*Acta Cryst.* (1972). A**28**, 303

# Application of Weighted Phase-Sum Formulae for Phase Determination

By Rolf Norrestam

*Institute of Inorganic and Physical Chemistry, University of Stockholm, S*–10405 *Stockholm, Sweden*

(*Received* 29 *December* 1971 *and in revised form* 22 *February* 1972)

The simple 'weighted sum formula' has been studied as regards its applicability to phase refinements. These studies show that proper weighting of the phase relations significantly improves the phase estimates from phase-sum formulae. The evaluation of weights used in the 'variance-weighted sum formula' is discussed, together with the problem of the $2\pi$ ambiguity. A simple procedure for selecting efficient basis sets is also given. The procedures described have been programmed and used to solve twelve unknown structures. Applications to somewhat complicated cases are described.

## Introduction

The investigations of *inter alios* Sayre (1952), Cochran (1952), Hauptman & Karle (1953) and Hughes (1953) show that the sum

$$\varphi(\mathbf{h})+\varphi(-\mathbf{k})+\varphi(\mathbf{k}-\mathbf{h})\,, \tag{1}$$

where $\varphi(\mathbf{h})$ is the phase of the structure factor $F(\mathbf{h})$, is distributed about zero (modulo $2\pi$). Studies by *inter alios* Cochran (1955) and Karle & Hauptman (1956) have shown (*cf.* Karle & Karle, 1966) that given $m$ sums $\varphi(-\mathbf{k}_r)+\varphi(\mathbf{k}_r-\mathbf{h})$ $(r=1,2,\ldots,m)$ and the corresponding normalized structure-factor values $|E(-\mathbf{k}_r)|$ and $|E(\mathbf{k}_r-\mathbf{h})|$ $(r=1,2,\ldots,m)$, the maximum of the probability distribution of the sums approximately satisfies

$$\sum_{r=1}^{m}|E(-\mathbf{k}_r)\cdot E(\mathbf{k}_r-\mathbf{h})|\cdot\sin\left[\varphi(\mathbf{h})+\varphi(-\mathbf{k}_r)+\varphi(\mathbf{k}_r-\mathbf{h})\right]=0\,. \tag{2}$$

Equation (2) becomes, by rearrangement and use of Friedel's law:

$$\tan\varphi(\mathbf{h})=\frac{\sum_{r=1}^{m}|E(\mathbf{k}_r)\cdot E(\mathbf{h}-\mathbf{k}_r)|\cdot\sin\left[\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)\right]}{\sum_{r=1}^{m}|E(\mathbf{k}_r)\cdot E(\mathbf{h}-\mathbf{k}_r)|\cdot\cos\left[\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)\right]}\,, \tag{3}$$

the 'tangent formula' (Karle & Hauptman, 1956).

On the other hand, expansion of the sine function of (2) as a Taylor series (Karle & Karle, 1966), with neglect of all terms except the first, yields

$$\varphi(\mathbf{h})=\frac{\sum_{r=1}^{m}|E(\mathbf{k}_r)\cdot E(\mathbf{h}-\mathbf{k}_r)|\left[\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)\right]}{\sum_{r=1}^{m}|E(\mathbf{k}_r)\cdot E(\mathbf{h}-\mathbf{k}_r)|}\,, \tag{4}$$

the 'weighted sum formula' (Karle & Karle, 1966). In applying this formula one must somehow deal with the $2\pi$ ambiguity of each contributing phase sum, $\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)$.

An expression given by Karle & Karle (1966) for the variance of phase sums, $\varphi(\mathbf{h})-\varphi(\mathbf{k})-\varphi(\mathbf{h}-\mathbf{k})$, has been tabulated by Germain, Main & Woolfson (1970) for the variance, $V(\mathbf{h},\mathbf{k})$, as a function of $\kappa(\mathbf{h},\mathbf{k})=2\sigma_3\sigma_2^{-3/2}|E(\mathbf{h})\cdot E(\mathbf{k})\cdot E(\mathbf{h}-\mathbf{k})|$. Here $\sigma_n=\sum_{j=1}^{N}Z_j^n$, where $N$ is the number of atoms in the unit cell and $Z_j$ the atomic number of the $j$th atom. $V(\mathbf{h},\mathbf{k})$ as a function of $\kappa(\mathbf{h},\mathbf{k})$, is given in Table 1.

Table 1. *Comparison of the theoretically estimated variances for single relations with the approximations* $1/\kappa$ and $1/(\kappa-0{\cdot}7)$

| $\kappa$ | Variance in (radian)$^2$ | $1/\kappa$ | $1/(\kappa-0{\cdot}7)$ |
|---|---|---|---|
| 1·0 | 1·604 | 1·000 | 3·333 |
| 1·5 | 1·091 | 0·667 | 1·250 |
| 2·0 | 0·764 | 0·500 | 0·769 |
| 2·5 | 0·563 | 0·400 | 0·556 |
| 3·0 | 0·437 | 0·333 | 0·435 |
| 3·5 | 0·354 | 0·286 | 0·357 |
| 4·0 | 0·298 | 0·250 | 0·303 |
| 5·0 | 0·227 | 0·200 | 0·232 |
| 6·0 | 0·184 | 0·167 | 0·188 |
| 8·0 | 0·134 | 0·125 | 0·136 |
| 10·0 | 0·106 | 0·100 | 0·107 |
| 14·0 | 0·074 | 0·071 | 0·075 |
| 20·0 | 0·051 | 0·050 | 0·052 |

As a rule most of the 'strongest' of the $m$ contributors $\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)$ to the phase $\varphi(\mathbf{h})$ are clustered within one quadrant of the phase-circle. For this reason it seems reasonable to determine the half-region of the phase-circle which contains the maximum number of contributors and to evaluate the value of $\varphi(\mathbf{h})$ only from these indications while considering the remaining indications as contradictions. Thus, it should be possible to recognize most of the inconsistencies among the contributors and also to take account of them in a proper way when determining $\varphi(\mathbf{h})$. An analysis of the distribution of the phase contributors is obviously easier with a linear phase-determining formula than with a non-linear phase-determining formula if only contributors within one half of the phase-circle are accepted. Furthermore, linear formulae provide simple relations to be used to calculate proper weights based on the estimated variances for the individual contributors. When using a more elaborate weighting with the tangent formula, *viz.* the 'weighted tangent formula', the weights applied are usually more or less empirical (*cf.* Germain, Main & Woolfson, 1970). Because of the fairly simple calculations that are needed with linear formulae, one can decrease significantly the computer time (and size) used. Thus, several computational aspects speak in favour of using a simple weighted linear phase-sum formula for computerized phase determination.

### Description of the formulae used

In the computerized procedure for phase determination described in this paper, formulae related to (4) are used exclusively. Assume that $m$ pairs $\varphi(-\mathbf{k}_r)+\varphi(\mathbf{k}_r-\mathbf{h})$ $(r=1,2,\ldots,m)$ are given together with the corresponding normalized structure-factor magnitudes $|E(\mathbf{h})|$, $|E(-\mathbf{k}_r)|$ and $|E(\mathbf{k}_r-\mathbf{h})|$.

If, for each individual relation $\varphi(\mathbf{h})+\varphi(-\mathbf{k}_r)+\varphi(\mathbf{k}_r-\mathbf{h})=0$, an individual weight $w(\mathbf{h},\mathbf{k}_r)$ is used, the weighted average value of $\varphi(\mathbf{h})$ is given by

$$\varphi(\mathbf{h})=\frac{\sum\limits_{r=1}^{m} w(\mathbf{h},\mathbf{k}_r)\,[\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)]}{\sum\limits_{r=1}^{m} w(\mathbf{h},\mathbf{k}_r)}\,. \tag{5}$$

A suitable weight would be $1/V(\mathbf{h},\mathbf{k}_r)$, the inverse of the variance of each phase relation used. As shown in Table 1, $V(\mathbf{h},\mathbf{k}_r)$ is roughly proportional to the inverse of $\kappa(\mathbf{h},\mathbf{k}_r)$ for large $\kappa$ values. If this approximation is substituted in (5), one immediately obtains (4). Thus, formula (4) can be interpreted not only as a rough first-order approximation to the tangent formula (3), but also as an approximation for solving the system of weighted linear phase relations given by (1). As seen from Table 1, a fundamental disadvantage of using the approximation $V(\mathbf{h},\mathbf{k})\simeq 1/\kappa(\mathbf{h},\mathbf{k})$, is that the variances of 'weaker' relations, those with smaller $\kappa$ values, will be underestimated as compared with those of the 'stronger' ones. This effect, quite in opposition to the principles usually adopted for proper weighting, is already appreciable at $\kappa=4$ and increases rapidly with decreasing $\kappa$. Since $\kappa$ is less than 4 for a large fraction of the relations used in typical phase determinations, a better approximation for estimating the variances used for calculating the weights in formula (5) is clearly desirable.

The approximation $V(\mathbf{h},\mathbf{k})\simeq 1/[\kappa(\mathbf{h},\mathbf{k})-0{\cdot}7]$ can be used to estimate the variances with good precision (relative errors below 3% when compared to the theoretically calculated values) for $\kappa$ larger than about 2, (Table 1), and has the further advantage of overestimating the variances of relations with smaller $\kappa$'s. This approximation does not serve well to estimate the variances of relations with extremely small $\kappa$'s; in the structures so far investigated using the procedure described in this paper, the lower limit of the $\kappa$'s of the relations used has been in the range 1·0 to 1·8. Use of the approximation to estimate $w(\mathbf{h},\mathbf{k})=1/V(\mathbf{h},\mathbf{k})$ leads to a modification of formula (4):

$$\varphi(\mathbf{h})=\frac{\sum\limits_{r=1}^{m}[\kappa(\mathbf{h},\mathbf{k}_r)-0{\cdot}7]\,[\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)]}{\sum\limits_{r=1}^{m}[\kappa(\mathbf{h},\mathbf{k}_r)-0{\cdot}7]}\quad\text{for } \kappa\text{'s}\geq 1 \tag{6}$$

with weighting in much better accord with the rigorous estimates of variance.

The contributors $\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)$ to a phase $\varphi(\mathbf{h})$ are usually not known exactly, but have been determined with some estimate of their variances, $v(\mathbf{k}_r)$ and $v(\mathbf{h}-\mathbf{k}_r)$. These may be taken into account in the calculation of the phase $\varphi(\mathbf{h})$ by modifications of the phase-determining formulae, which are fortunately simple for linear relations such as (5). The variance of each estimate of $\varphi(\mathbf{h})$ is no longer given by $V(\mathbf{h},\mathbf{k}_r)$, but by $V(\mathbf{h},\mathbf{k}_r)+v(\mathbf{k}_r)+v(\mathbf{h}-\mathbf{k}_r)$. Thus, the individual weights $w(\mathbf{h},\mathbf{k}_r)$ to be used in (5) are

$$w(\mathbf{h},\mathbf{k}_r)=1/[V(\mathbf{h},\mathbf{k}_r)+v(\mathbf{k}_r)+v(\mathbf{h}-\mathbf{k}_r)]\,. \tag{7}$$

It is prudent to overestimate the variance $v(\mathbf{h})$ of $\varphi(\mathbf{h})$ when the $m$ contributors, $\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)$ $(r=1,2,\ldots,m)$, have large inconsistencies, thus reducing the propagation of errors during the phase determining process. A modified 'variance', $v(\mathbf{h})$, has accordingly been calculated as

$$v(\mathbf{h})=\frac{\sum\limits_{r=1}^{m} w(\mathbf{h},\mathbf{k}_r)\cdot\Delta_r^2}{(p-1)\sum\limits_{r=1}^{m} w(\mathbf{h},\mathbf{k}_r)} \tag{8}$$

where $\Delta_r=\varphi(\mathbf{h})-\varphi(\mathbf{k}_r)-\varphi(\mathbf{h}-\mathbf{k}_r)$ is calculated to be in the range 0 to $\pi$, and where $p$ is the number of contributors having $\Delta$ less than $\pi/2$.

If there is only one contributor, the variance of $\varphi(\mathbf{h})$ is calculated as $V(\mathbf{h},\mathbf{k}_1)+v(\mathbf{k}_1)+v(\mathbf{h}-\mathbf{k}_1)$.

The procedure adopted for evaluating $\varphi(\mathbf{h})$ from the $m$ contributors $\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)$ is not rigorously that given by (5). Inspection of the phase distribution of the contributors in a real case shows that with only a few exceptions these are usually clustered within a phase range of $\pi$ radians. The exceptions are considered as contradictions and the summations to be done in formula (5) are performed only over the $p$ contributors having $\Delta$'s less than $\pi/2$. These $p$ contributors are extracted from the $m$ by an iterative calculation of $\varphi(\mathbf{h})$. The first value of $\varphi(\mathbf{h})$ is calculated by (5) using all the $m$ contributors, and the corresponding $\Delta_r$'s are calculated. The largest $\Delta$ is examined and if its value is larger than $\pi/2$, a second value of $\varphi(\mathbf{h})$ is calculated by (5), without the corresponding contributor. All the remaining $(m-1)$ $\Delta$'s are recalculated and examined. Such iterative calculations of successive new values of $\varphi(\mathbf{h})$ are performed $n$ times until either all the $p=m-n$ remaining $\Delta$'s are less than $\pi/2$, or $n$ equals $m/2$. However, as stated above, all the $m$ contributors are used to calculate the estimated variance $v(\mathbf{h})$ of the $\varphi(\mathbf{h})$ value finally obtained. If the phase has a value restricted by space group symmetry, the nearest allowed value is assigned to $\varphi(\mathbf{h})$ before the calculation of its variance.

Thus, the formula used for phasing can be written as

$$\varphi(\mathbf{h})=\frac{\sum_{s=1}^{n} w(\mathbf{h},\mathbf{k}_s)\,[\varphi(\mathbf{k}_s)+\varphi(\mathbf{h}-\mathbf{k}_s)]}{\sum_{s=1}^{n} w(\mathbf{h},\mathbf{k}_s)} \tag{9}$$

where the contributors $\varphi(\mathbf{k}_s)+\varphi(\mathbf{h}-\mathbf{k}_s)$ $(s=1,2,\ldots,p)$, a subset of all the $m$ contributors, are chosen so that $m/2\le p\le m$, where $p$ is the maximum number of contributors having $\Delta$'s less than $\pi/2$.

### The problem of the 2π ambiguity

Application of linear phase relations such as formulae (4) or (5) has one disadvantage as compared with *e.g.* the tangent formula (3), *viz.* the $2\pi$ ambiguity inherent in the periodicity of the phase sums. Karle & Karle (1966) have shown one way out of this dilemma: another way, to be used with formula (9), will be described here. Suppose that there are $m$ contributors, normalized (by adding or subtracting $2\pi$) so that $-\pi<\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)\le\pi$, with the weights $w(\mathbf{h},\mathbf{k}_r)$ $(r=1,2,\ldots,m)$. The quantity $\alpha(\mathbf{h})$ calculated as

$$\alpha(\mathbf{h})=\frac{\sum_{r=1}^{m} w(\mathbf{h},\mathbf{k}_r)|\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)|}{\sum_{r=1}^{m} w(\mathbf{h},\mathbf{k}_r)} \tag{10}$$

can then be used to test whether the $2\pi$ ambiguity needs to be taken into consideration or not. If $\alpha(\mathbf{h})\le\pi/2$ formula (9) can be used directly. If $\alpha(\mathbf{h})>\pi/2$, a phase shift of $\pi$ is subtracted from all the $m$ contributors and the results are normalized to be in the range from $-\pi$ to $\pi$. Now formula (9) can be used to determine $\varphi(\mathbf{h})-\pi$ instead of $\varphi(\mathbf{h})$.

### Selection of basis set

Various procedures for selecting basis sets have been described by *inter alios* Karle & Karle (1966), Hauptman, Fisher & Hancock (1969) and Germain, Main & Woolfson (1970). The alternative computerized procedure described here, based on an analysis of the results obtained during a simulated phasing process without the use of any explicit phase values, was written to select a small but efficient basis set.

First, for every reflexion the number of triple phase relations ('frequency') and the sum of the corresponding estimated reciprocals of the variances, $1/V$'s, is determined and filed. From this file the first origin-specifying reflexion is selected, following the general rules given by Hauptman & Karle (1956) and Karle & Hauptman (1961). The criterion used for selecting a reflexion is that it should enter into as many reliable triple relations as possible, *viz.* have high values of 'frequency' and sum of $1/V$'s. This reflexion is now considered to have a 'known' phase, and will thus yield a large number of phase relations containing two instead of three 'unknowns', which will be termed double relations. The distribution for the reflexions of frequency and sum of $1/V$'s in both triple and double relations is calculated. For a structure with a primitive unit cell having orthorhombic or lower symmetry as an example, two additional origin-defining reflexions are required. These are selected one at a time from the reflexions currently having high values of frequency and sum of $1/V$'s in triple relations. Thus, the origin is defined by the use of reflexions involved in as many

Table 2. *Sign shifts of the real (A) and imaginary (B) parts of structure factors, caused by origin shifts (Δ) and by origin shifts together with inversion (i)*

| | Parity group: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\Delta$ | *ggg* | *ggu* | *gug* | *ugg* | *guu* | *ugu* | *uug* | *uuu* |
| $(0, 0, 0)$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $A$, $B$ |
| $(0, 0, \frac{1}{2})$ | $A$, $B$ | $-A$, $-B$ | $A$, $B$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $A$, $B$ | $-A$, $-B$ |
| $(0, \frac{1}{2}, 0)$ | $A$, $B$ | $A$, $B$ | $-A$, $-B$ | $A$, $B$ | $-A$, $-B$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ |
| $(\frac{1}{2}, 0, 0)$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $-A$, $-B$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $-A$, $-B$ |
| $(0, \frac{1}{2}, \frac{1}{2})$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $A$, $B$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $A$, $B$ |
| $(\frac{1}{2}, 0, \frac{1}{2})$ | $A$, $B$ | $-A$, $-B$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $A$, $B$ | $-A$, $-B$ | $A$, $B$ |
| $(\frac{1}{2}, \frac{1}{2}, 0)$ | $A$, $B$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $-A$, $-B$ | $-A$, $-B$ | $A$, $B$ | $A$, $B$ |
| $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ | $A$, $B$ | $-A$, $-B$ | $-A$, $-B$ | $-A$, $-B$ | $A$, $B$ | $A$, $B$ | $A$, $B$ | $-A$, $-B$ |
| $(0, 0, 0)+i$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ |
| $(0, 0, \frac{1}{2})+i$ | $A$, $-B$ | $-A$, $B$ | $A$, $-B$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $A$, $-B$ | $-A$, $B$ |
| $(0, \frac{1}{2}, 0)+i$ | $A$, $-B$ | $A$, $-B$ | $-A$, $B$ | $A$, $-B$ | $-A$, $B$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ |
| $(\frac{1}{2}, 0, 0)+i$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $-A$, $B$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $-A$, $B$ |
| $(0, \frac{1}{2}, \frac{1}{4})+i$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $A$, $-B$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $A$, $-B$ |
| $(\frac{1}{2}, 0, \frac{1}{2})+i$ | $A$, $-B$ | $-A$, $B$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $A$, $-B$ | $-A$, $B$ | $A$, $-B$ |
| $(\frac{1}{2}, \frac{1}{2}, 0)+i$ | $A$, $-B$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $-A$, $B$ | $-A$, $B$ | $A$, $-B$ | $A$, $-B$ |
| $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})+i$ | $A$, $-B$ | $-A$, $B$ | $-A$, $B$ | $-A$, $B$ | $A$, $-B$ | $A$, $-B$ | $A$, $-B$ | $-A$, $B$ |

different reliable triple relations as possible, so as to give a large number of relations involving two 'unknowns' (double relations). The number of relations involving one 'unknown' (single relations) will be very small at this stage.

Any other reflexions having known phases, *e.g.* from $\sum_1$ relations, are now introduced, the distributions are reanalysed, and an enantiomorph determining reflexion is selected from the reflexions having the highest values of frequency and sum of $1/V$'s in double relations. Then the distribution analysis is repeated. The selected enantiomorph-defining reflexion will, as it enters into several double relations, produce a number of single relations.

The distribution of the reflexions in single relations is now examined and those reflexions having the sums of $1/V$'s in single relations above a given limit are considered as 'known'. These reflexions are now used to further simplify the equation system by producing new double and single relations from triple and double relations respectively. These operations constitute one cycle of a cyclic procedure which is repeated until no new single relations are produced. The reflexion having the highest values of frequency and sum of $1/V$'s in double relations in the first few cycles is selected and considered as a variable to be added to the basis set. This reflexion is used to reduce the system of phase relations through application of the cyclic procedure described above. As many such variables are selected one at a time as are needed to yield single relations for most of the reflexions within five cycles.

Thus, the basis set selected by this simply programmable procedure yields as many new single relations as possible directly from the basis set, and most of the remaining reflexions will occur in single relations within a few cycles. This last property reduces both the propagation of errors and computer time.

Space group symmetry restrictions can be used to reduce the number of alternative solutions. Furthermore, relations containing two unknown but restricted phase values and an unknown general phase yield, for orthorhombic or lower symmetry, the absolute value of the latter, modulo $\pi$, by the use of formula (6), since each restricted contributor $\varphi(\mathbf{k}_r)+\varphi(\mathbf{h}-\mathbf{k}_r)$ is known to be 0 or $\pi/2$, modulo $\pi$. The same kind of information is obtained in a more general way from the 'squared tangent formula' (Hauptman, 1970).

When reflexions of unknown general phase are included in the basis set to specify origin or enantiomorph, tables such as Table 2 are convenient to facilitate specification of the ranges to be investigated for these phases. If proper rows and columns of Table 2 are deleted, tables for various centred cells are easily obtained; the permitted equivalent origins for *C*-centred cells of crystal class {222} for example, are $0,0,0$, $0,\frac{1}{2},0$, $\frac{1}{2},0,\frac{1}{2}$ and $0,0,\frac{1}{2}$. For primitive cells of crytals class {222} one finds from Table 2 that, if two reflexions from the parity groups *ggu* and *gug* are assigned phase value 0 for origin specification, the necessary third reflexion must distinguish between the origins $0,0,0$ and $\frac{1}{2},0,0$, and thus must belong to parity groups *ugg*, *ugu*, *uug* or *uuu*. Let it be chosen from parity group *ugg* and assigned as an example the phase $\pi/2$. A fourth reflexion is then necessary to specify the enantiomorph, since a shift of $\frac{1}{2},0,0$ and a subsequent inversion causes no changes of the phases of the three chosen reflexions. From Table 2 we obtain the following restrictions on the ranges of the phase values, for reflexions of the eight different parities, that must be fulfilled if the enantiomorph is to be determined: $0<\varphi(ggg)<\pi$, $0<\varphi(ggu)<\pi$, $0<\varphi(gug)<\pi$, $-\pi/2<\varphi(ugg)<\pi/2$, $0<\varphi(guu)<\pi$, $-\pi/2<\varphi(ugu)<\pi/2$, $-\pi/2<\varphi(uug)<\pi/2$ and $-\pi/2<\varphi(uuu)<\pi/2$.

### Description of phasing procedure

The procedures discussed above have been programmed in basic FORTRAN IV as an integrated set of programs, to fit into the 14 K 16-bit words available for background calculations in our IBM 1800 computer.

The first program of the set generates triple relations with estimated variances and symmetry-imposed phase restrictions, from input of $|E|$ values, unit cell composition, and equivalent positions.

The second program facilitates the selection of the basis set which yields the most single indications within a minimum number of phasing cycles (*cf.* previous paragraph).

The third program solves the variance-weighted linear phase relations [*cf.* formula (7)] from an input basis set. The strategy used in this program is similar to that which proved successful in a previous program (Norrestam, 1971) limited to centrosymmetric structures. The input basis set is used to initialize the phasing. New phases are approved if their e.s.d.'s are below $\pi/4$ radians. From this preliminary set of phases

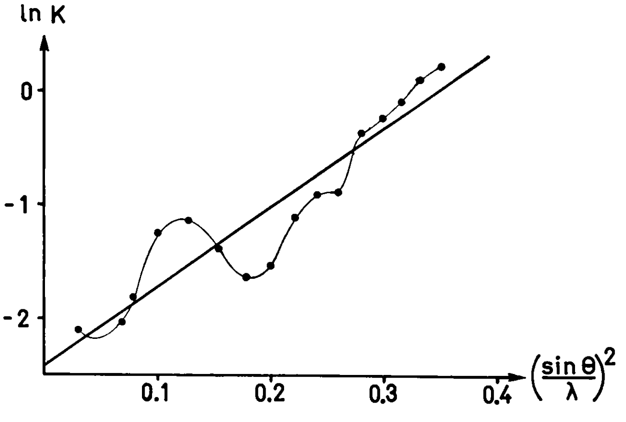

Fig. 1. Plot of ln *K versus* $(\sin\theta/\lambda)^2$ together with the least-squares line, for 4a-allyl-3,5,7,8,10-pentamethyl-4a,5-dihydroisoalloxazine. *K* was evaluated as $\langle\sum f^2\rangle/\langle|F|^2_{\rm obs}\rangle$ for sin $\theta$ intervals containing equal numbers (80) of reflexions. The curve joining the points was plotted from a second-order Lagrange interpolation using Aitkens's method.

for the reflexions, an expansion of the basis set is performed stepwise to reduce the propagation of errors. Only those phases having e.s.d.'s below $\pi/6$ and multiplicities [$p$ in formula (9)] greater than or equal to a given number $q$ in the previous step, are added to the basis set. The extended basis set thus obtained is then used to solve the triple relations once more. New phases obtained during this phasing procedure are approved only if their e.s.d.'s are below $\pi/6$ radians and if they have multiplicities greater than or equal to $q$. When no more new phases are approved, the value of $q$ is decreased, a new, more extensive basis set is formed and the above procedure is repeated. Usually three iterations, with $q$ equal to 4, 3 and 2, have been performed. Finally the phases are refined for two cycles with formula (9), one reflexion at a time. For a given reflexion, all phase indications are used which arise from reflexions whose phases currently have e.s.d.'s below $\pi/6$ radians and multiplicities greater than 2.

Four different figures of merit are evaluated to facilitate the choice of the best solution, *viz.* the number of phases determined with e.s.d.'s below $\pi/6$, the number of triple relations having phase sums (calculated from the predicted phases) between $-\pi/6$ and $\pi/6$, the r.m.s. deviation of the phase sums from zero, and the peak height at the origin $\sum_i |E_i| \cos \varphi_i$ (where $\varphi_i$ is the predicted phase for reflexion number $i$). Thus, high values for the first two figures of merit and low values for the last two are expected for the best solutions.

## Applications

The procedures described above have been used at our Institute to solve twelve unknown noncentrosymmetric crystal structures, whose sizes range from small inorganics such as copper(II)molybdate $Cu_3Mo_2O_9$ (Kihlborg, Norrestam & Olivecrona, 1971) to fairly complex organic structures such as the cyclic hexapeptide ferrichrysin and the antibiotic oligomycin (von Glehn, Norrestam, Kierkegaard, Maron & Ernster, 1972). The composition of the asymmetric unit for ferrichrysin is $C_{29}H_{48}FeN_9O_{15}$, while oligomycin B has an asymmetric unit of 60 non-hydrogen atoms, with a probable composition, indicated by the X-ray structure, of about $C_{47}H_{76}O_{13}$. The remaining 9 structures have asymmetric units of 15–25 non-hydrogen atoms (see *e.g.* Norrestam, 1972).

For the moderately complex structures, about ten times as many high $|E|$ values as there are non-hydrogen atoms in the asymmetric unit have been used to generate triple relations. Usually the lower limit of the allowed e.s.d.'s of the triple relations has been set to yield about 4–5 times as many relations as the number of $|E|$ values used. Thus, the average number of terms used in the summations in formula (9) has been about 12–15.

The average deviations between the generated phases and those calculated after the completion of the structure determinations have been in the range 8–15° for the 'equal atom' cases. The average deviations, with symmetry-restricted phase values omitted, have been in the range 10–20°. However, for the structure of 4a-allyl-3,5,7,8,10-pentamethyl-4a,5-dihydroisoalloxazine ($C_{18}H_{22}N_4O_2$) (Norrestam, 1972), somewhat higher average deviations were obtained, plausibly attributable to extensive vector overlap in Patterson space (see *e.g.* Hauptman, 1964) arising from the roughly planar molecules formed of hexagons. This vector overlap is also indicated in the Wilson plot, Fig. 1, by the prominent peak corresponding to an interatomic separation of 1·4–1·5 Å. For this structure both the usual weighted sum formula (4) and the modified variance-weighted sum formula (9) were used to refine phases. The average deviations obtained (symmetry-restricted phase values omitted) of 31 and 23° respectively, show that the latter weighting [formula (7)] produces significantly improved phase estimates.

In the first attempts to solve the oligomycin B structure (space group $P2_12_12_1$), the basis sets were selected by conventional manual techniques. Although the basis sets seemed to yield satisfactory solutions of the phase-relations, the calculated $E$ maps were discouraging in that they showed no recognizable partial structures. Furthermore, the number of alternative solutions was overwhelming since the basis sets included several reflexions with unrestricted phase values. Use of the stepwise procedure described above to select an efficient basis set enabled a choice of starting reflexions (all having symmetry-restricted phase values) for which the number of alternative solutions was reduced to four. The figures of merit obtained for the different solutions clearly revealed the best solution. Two consecutive phase refinements incorporating partial structure information, subsequent difference electron-density maps and least-squares refinements yielded the complete structure. A more detailed description of the structure determination is given elsewhere (Norrestam & von Glehn, 1972). It is notable that although the asymmetric unit of this structure contained 60 non-hydrogen atoms, the 290 highest $|E|$ values, and the 1500 best triple relations among them, sufficed.

In determining the structure of a photolysis product (space group $P2_12_12_1$) by direct methods, Karle, Karle & Estlin (1967) encountered difficulties arising from bad phase estimates for some reflexions. Using their data, the author used the 100 highest $|E|$ values to generate 556 phase relations. With a basis set selected as above, four alternative solutions were obtained. For the best solution, and the 90 reflexions having e.s.d.'s below $\pi/6$ radians, the average deviation of the generated phases from the final calculated ones was 15° when using formula (4), and 12° with formulae (7) and (9).

No more than four alternative solutions have been generated for the structures dealt with so far, when the basis sets were selected as described above. The best solution has always been one of the alternatives having the best figures of merit.

I am greatly indebted to Professor Peder Kierkegaard for his encouraging interest in this work and for the facilities placed at my disposal. Thanks are also due to Dr Don Koenig for valuable comments on the manuscripts and for his correction of the English of this paper.

### References

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

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274.

GLEHN, M. VON, NORRESTAM, R., KIERKEGAARD, P., MARON, L. & ERNSTER, L. (1972). *F.E.B.S. letters*, **20**, 267.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421.

HAUPTMAN, H. (1970). *Acta Cryst.* **B26**, 531.

HAUPTMAN, H., FISHER, J. & HANCOCK, H. (1969). *Acta Cryst.* **B25**, 811.

HAUPTMAN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136.

HAUPTMAN, H. & KARLE, J. (1956). *Acta Cryst.* **9**, 45.

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

KARLE, I. L., KARLE, J. & ESTLIN, J. A. (1967). *Acta Cryst.* **23**, 494.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.

KARLE, J. & HAUPTMAN, H. (1961). *Acta Cryst.* **14**, 217.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.

KIHLBORG, L., NORRESTAM, R. & OLIVECRONA, B. (1971). *Acta Cryst.* **B27**, 2066.

NORRESTAM, R. (1971). *Acta Chem. Scand.* **25**, 1040.

NORRESTAM, R. (1972). *Acta Cryst.* **B28**, 1713.

NORRESTAM, R. & VON GLEHN, M. (1972). To be published.

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

---

*Acta Cryst.* (1972). A**28**, 308

# The One-Dimensional Anti-Phase Domain Structures. III. An Alternative Interpretation of the Structure with a Non-Integral Value of the Half Period, $\tilde{M}$

BY TERUAKI MINAGAWA*

*Department of Physics, Faculty of Science, Osaka City University,* 459 *Sugimoto-cho, Sumiyoshi-ku, Osaka, Japan*

(*Received* 24 *August* 1971 *and in revised form* 17 *November* 1971)

The non-integral value for the half period, $\tilde{M}(M<\tilde{M}<M+1$, $M$:integer), as experimentally found in one-dimensional anti-phase domain structures, is explained by a disordered structure consisting of a mixture of two kinds of structure units with thicknesses of $M$ and $M+1$ layers. The diffraction intensity for the disordered structure is calculated by the use of the general diffraction theory for a one-dimensionally disordered crystal given by Kakinoki & Komura [*Acta Cryst.* (1965), **19**, 137]. With this model the non-integral value of the half period, $\tilde{M}$, can be explained as due to the shift between two peaks, one due to a simple APD (anti-phase domain) structure with half period, $M$, and the other to another simple APD structure with half period, $M+1$. It is not necessary to consider a very large value of the period as was required in Fujiwara's model [Fujiwara, *J. Phys. Soc. Japan* (1957), **12**, 7], which was proposed assuming a disordered structure, deviating from the standard structure defined by a step function.

## 1. Introduction

In some examples of one-dimensional anti-phase domain structures with an out-of-step vector $\mathbf{u}=(\mathbf{a}+\mathbf{b})/2$, the half period, $\tilde{M}$, has experimentally been found to be non-integral, as shown in Table 1. Fujiwara (1957) explained this by assuming a disordered structure deviating from a standard structure which is defined by the use of a step function. The interpretation of the non-integral structure and the relevant intensity equations were discussed in detail in part II of this series (Kakinoki & Minagawa, 1972). In Fujiwara's interpretation, however, the period $P$ of the standard structure should be subject to the relation

$$P=2\nu\tilde{M} \qquad (1)$$

where $\nu$ is the minimum positive integer to make $2\nu\tilde{M}$ equal to an integer $P$ [refer to equation (F–4)†]. Therefore, if we put

$$\tilde{M}=M+\Delta M \quad \text{with} \quad 0<\Delta M<1 \qquad (2)$$

where $M$ is an integer, then we have to assume a very high value of $P$, *e.g.*, $P=321$ for $\tilde{M}=3{\cdot}21$ ($M=3$,

---

* Present address: Department of Physics, Osaka Kyoiku University, Tennoji, Osaka, Japan.

† Equation (F–4) means equation (4) in Fujiwara's (1957) paper.