Acta Cryst. (1954). 7, 721

Sign Determination for *pgg*, *p4g* and certain other Space Groups

By M. M. Woolfson

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

(*Received* 11 *March* 1954 *and in revised form* 19 *June* 1954)

An analytical method is described for finding the probable signs of structure factors of even indices for the (two-dimensional) space groups *pgg* and *p4g* from a knowledge of the intensities alone.

It is shown that the sign of $U_{2h,2k}$ is probably the sign of $\sum_{h'} \sum_{k'} W_{h'k'}(-1)^{h'+k'}$, where

$$\begin{aligned} W_{h'k'} = & \tanh(N|U_{2h,2k}U_{h+h',k}U_{h-h',k}|) \\ & \times \tanh(N|U_{h+h',k}U_{h,k+k'}U_{h'k'}|) \\ & \times \tanh(N|U_{h-h',k}U_{h,k+k'}U_{h'k'}|) \\ & + \tanh(N|U_{2h,2k}U_{h,k+k'}U_{h,k-k'}|) \\ & \times \tanh(N|U_{h,k+k'}U_{h+h',k}U_{h'k'}|) \\ & \times \tanh(N|U_{h,k-k'}U_{h+h',k}U_{h'k'}|) \end{aligned}$$

and the summation is carried out over all h' and k' contained in two adjacent quadrants.

The principles are extended to other space groups and it is shown that the restriction of even indices may be overcome.

An example of the application of the method is given, the reliability of the results is discussed and some possible extensions of the method are suggested.

1. Introduction

Definite or probable signs for the structure factors of a centrosymmetrical crystal may sometimes be obtained from a knowledge of the reflexion intensities alone. The Harker–Kasper inequalities give signs unequivocally when large enough unitary structure factors are available. When the unitary structure factors are smaller, probable signs may sometimes be found by applying the sign relationship $s(h) \approx s(h')s(h+h')$, (Cochran, 1952; Zachariasen, 1952), where $\approx$ is taken to mean 'probably equals'.

A special case of the sign relationship in two dimensions is

$$s(2h,0) \approx s(hk)s(h\bar{k}) . \qquad (1)$$

For the two-dimensional space groups *pgg* and *p4g*,

$$s(hk) = (-1)^{h+k}s(h\bar{k}) , \qquad (2)$$

whence $s(2h,0) \approx (-1)^{h+k}$ if the involved structure factors are all large. For a constant h, contradictory indications of sign for $F_{2h,0}$ may be found, depending on whether a chosen value of k makes $h+k$ even or odd. If one set of unitary structure factors for which h is constant and for which $h+k$ is even, say, is found to be stronger than the other set then a definite sign indication is given for $F_{2h,0}$.

The method to be described, by which signs may be determined for general terms of the type $F_{2h,2k}$ is an extension of the above idea.

2. Sign determination for reflexions of even indices

For a structure of the two-dimensional space group *pgg* (or *p4g*), let us consider two reflexions of indices $(h,k-k')$ and $(h'k')$ whose signs are given by

$$s(h,k-k') = a, \quad s(h'k') = b .$$

Then

$$s(h+h',k) \approx s(h,k-k')s(h'k') = ab \qquad (3)$$

and

$$s(h-h',k) \approx s(h,k-k')s(h'\bar{k}') .$$

From (2),

$$s(h'\bar{k}') = (-1)^{h'+k'}s(h'k') ,$$

which gives

$$s(h-h',k) \approx (-1)^{h'+k'}ab . \qquad (4)$$

Equations (3) and (4) combine to give

$$\begin{aligned} s(2h,2k) & \approx s(h+h',k)s(h-h',k) \\ & \approx (-1)^{h'+k'} \end{aligned} \qquad (5)$$

if the structure factors involved are all large.

It will be noticed that the derivation of (5) is not symmetrical with respect to h and k. By interchanging the rôles of h and k, (5) is derived through different intermediate indices although the same $(h'k')$ has been used.

The probability relation (5) resembles (2) in that changing the variables h' and k' leads to contradictory indications for $s(2h,2k)$. The weighting factor which should be given to each individual determination is

not immediately obvious and the more probable sign cannot easily be found. The probability that (5) gives a correct sign depends on the validity of three separate sign relationships and in general this probability will be small. However, since the variables h' and k' may be changed independently, and such a large number of applications of (5) may be made for any one reflexion, we may consider the aggregate of information to be statistically significant.

It is possible to calculate the probability of a given sign relationship being true (Woolfson, 1954*a*) and a weight may be given to each result of the application of (5).

If the weight for the pair of variables h' and k' is $W_{h'k'}$ then we may write

$$s(2h,2k) \approx s\{\sum_{h'} \sum_{k'} W_{h'k'}(-1)^{h'+k'}\} . \qquad (6)$$

It remains to determine the form of $W_{h'k'}$.

3. The calculation of the weighting function

The determination of a sign by the method outlined in § 2 depends on the validity of the three separate sign relationships occurring in (3), (4) and (5). A correct sign will be found not only if all three sign relationships hold but also *if any two of them fail.*

Let the probability that (3), (4) and (5) hold be P_1, P_2 and P_3 respectively. Then the probability that they all hold is $P_1P_2P_3$ while the probability that one holds and two fail is

$$P_1(1-P_2)(1-P_3)+P_2(1-P_3)(1-P_1) +P_3(1-P_1)(1-P_2) .$$

Hence the total probability of obtaining a correct sign is

$$P_1P_2P_3+P_1(1-P_2)(1-P_3) +P_2(1-P_3)(1-P_1)+P_3(1-P_1)(1-P_2)$$

or

$$(P_1+P_2+P_3)-2(P_1P_2+P_2P_3+P_3P_1)+4P_1P_2P_3 .$$

It is convenient to consider the excess of the probability of the sign relationship holding over that of it failing. For a probability P this is $P-(1-P)=2P-1$, and this will be called the 'excess probability'. For each of the sign determinations the excess probability is thus

$$2(P_1+P_2+P_3)-4(P_1P_2+P_2P_3+P_3P_1)+8P_1P_2P_3-1 = (2P_1-1)(2P_2-1)(2P_3-1) .$$

This is a useful and interesting result and shows that the overall excess probability for a sign determination is the product of the excess probabilities for each of the three stages.

If the probabilities of the three stages of the alternative derivation of (6) (where the rôles of h and k are reversed) are P_1', P_2' and P_3' respectively, we may use for the weighting function

$$W_{h'k'} = (2P_1-1)(2P_2-1)(2P_3-1) +(2P_1'-1)(2P_2'-1)(2P_3'-1) . \qquad (7)$$

It has been shown that the probability of the sign relationship $s(h) = s(h')s(h+h')$ being true is

$$P = \frac{\exp(2N|U_hU_{h'}U_{h+h'}|)}{1+\exp(2N|U_hU_{h'}U_{h+h'}|)} , \qquad (8)$$

where there are N resolved equal atoms in the unit cell (Woolfson, 1954*a*). The excess probability is then

$$2P-1 = \frac{\exp(2N|U_hU_{h'}U_{h+h'}|)-1}{\exp(2N|U_hU_{h'}U_{h+h'}|)+1} = \tanh(N|U_hU_{h'}U_{h+h'}|) .$$

Inserting the appropriate values for the unitary structure factors, we find

$$\begin{aligned} W_{h'k'} = &\tanh(N|U_{2h,2k}U_{h+h',k}U_{h-h',k}|) \\ &\times\tanh(N|U_{h+h',k}U_{h,k+k'}U_{h'k'}|) \\ &\times\tanh(N|U_{h-h',k}U_{h,k+k'}U_{h'k'}|) \\ &+\tanh(N|U_{2h,2k}U_{h,k+k'}U_{h,k-k'}|) \\ &\times\tanh(N|U_{h,k+k'}U_{h+h',k}U_{h'k'}|) \\ &\times\tanh(N|U_{h,k-k'}U_{h+h',k}U_{h'k'}|) . \end{aligned}$$

4. The application of the method

The summation (6) may be calculated in a systematic way and if some preliminary work is done the calculation is greatly simplified. It is repeatedly necessary to find the excess probability, $2P-1$, for a given value of the triple product of the unitary structure factors. It is convenient to multiply the U's by the factor 100/3 and to express them to the nearest whole number; the resulting small integers are more easily multiplied together. A table or graph of $2P-1$ may then be prepared for varying values of the triple product of the modified unitary structure factors.

The modified U's are plotted on two reciprocal-lattice charts, one on paper and the other on tracing paper or other transparent material. The transparent chart is placed with its origin on the point $(2h,2k)$ of the opaque chart; one line of the superimposed parallel charts shows all possible pairs of indices $(h+h',k)$ and $(h-h',k)$ corresponding to those on the right-hand side of (5). From the previously prepared table the value of $2P_3-1$ is found and noted for each pair. The origin of the upper chart is now moved to a point $(h-h',k)$ of the lower one; all the pairs of indices $(h,k+k')$ and $(h'k')$ which occur in equations (3) and (4) may then be read off along one line of the superimposed charts. The values of $(2P_1-1)(2P_2-1)$ are then tabulated in two groups, corresponding to $(-1)^{h'+k'}$ being positive or negative. This is done for all the points $(h-h',k)$.

The same process as that described above may be repeated with the initial step of finding pairs of indices such as $(h,k+k')$ and $(h,k-k')$ corresponding to the right-hand side of (5). In this way the summation (6) may be carried out systematically for all h' and k'.

In the following example there were about 300 values of $W_{h'k'}$ for each $(2h,2k)$ and each sign determination took about an hour. Where the method gave a strong sign indication it often became evident long before the summation was complete, and some time could be saved by not including all the terms in (6).

The method was used to determine the signs of some structure factors of α-glucose (McDonald & Beevers, 1952). Nine reflexions were chosen with $\sin\theta < 0{\cdot}8$ and $|U_{2h,2k}| > 0{\cdot}20$. Table 1 shows the results obtained.

Table 1

$2h$	$2k$	$\lvert U_{2h,2k}\rvert$	$\sum_{h'}\sum_{k'} W_{h'k'}(-1)^{h'+k'}$	True sign
2	12	0·24	−0·05	−
4	2	0·21	−0·46	+
6	8	0·24	−0·82	−
6	10	0·48	−0·40	−
8	4	0·36	+1·20	+
8	8	0·37	+0·41	+
8	10	0·29	+0·33	+
10	4	0·37	−1·59	−
12	4	0·40	−0·69	−

Eight of the nine signs were correctly given although one of these was only feebly indicated.

5. The reliability and usefulness of the method

It would be both interesting and useful to be able to calculate the probability that the sign found for the summation (6) is the correct sign. Attempts to calculate this probability have been unsuccessful and, even if it could be done, the answer would depend on the expression (7) for $W_{h'k'}$. This cannot be rigorously justified since it is likely that the probabilities P_1, P_2 and P_3 are interdependent and their independence has been tacitly assumed in deriving (7). Equation (7) will certainly give the excess probability fairly accurately and in practice this is all that is required. Signs were derived for eight reflexions of salicylic acid, using as the expression for the probability

$$P = K|U_h U_{h'} U_{h+h'}| \,.$$

Six of the eight signs were determined correctly, but when probabilities given by (8) were used one small indication of an incorrect sign changed to a small correct indication. It seems that any sensible weighting system will show those signs determined strongly by the method and, in particular, the weighting factor suggested in (7) is convenient to use and has some theoretical justification.

Incorrect signs will sometimes be found for a reason advanced by Cochran (1953). If the atoms of a structure are unequal (or there is overlap in a projection) it is possible that the mean value of $s(h')s(h+h')$ has the opposite sign to $s(h)$. This happens when a heavy atom (or overlapped atoms) gives a contribution to an F_h opposite to it in sign. If this condition occurs for a reflexion $(2h,2k)$ then (5) fails more often than not and the signs found from the summation (6) will probably be incorrect.

For the projection of α-glucose on (001) two atoms completely overlap. Table 2 gives the contribution of these atoms to the unitary structure factors of the reflexions whose signs have been determined.

Table 2

$2h$	$2k$	$U_{2h,2k}$	Contribution of overlapped atoms
2	12	−0·24	+0·032
4	2	+0·21	−0·044
6	8	−0·24	−0·060
6	10	−0·48	−0·190
8	4	+0·36	+0·016
8	8	+0·37	−0·006
8	10	+0·29	−0·020
10	4	−0·37	−0·140
12	4	−0·40	−0·162

The incorrect sign for (4,2) and the feeble sign indication found for (2,12) are explained by Table 2, while the adverse contributions of the overlapped pair to (8,8) and (8,10) appear to be too small to give an incorrect result.

It would seem from the above considerations that the method would be most successful for a structure with equal resolved atoms, although there is then the compensating disadvantage that the normal sign relationship $s(h) \approx s(h')s(h+h')$ between large unitary structure factors holds less often in this case (Cochran, 1952; Woolfson, 1954*a*).

Signs found by the technique described above must be used with some caution, especially if they are used as known signs in a chain process involving the normal sign relationship. It is common experience that the introduction of incorrect signs at an early stage plays havoc with such methods. However, if the signs are used as constant known terms with the method of permutation syntheses (Woolfson, 1954*b*) they can be most valuable. The fraction of correct signs for the calculated terms will probably equal or exceed the fraction correct in the best combination of the permuted terms and the quality of the syntheses examined will be improved.

6. A critical survey of possible extensions of the method

(a) The determination of $s(2h,2k)$ for other two-dimensional space groups

Sign relationships for special sets of reflexions may be combined in such a way that other assessments, not previously considered, may be made of $s(2h,2k)$.

Let us consider the following three statistical equations:

$$s(h-h',k-k') \approx s(hk)s(h'k') \,,$$
$$s(h+h',k+k') \approx s(hk)s(h'k') \,,$$
$$s(2h,2k) \approx s(h+h',k+k')s(h-h',k-k') \,,$$

from which we find

$$s(2h,2k) \approx + \qquad (9)$$

if the involved structure factors are all large. The sign indicated does not depend on the choice of $(h'k')$ although, if U_{hk} is small, the probability of (9) being valid will always be very little larger than $\frac{1}{2}$. Equation (9) is obviously not generally true for all the reflexions of a certain class and there is what may be called a 'positive sign paradox'. This paradox is a recurring feature of many sign-determining methods. The Harker–Kasper inequality

$$U_h^2 \leqslant \tfrac{1}{2}(1+U_{2h})$$

can only prove that U_{2h} is positive. The special case of the simple sign relationship

$$s(2h) \approx s(h)s(h)$$

also has this restriction and, however small the value of U_h, a greater probability of a positive sign is found for U_{2h} when the values are inserted in equation (8). It should be mentioned, however, that the derivation of (8) excludes the case where two reflexions are identical.

A more interesting statistical equation has been given by Hauptman & Karle (1953) and may be written in the form

$$s(2h) \approx s(U_h^2-\overline{U^2}) . \qquad (10)$$

This does allow a negative sign indication for U_{2h} but the positive sign paradox is not completely eliminated. The value of $U_h^2-\overline{U^2}$ is capable of being much larger as a positive quantity than as a negative one, and the Hauptman & Karle probability result for (10) shows that, for moderately complex structures, negative signs may be indicated only very weakly.

It seems from the above considerations that the determination of both positive and negative signs from intensities for the general case $P\bar{1}$ is not practically feasible at the present time.

For all rectangular and square space groups the procedure outlined in § 2 may be used, although, for *pmm*, *cmm* and *p4m*, $s(h'k') = s(h'\bar{k}')$ and $s(2h,2k)$ will always be found as positive—another example of the positive sign paradox.

For *pmg*, however, we have

$$s(h'\bar{k}') = (-1)^{h'}s(h'k') ,$$

and both positive and negative signs may be found. The equation corresponding to (6) in this case is

$$s(2h,2k) \approx s\{\sum_{h'}\sum_{k'} W_{h'k'}(-1)^{h'}\} ,$$

where $W_{h'k'}$ is given by equation (8).

(*b*) *Application of the method in three dimensions*

The process described in § 2 may readily be applied to certain centrosymmetrical orthorhombic space groups to determine the probable signs of reflexions with even indices.

The corresponding set of probability relations is

$$\begin{aligned} s(h+h',k,l+l') &\approx s(h,k-k',l)s(h'k'l') , \\ s(h-h',k,l-l') &\approx s(h,k-k',l)s(h'\bar{k}'l') , \\ s(2h,2k,2l) &\approx s(h+h',k,l+l')s(h-h',k,l-l') . \end{aligned}$$

From these we find

$$s(2h,2k,2l) \approx s(h'k'l')s(h'\bar{k}'l') \qquad (11)$$

with a probability which may be readily found. Interchanging the rôles of h', k' and l' in the above equations produces two similar probability relationships which give the complete set for a chosen $(h'k'l')$.

If the space group is such that

$$F_{hkl} = F_{\bar{h}kl} = F_{h\bar{k}l} = F_{hk\bar{l}} \qquad (12)$$

then all the sign indications are positive and the positive sign paradox is encountered once more. On the other hand there are many orthorhombic and tetragonal space groups for which the reflexions listed in (12) are not all related by a positive sign.

As an example we may consider the space group *Pban* for which the following relationships are valid:

$$\begin{aligned} &h \text{ even, } k \text{ even:} && F_{hkl} = F_{\bar{h}kl} = F_{h\bar{k}l} = F_{hk\bar{l}}; \\ &h \text{ even, } k \text{ odd:} && F_{hkl} = -F_{\bar{h}kl} = F_{h\bar{k}l} = -F_{hk\bar{l}}; \\ &h \text{ odd, } k \text{ even:} && F_{hkl} = F_{\bar{h}kl} = -F_{h\bar{k}l} = -F_{hk\bar{l}}; \\ &h \text{ odd, } k \text{ odd:} && F_{hkl} = -F_{\bar{h}kl} = -F_{h\bar{k}l} = F_{hk\bar{l}}. \end{aligned}$$

Equation (11) will give both positive and negative sign indications and this happens for many other orthorhombic and tetragonal space groups.

Corresponding to equation (6) there will be a triple summation with h', k' and l' varying independently. There are a very large number of terms in the summation and the aggregate of information should be statistically very significant although, for the reasons advanced in § 5, it has not been found possible to estimate the true value of the results.

The greatest drawback to working in three dimensions is the amount of computation which is necessary. Punched card or electronic computers should provide an answer to this difficulty.

(*c*) *Sign determination for other classes of structure factors*

This paper has so far dealt with the determination of signs for reflexions with even indices. These reflexions are the structure invariants; their signs do not depend on the choice of origin. In general the sign of a reflexion *will* change when the origin is moved but the signs of reflexions with indices such as (H,K) and $(H+2h,K+2k)$ will vary together so that their *product* will be a structure invariant. The following discussion will show how the probable sign of such structure invariants may be found for the two-dimensional space groups *pgg* and *p4g*.

The appropriate equations are:

$$s(H+h,K+k') \approx s(HK)s(hk') , \quad (13)$$

$$s(H+h,K+k')s(h\bar{k}') \approx s(H+2h,K) , \quad (14)$$

$$s(H+2h+h',K+k) \approx s(H+2h,K)s(h'k) , \quad (15)$$

$$s(H+2h+h',K+k)s(h'\bar{k}) \approx s(H+2h,K+2k) . \quad (16)$$

By equating the product of the left-hand side of the four equations to the product of the right-hand side it is found that

$$s(H+2h,K+2k) \approx s(HK)(-1)^{h+h'+k+k'} . \quad (17)$$

There are other routes by which $s(H+2h,K+2k)$ may be related to $s(HK)$ for a given h' and k'. For example:

$$s(H+h,K+k') \approx s(HK)s(hk') ,$$
$$s(H+h,K+k')s(h'k) \approx s(H+h+h',K+k+k') ,$$
$$s(H+2h+h',K+k) \approx s(H+h+h',K+k+k')s(h\bar{k}') ,$$
$$s(H+2h+h',K+k)s(h'\bar{k}) \approx s(H+2h,K+2k) .$$

These again lead to (17), although the probabilities of the various steps are quite different from those of equations (13)–(16).

The weight $W_{h'k'}$ will be the sum of the composite excess probabilities for the six possible routes for a given h' and k'.

The disadvantage of this procedure is that each sign indication is based on four steps and the probability of the indication will be correspondingly weaker. A slightly compensating advantage is the increase of the number of terms in $W_{h'k'}$.

The extension of the results of this section to three dimensions is easily accomplished and more powerful sign indications may be found in this way.

7. Conclusions

The above results would seem to indicate that sign relationships assisted by structural symmetry can sometimes yield useful information from the intensities alone. However, it cannot be too strongly stressed that the signs determined by these methods should be treated circumspectly until experience demonstrates more clearly the reliability of the results they yield. The most straightforward method, at the present time, of estimating this reliability for a particular crystal-structure investigation seems to be to calculate signs for a known structure of the same complexity.

The results given in this paper bear a superficial resemblance to those given by Hauptman & Karle (1953). The claims of these authors are disputed by Vand & Pepinsky (1953) and similarly by Cochran & Woolfson (1954), who claim that important equations given by Hauptman & Karle have a simple interpretation in terms of Patterson syntheses and that this shows that their method cannot be a general solution of the phase problem. The methods developed here do not appear to have a similar interpretation.

Not all the possible applications of the symmetry of the reflexions have been considered in the above discussion. Each problem should be studied individually and equations should be developed as they are required. Care should be taken to avoid the positive sign paradox since, if a proportion of the equations always indicate a positive sign while the remainder may give either sign, the results will be biased in favour of positive signs.

The author wishes to thank Dr W. Cochran for many discussions during which the ideas incorporated in this paper were evolved. The referee is also thanked for his many constructive criticisms and suggestions which led to the preparation of an improved version of this paper. The financial assistance of the Department of Scientific and Industrial Research is gratefully acknowledged.

References

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65.
COCHRAN, W. (1953). *Acta Cryst.* **6**, 810.
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.
MCDONALD, T. R. R. & BEEVERS, C. A. (1952). *Acta Cryst.* **5**, 654.
VAND, V. & PEPINSKY, R. (1953). *The Statistical Approach to X-ray Structure Analysis.* State College: X-ray and Crystal Analysis Laboratory of the Pennsylvania State University.
WOOLFSON, M. M. (1954*a*). *Acta Cryst.* **7**, 61.
WOOLFSON, M. M. (1954*b*). *Acta Cryst.* **7**, 65.
ZACHARIASEN, W. H. (1952) *Acta Cryst.* **5**, 68.