3 = 0{\cdot}8,\quad \beta = 0{\cdot}88\,,$$

assuming an effective value for N of 50 atoms.

The probability that b and h have the same sign is, therefore,

$$p(=) = 0{\cdot}70\,.$$

Similarly, from square $(8g, 3\mathscr{L})$, we have $a \approx -u$, with $p(+) = 0{\cdot}59$.

If there are two coincidences between the same pair of terms (as, for example, in square $(3\mathscr{R}, 4\mathscr{D})$ in Fig. 3) and the probabilities of the signs agreeing in each case are $p_1(=)$ and $p_2(=)$ respectively, then the resultant probability of the signs agreeing for the double coincidence is $P(=)$, where

$$P(=) = \frac{p_1(=)p_2(=)}{p_1(=)p_2(=) + p_1(+)p_2(+)}\,, \quad (9)$$

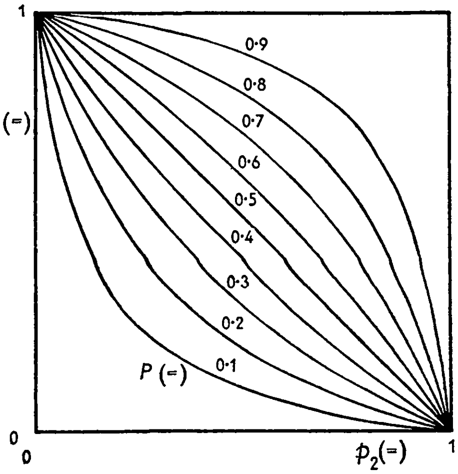

Fig. 6. A graphical representation of (9), giving the resultant conjoint probability, $P(=)$, of a duplicated coincidence, the individual probabilities for each coincidence being $p_1(=)$ and $p_2(=)$.

which is greater than $p_1(=)$ if $p_1(=) \geq p_2(=) \geq \frac{1}{2}$, and is illustrated graphically in Fig. 6.

More generally, for n coincidences the resultant probability of the positive sign is deducible from the equation

$$\frac{P(=)}{P(\mp)} = \prod_{j=1}^{n} \left(\frac{p(=)}{p(\mp)} \right)_j . \tag{10}$$

$P(=)$ is very conveniently evaluated stepwise by repeated use of Fig. 6.

As an example, consider the coincidences in square $(4\mathscr{D}, 4\mathscr{E})$ of the (e, e) array (Fig. 3 and § 3·1). Two different levels of probability for each coincidence will be considered, corresponding to 16 and 10 atoms per molecule—to allow for the overlapping in projection:

	16 atoms	10 atoms
$p_1(=)$	0·66	0·57
$p_2(=)$	0·20	0·34
$p_3(=)$	0·29	0·40

From Fig. 6,

$P(=)$	0·17	0·32
$P(\mp)$	0·83	0·68

Even at the lower level these are useful.

Had all three been of the same sign, as often happens, the resultant probabilities for $P(\mp)$ would have been 0·95 and 0·75 respectively.

The most appropriate quantity to use for $W(\mathrm{C}_1, \mathrm{C}_2)$, the resultant conjoint weight, when maximizing (12) below, is obviously $(2P(=)-1)$, but an exact evaluation of this is rather tedious and often unnecessary, for the form mentioned in § 3·1 is a reasonable approximation for all but the larger coincidences. Thus if $\tanh \varphi_1$ is substituted for $(\tanh \theta_1 \tanh \theta_2)$ in equation (8), where θ_1 and θ_2 have the values given above, and if similar quantities are defined for each of the coincidences between these two terms, it is easy to show that

$$W = \tanh \left[\sum_{j=1}^{n} \varphi_j \right] , \tag{11}$$

whence, if $\Sigma\varphi_j$ and all its contributing terms are not large, the approximation of § 3·1 is obtained, apart from some constant factors.

The solution of the four coincidence arrays could easily be mechanized, but so far nothing sufficiently difficult to warrant this has arisen. It would involve the maximization of

$$\Gamma = \sum_{\mathrm{C}_1} \sum_{\mathrm{C}_2} S(\mathrm{C}_1) S(\mathrm{C}_2) W(\mathrm{C}_1, \mathrm{C}_2) , \tag{12}$$

where $W(\mathrm{C}_1, \mathrm{C}_2)$ is the resultant conjoint weight (including sign) for the coincidence between $S(\mathrm{C}_1)$ and $S(\mathrm{C}_2)$.

Note added in proof, 19 *May* 1957.—No reference has been made here to the paper recently published by Woolfson (1957), as the authors were unaware of its existence until it appeared in print after this paper went to press.

We wish to record our appreciation of the cooperation and helpful criticisms of Dr W. Cochran.

References

Cochran, W. (1952). *Acta Cryst.* **5**, 62.

Cochran, W. & Douglas, A. S. (1955). *Proc. Roy. Soc.* A, **227**, 486.

Cochran, W. & Woolfson, M. M. (1955). *Acta Cryst.* **8**, 1.

Grant, D. F. (1957). *Acta Cryst.* **10**, 498.

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

Howells, R. G. (1954). *Acta Cryst.* **7**, 627.

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

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

Vand, V. & Pepinsky, R. (1956). *Z. Kristallogr.* **107**, 202.

Woolfson, M. M. (1954). *Acta Cryst.* **7**, 721.

Woolfson, M. M. (1957). *Acta Cryst.* **10**, 116.

Zachariasen, W. H. (1952). *Acta Cryst.* **5**, 65.