

## On the Application of Phase Relationships to Complex Structures. X. *MAGLIN* – A Successor to *MULTAN*

BY M. M. WOOLFSON

*Department of Physics, University of York, England*

(Received 11 June 1976; accepted 12 September 1976)

The *MULTAN* system has almost reached the limit of possible development. For complicated structures the starting set must be so large that even the largest computers cannot handle the number of phase permutations required. Another difficulty is that for some structures a correct set of phases is unstable under tangent-formula refinement. In the *MAGLIN* program now being developed initial sets of phases will be found for 30 or so reflexions by an application of magic integers to Karle–Hauptman determinants with the use of the Tsoucaris maximum-determinant rule. Further phases are then found by repeated application of magic integers. Phase refinement is carried out by a least-squares solution of a set of linear equations with each  $\Sigma_2$  relationship represented by one equation. A novel technique is described whereby a set of equations involving  $M$  reflexions may be solved for the phases of only  $m (< M)$  of them. This leads to a considerable saving of time in running the *MAGLIN* process. For complicated structures it is expected that *MAGLIN* will not only be more effective than *MULTAN* but also considerably less time-consuming.

### The limits of *MULTAN*

The *MULTAN* (MULTiple-solution TANGent-formula) computer program for phase determination has been widely used in the past few years for solving both centrosymmetric and non-centrosymmetric structures. While there may be dispute about the relative merits or efficacy of one method or another there can be little doubt that *MULTAN* is unrivalled at the present time in the total package of power plus convenience that it offers.

However, *MULTAN* sometimes fails and no  $E$  map can be found that shows a fragment, recognizable enough to be recycled to give the complete structure. The pattern of failure is rather difficult to interpret. In general the fewer atoms there are in the asymmetric unit the easier and more straightforward is the solution but this is not always the case. While complicated structures with 80 or more atoms in the asymmetric unit have sometimes been solved with ease some of the difficult or unsolvable structures have had fewer than 30 atoms in the asymmetric unit.

The analysis of a number of *MULTAN* failures, reported by Lessinger (1976), has led to important improvements in the *MULTAN* procedure but, no matter how *MULTAN* is modified, there are two inherent limitations on its ability to solve structures.

The first limitation concerns the size of the starting set. The basic aim is to use a starting set of reflexions small enough to give a manageable number of phase permutations yet large enough to lead to a strong development of new phase information. The *CONVERGENCE* procedure has been designed to achieve this but it is nevertheless sometimes impossible to do and it is found that one or more weak links exist in the phase-development chain. Of course, this can usu-

ally be corrected by increasing the size of the starting set but the penalty is an increase by a factor of four (sometimes two) in the computer time required for each extra reflexion. This price soon becomes too penal for even the most powerful computers.

While the phase-permutation procedure is an intrinsic component of the present *MULTAN* system it could be replaced by a more efficient alternative. But that would still leave the other, more important and rather more basic limitation, the use of the tangent formula itself. The alarming fact is that for some structures even the true phases are unstable under the process of tangent-formula refinement. While this is frequently observed for structures with no translational symmetry, where phases can all steadily drift towards zero, it has also been found for structures with space groups such as  $Pca2_1$ ,  $P2_1$  etc. (Lessinger, 1976).

Although *MULTAN* remains an extremely useful system and while further improvements of it may still be possible it must be recognized that any new leap forward in automated direct methods will require a radical reappraisal of the whole process of phase determination.

Such a process of reappraisal has been in progress for about two years and has led to a number of novel ideas for phase determination. The assembly of these ideas into a new system of phase determination will now be described.

### The starting-set problem

There has been very rapid progress in the use of magic integers in the recent past (White & Woolfson, 1975; Declercq, Germain & Woolfson, 1975; Taylor & Woolfson, 1975). It is easy to interpret the use of magic integers as a way of obtaining larger starting sets. A

typical magic integer approach would consist of:

- (i) Select, say, 15 reflexions of large  $E$  and represent their phases by elements of the matrix  $\{n_1, n_2, n_3, n_4, n_5\}$  ( $xyz$ ) where the  $n$ 's are magic integers.
- (ii) By the use of single strong relationships involving two of the basic set of 15 reflexions, represent the phase of various third reflexions by magic integers.
- (iii) For the complete set of reflexions represented in magic-integer form find all the  $\Sigma_2$  relationships [other than those used in stage (ii)] and construct a  $\psi$  map

$$\psi(x, y, z) = \sum_{r=1}^m \kappa_r \cos \{2\pi(H_r x + K_r y + L_r z + b_r)\} \quad (1)$$

where there are  $m$  relationships with weights  $\kappa_r$  and  $H_r x + K_r y + L_r z + b_r$  is the magic-integer representation of the sum of three phases in the relationship.

The peaks of the  $\psi$  map give values of ( $xyz$ ) which translate into probable sets of phases for all the reflexions considered in (i) and (ii). Typically one may find 50–100 possible sets of phases for 25–40 reflexions and experience shows that one, if not several, of the sets usually gives a good approximation to the correct phases. This is a vastly superior procedure to phase permutation. It is not an expensive operation – two or three minutes on even a modest computer – but it is orders of magnitude more efficient than the present *MULTAN* practice.

A magic-integer approach of great promise involves the use of Karle–Hauptman determinants (Karle & Hauptman, 1950). One of these determinants is of the form

$$\begin{vmatrix} E_0 & E_{h_1} & E_{h_2} & E_{h_n} \\ E_{-h_1} & E_0 & E_{h_2-h_1} & E_{h_n-h_1} \\ E_{-h_2} & E_{h_1-h_2} & E_0 & E_{h_n-h_2} \\ E_{-h_n} & E_{h_1-h_n} & E_{h_2-h_n} & E_0 \end{vmatrix} \quad (2)$$

This determinant is real, must be non-negative but, in addition, de Rango (1969) and Tsoucaris (1970) have shown that the most probable set of phases for the reflexions in the determinant is that which maximizes its value. As an example, for a determinant of order eight there are 28 separate elements. Main (unpublished) has designed a computer algorithm for finding determinants such that the average magnitude of the elements is as large as possible but also giving as few as possible independent involved  $E$ 's.

As an example of the application of the method we take lithocholic acid ( $C_{24}H_{30}O_3$ ) (Arora & Bates, 1977). The space group is  $P2_12_12_1$  with  $Z=4$ . The chosen determinant is illustrated in Fig. 1. The origin

and enantiomorph are defined as follows:

$ E_h $	$h$	$k$	$l$	Phase
2.55	3	$\bar{5}$	0	$90^\circ$
1.73	1	0	5	$90^\circ$
2.54	0	2	23	$180^\circ$
1.61	0	1	6	$90^\circ$

Ten reflexions (primary set) have their phases represented by magic integers as  $\{3\ 5\ 7\ 11\ 19\}$  ( $x_1$ ) as shown in Table 1. The phases of other reflexions (secondary set) appearing in the determinant are expressed in magic-integer form through linkage with two of the above in a  $\Sigma_2$  relationship. These linkages are shown in Table 2.

Table 1. *Magic-integer representation of the ten reflexions in the primary set*

$ E_h $	$h$	$k$	$l$	Symbol	$ E_h $	$h$	$k$	$l$	Symbol
3.11	1	1	12	$3x$	2.99	$\bar{2}$	4	1	$3y$
2.38	$\bar{2}$	1	6	$5x$	2.77	2	$\bar{3}$	17	$5y$
2.28	$\bar{2}$	4	11	$7x$	2.15	0	0	$\bar{10}$	$7y$
2.16	1	$\bar{1}$	$\bar{1}$	$11x$	2.01	1	$\bar{1}$	$\bar{1}\bar{1}$	$11y$
1.65	$\bar{1}$	2	17	$19x$	1.78	1	0	6	$19y$

The value of the determinant,  $D_8(x, y)$ , is now a function of  $x$  and  $y$  and this is evaluated and output by a computer as a grid of numbers. The peaks of this map give values of ( $x, y$ ) which can be translated into phases for each of the reflexions. These phase sets can then be refined by a method due to Main. The technique is based on matrix algebra and simultaneously modifies all the phases in the determinant by maximizing the largest eigenvalue of the associated matrix. The determinant is finally recalculated and its

0,0,0 10.6	0,0, $\bar{10}$ 2.2	$\bar{2},4,\bar{1}$ 2.3	$1,\bar{1},\bar{1}$ 2.0	$\bar{1},0,\bar{5}$ 1.7	$\bar{2},4,1$ 3.0	0,1,6 1.6	1,1,12 3.1
0	7y	180-7x	11y	270	3y	90	3x
0,0,0 10.6	$\bar{2},4,\bar{1}$ 3.0	$1,\bar{1},\bar{1}$ 2.2	$\bar{1},0,\bar{5}$ 1.7	$\bar{2},4,11$ 2.3	0,1,16 0.8	1,1,22 0.4	
0	180-3y	11x	90	7x	90-7y	180+11x	
0,0,0 10.6	3,5,0 2.6	1,4,6 1.3	0,0,12 1.2	2,3,172.8	3,3,23 1.1		
0	90	7x+90	3y, 7x-180	5y	270		
0,0,0 10.6	$\bar{2},1,6$ 2.4	$\bar{3},5,12$ 1.3	$\bar{1},2,17$ 1.7	0,2,23 2.5			
0	5x	-8y	19x	180			
0,0,0 10.6	1,4,6 1.3	1,1,1 2.0	2,1,17 1.2				
0	7x+270	11y	3x+90				
0,0,0 10.6	2,3,5 1.4	3,3,11 1.4					
0	90-3y	3x-3y					
0,0,0 10.6	1,0,6 1.8						
0	19y						
0,0,0 10.6							
0							

Fig. 1. A Karle–Hauptman determinant of order eight for lithocholic acid. Each element shows the indices, value of  $|E|$  and representation of phase in magic-integer form or in degrees.

value used to give an order of merit to the phase sets under consideration.

For lithocholic acid the top seven peaks of the determinant map were considered as significant and after refinement five of the sets gave very high values for the determinant. Of these five, two corresponded to reasonably small weighted average phase errors

$$\overline{\Delta\varphi^E} = \left( \sum_h |E_h| \Delta\varphi_h \right) / \left( \sum_h |E_h| \right) < 30^\circ .$$

Experience so far with determinants has been very encouraging. They seem to have a tremendous reserve of discriminating power for correct phase sets which can cope with even the crudeness of magic-integer phase representation. Frequently it is found that the highest peak in the determinant map corresponds to a reasonable set of phases and when the *top* peak does not give good phases then they are usually given by one or more of the top few peaks in ranking order of peak magnitude.

Further improvements in 'magic determinants' are indicated by recent tests. Sequences of magic integers can now be used which are optimized in the sense that, while keeping the integers reasonably small, large numbers of phases can be represented with acceptable errors and with expected errors uniformly spread out over the represented phases (Main, unpublished). Using the power of the latest computers this makes possible the use of 'magic determinants' of order 20 or more.

There is no doubt that at this time, by the use of magic integers, the 'starting set' problem has been substantially solved. The crudest applications of magic integers are appreciably better than phase permutation and much more sophisticated and effective methods of application are under active development.

### Refining phases

The tangent formula has long served the needs of the direct-methods practitioner wishing to extend phase information or to refine a set of phases to self-consistency. We shall now consider an alternative procedure for using  $\Sigma_2$  relationships to refine phases.

A typical triple-phase relationship may be written as

$$\varphi_1 \pm \varphi_2 \pm \varphi_3 + b \approx 0 \pmod{2\pi} . \quad (3)$$

If phases are expressed in cycles then this becomes

$$\varphi_1 \pm \varphi_2 \pm \varphi_3 + b \approx 0 \pmod{1} \quad (4)$$

or

$$\varphi_1 \pm \varphi_2 \pm \varphi_3 + b \approx n \quad (5)$$

where  $n$  is some integer. The symbol  $\approx$  should be interpreted as 'tend to be close to' in these equations.

If there is a set of  $m$   $\varphi$ 's related by  $q$  relationships then these may be expressed as

$$\varphi_{1,s} \pm \varphi_{2,s} \pm \varphi_{3,s} + b_s \approx n_s \quad (s = 1 \text{ to } q) . \quad (6)$$

Normally  $q > m$  in any case of interest and for some set of correct  $\varphi$ 's the  $n$ 's will tend to cluster around some set of unknown integers.

It can be shown that  $m$  of the integers, corresponding to the right-hand sides of  $m$  linearly-independent relationships can be chosen arbitrarily (as zero for example). This comes about because phases are not defined exactly but are subjected to the '2 $\pi$  ambiguity'.

Now let us suppose that we have a set of nearest integers (N.I's) – do these enable us to get estimates for the phases? The answer is that they do and that one gets good estimates for phases from a least-squares solution of the linear equations (6) with  $\approx$  replacing  $=$ .

This may be illustrated by taking as an example some data from lithocholic acid. Reflexions are indicated by code numbers and the origin and enantiomorph were fixed by

Code	Phase
2	270
3	90
4	90
91	90 .

An additional 31 reflexions were taken and the 35 phases in the complete set were linked by 67 triple-phase relationships. From a *CONVERGENCE* map it appeared that it would be necessary to know one extra phase to determine all the others and this was

Table 2. Linkages of phases – the primary set giving phases of the secondary set

$ E_h $	$h k l$	$h k l$	$h k l$	Symbol
1.37	$\varphi(3 \bar{3} 11) \approx \varphi(1 1 12)(3x)$	$+\varphi(2 \bar{4} \bar{1})(-3y)$		$3x - 3y$
1.35	$\varphi(2 \bar{3} 5) \approx \varphi(0 1 6)(90^\circ)$	$+\varphi(2 \bar{4} \bar{1})(-3y)$		$90^\circ - 3y$
1.31	$\varphi(\bar{1} 4 6) \approx \varphi(2 4 11)(7x)$	$+\varphi(1 0 \bar{5})(270^\circ)$		$7x + 270^\circ$
1.28	$\varphi(\bar{3} 5 12) \approx \varphi(2 4 1)(3y)$	$+\varphi(\bar{1} 1 11)(-11y)$		$-8y$
1.18	$\varphi(0 0 12) \approx \varphi(2 4 1)(3y)$	$+\varphi(2 \bar{4} 11)(7x + 180^\circ)$		$3y + 7x + 180^\circ$
1.15	$\varphi(2 1 17)$	$\varphi(1 1 12)(3x)$	$+\varphi(1 0 5)(90^\circ)$	$3x + 90^\circ$
1.05	$\varphi(3 \bar{3} 23)$	$\varphi(0 2 23)(180^\circ)$	$+\varphi(3 \bar{3} 0)(90^\circ)$	$270^\circ$
0.84	$\varphi(0 1 16)$	$\varphi(0 1 6)(90^\circ)$	$+\varphi(0 0 10)(-7y)$	$90^\circ - 7y$
0.40	$\varphi(1 1 22)$	$\varphi(0 2 23)(180^\circ)$	$+\varphi(1 \bar{1} \bar{1})(11x)$	$180^\circ + 11x$

for reflexion 39. The true phase of this reflexion is  $210^\circ$  but, in accordance with the *MULTAN* philosophy, it was considered with the four possible values, 45, 135, 225 and  $315^\circ$ .

The first few phase relationships are shown below with constant angles shown in cycles.

$$\begin{array}{llll} \varphi_{39} + 0.375 \approx n_1 & (0.96) & \kappa = 5.00 & \\ \varphi_5 - \varphi_{39} - \varphi_4 & \approx n_2 & (0.11) & 3.94 \\ \varphi_6 - \varphi_3 + \varphi_5 & \approx n_3 & (0.95) & 3.64 \\ \varphi_7 + \varphi_3 - \varphi_4 & \approx n_4 & (0.92) & 3.66 \\ \varphi_7 - \varphi_{39} - \varphi_6 & \approx n_5 & (0.08) & 2.70 \\ \varphi_8 - \varphi_2 + \varphi_6 & \approx n_6 & (-0.85) & 3.42 \\ \varphi_9 - \varphi_2 - \varphi_5 & \approx n_7 & (0.01) & 3.17 \\ \varphi_9 - \varphi_3 - \varphi_8 & \approx n_8 & (1.06) & 3.04 \\ \varphi_{10} - \varphi_6 + \varphi_8 + 0.500 \approx n_9 & (2.03) & & 3.31 \\ \varphi_{10} + \varphi_5 + \varphi_9 & \approx n_{10} & (1.13) & 2.99 \end{array}$$

The first of these equations corresponds to  $\varphi_{39} \approx 225^\circ$  and the  $\kappa$  value is an arbitrary one but made to correspond to a strong relationship. Since the phases are known it is possible to calculate the actual values of the invariants and these values are shown in brackets. For the complete set of relationships they have a usual wide range of values and of the 67 relationships five deviate by more than 0.25 from integers.

Next the relationships are expressed as linear equations with weights  $\kappa$  and right-hand sides made equal to the nearest integer. Thus the second relationship appears as

$$3.94\varphi_5 - 3.94\varphi_{39} - 3.94\varphi_4 = 0 \quad (7)$$

and the ninth relationship as

$$3.31\varphi_{10} - 3.31\varphi_6 + 3.31\varphi_8 + 3.31 \times (0.5) = 3.31 \times 2 \quad (8)$$

or

$$3.31\varphi_{10} - 3.31\varphi_6 + 3.31\varphi_8 = 4.965 \quad (9)$$

The complete set of equations, in matrix algebra notation is

$$A\varphi = C \quad (10)$$

with least-squares solution

$$\varphi = (A^T A)^{-1} A^T C \quad (11)$$

The solution of the equations, with special phases taken as the nearest allowed value, is shown in Table 3.

The mean error is  $20^\circ$  and it is interesting to note that if the correct phases are fed into the tangent formula and refined to self-consistency then the mean error is  $27^\circ$ .

A series of experiments with the least-squares (l-s) equations has shown:

(i) The errors from solving the l-s equations are less than those obtained by the use of the tangent formula.

Table 3. Phases found for 31 lithocholic acid reflexions from least-squares equations with correct right-hand-side integers

Code	True phase	Calculated phase	Error
39	210	235	25
5	338	334	4
6	95	100	5
7	333	349	16
8	208	183	25
9	303	258	45
10	90	90	0
14	358	327	31
11	100	67	33
12	79	64	15
25	63	80	17
23	95	65	30
20	221	254	33
56	198	190	8
28	94	116	22
31	38	107	69
26	2	12	10
27	302	261	41
47	224	219	5
85	240	288	48
66	189	175	14
89	131	132	1
94	319	318	1
95	80	79	1
77	144	146	2
18	180	180	0
36	256	226	30
79	23	27	4
74	258	287	29
35	187	131	56
87	180	180	0

- (ii) The l-s equations are stable and a correct set of phases will not gradually drift a long way from true values. A cyclic refinement procedure with the l-s equations consists of recalculating N.I's with the last estimates of  $\varphi$  until the N.I's no longer change and this seems to stabilize quite quickly.
- (iii) The l-s equations have a remarkable ability to refine phases from values far from their true values.

### The basic MAGLIN concept

In order to establish the background for a discussion of some aspects of the practical application of the least-squares equations, including some new theoretical developments, it is desirable now to give the basic outline of the *MAGLIN* (MAGic integer - LINear equations) phase-development scheme.

The steps will be:

- (a) A determinant of order, say, 10-14 will be found containing 30-40 independent reflexions. This process, developed by Main, and to which reference has already been made, starts with a very large determinant which is reduced by the removal of rows and columns according to various

criteria until a determinant of the required order and with the required properties is reached. This process has a family resemblance to the *CONVERGE* procedure in *MULTAN*.

- (b) Magic-integer phases will be assigned to all phases in the determinant other than those which define the origin and enantiomorph. The algorithm for this will select primary and secondary set reflexions in such a way that the number of primary reflexions is kept reasonably small so that the magic integers do not get too great in magnitude.
- (c) The  $D(x,y)$  map will be calculated over a grid. An economical way of doing this has been developed whereby the first calculation is done over a fairly coarse grid which is later supplemented by further calculations on a finer grid in regions of interest.
- (d) A peak-search procedure will select a number of the highest peaks, say 10, and the phases corresponding to these will be refined by the 'maximum-greatest-eigenvalue' criterion.
- (e) About 8–10 further reflexions will be selected, which link strongly with those whose phases are already known, and represented by magic integers in one dimension. These extra reflexions will at each stage be indicated as a by-product of the *CONVERGE*-like process referred to in (a).
- (f) For each trial set of phases a  $\psi$  map will be calculated for the further reflexions by a fast-Fourier-transform program. A peak search program will find the 50 highest peaks and the phases from these will be calculated.
- (g) For each of the 500 ( $10 \times 50$ ) trial sets of phases, right-hand-side integers will be calculated for the triple-phase relationships linking *all* the phases in each set.
- (h) Each of the 500 sets of phases will be refined by the least-squares process. Experience shows that rarely is more than one cycle of refinement required.
- (i) A sorting program will eliminate all duplicate sets of phases leaving only those which are significantly different from each other. Again, it is a matter of experience that this drastically reduces the number of phase sets to be considered.
- (j) For all remaining sets of phases various figures of merit will be calculated. These will include all those presently used in *MULTAN* plus negative quartets (Hauptman, 1975*a,b*) where available. A composition of these figures of merit will be made to establish a ranking order for the phase sets.
- (k) For the 20 top-ranking sets of phases the process described above from (e) will be repeated so that 1000 sets of phases will be explored at each stage. This will continue until a number of phases have been determined sufficient to define the structure.
- (l) All the phases determined at this last stage will be contained in a determinant (order  $\sim 50$ ) found during process (a). For a large number of the most

plausible final sets of phases this determinant will be evaluated to provide a highly sensitive figure of merit.

- (m) Sets of phases will be examined by the automatic procedures now available in the *MULTAN* system.

Some practical implications of this process will now be considered.

### The efficient implementation of *MAGLIN*

All the steps described in the *MAGLIN* system are capable of realization. Although the future tense has been used in reference to many of them they are all tried and tested, albeit in a somewhat *ad hoc* manner. Further trials and development may well give minor modifications of the scheme as envisaged at present but major changes are unlikely and the process of assembling *MAGLIN* into a system as sophisticated and as convenient to use as *MULTAN* has already begun.

A characteristic of *MAGLIN* that will not have escaped the notice of the reader is the great 'depth-of-search' which is being envisaged with up to 1000 sets of phases being refined by the least-squares process at each stage. To those accustomed to *MULTAN*, where to explore more than 64 sets of phases is rare, this number of phase sets may seem extravagant in the extreme and perhaps even impossible to handle.

Returning to the least-squares solution of the equations for phases,

$$\varphi = (A^T A)^{-1} A^T C, \quad (11)$$

it will be noted that the only change in going from one phase set to the next is the column vector  $C$ . In a modern computer with fast-access backing store it is possible to evaluate, store and quickly recover the complete matrix

$$L = (A^T A)^{-1} A^T. \quad (12)$$

It thus turns out that the major single item of time consumption at each stage of the phase-refinement process is the inversion of the matrix  $A^T A$ . Thereafter, once the initial overhead is paid there is little penalty in exploring a large number of phase sets. This is in contrast to the use of the tangent formula where the cost is exactly proportional to the number of phase sets being developed. While the development of a single phase set by the least-squares method may be much more expensive than a tangent-formula development, the generation of a large number of sets is incomparably cheaper.

However it may be argued that the inversion of a large matrix is by no means a trivial operation, even on a large computer. Typically the inversion time for an order 300 matrix is about 4 min but the complete *MAGLIN* process as described may require the inversion of a large number of matrices of steadily in-

creasing order, thus 20, 30, ..., 40, 290, 300. Since the time of inversion of a matrix is proportional to the cube of its order it will be found that over 30 min of matrix-inversion time will be required.

To solve this problem a novel process of solving the equations has been developed whereby the number of matrix inversions is greatly reduced. Consider a situation where there are  $N$  relationships involving  $M$  reflexions but estimates of phase are known only for  $m(\leq M)$  reflexions. An individual phase relationship will, in general, involve three phases and be of the form

$$\varphi_p \pm \varphi_q \pm \varphi_r + b \approx 0. \quad (13)$$

We shall take a particular case

$$\varphi_3 - \varphi_{19} + \varphi_{33} + 0.5 \approx 0 \quad (14)$$

where the constant phase is expressed in cycles.

If estimated values are available for all three phases,

$$\varphi_3 \approx 0.42, \varphi_{19} \approx 0.61, \varphi_{33} \approx 0.84,$$

then we find

$$\varphi_3 - \varphi_{19} + \varphi_{33} + 0.5 \approx 1.15. \quad (15)$$

The equation for the least-squares solution

$$\text{is} \quad \varphi_3 - \varphi_{19} + \varphi_{33} + 0.5 = 1 \quad (16a)$$

$$\text{or} \quad \varphi_3 - \varphi_{19} + \varphi_{33} = 0.5. \quad (16b)$$

A comparison of (15) and (16a) and an examination of the coefficients of  $\varphi_3$ ,  $\varphi_{19}$  and  $\varphi_{33}$  leads to the conclusion that equation (16b) points towards lower values for  $\varphi_3$  and  $\varphi_{33}$  and a higher value for  $\varphi_{19}$ . The final outcome for  $\varphi_3$  of the complete least-squares solution will depend, to first order, on the aggregated effects of all the equations involving  $\varphi_3$  looked at in this way.

We now come to the treatment of a phase relationship where there is no estimate for one or more of the included phases. Let this be

$$\varphi_3 + \varphi_{21} + \varphi_{37} + 0.5 \approx 0 \quad (17)$$

with no estimates available for  $\varphi_{21}$  and  $\varphi_{37}$ . In this case  $\varphi_{21}$  and  $\varphi_{37}$  are given arbitrary values – two cycles has been used – to give

$$\varphi_3 + \varphi_{21} + \varphi_{37} + 0.5 \approx 4.92 \quad (18)$$

$$\text{or} \quad \varphi_3 + \varphi_{21} + \varphi_{37} \approx 4.42. \quad (19)$$

The exact calculated value of (19) is now used as an element in the right-hand-side column vector  $C$  in equation (11). In the least-squares solution, equation (19) will have the effect of damping down the tendency of  $\varphi_3$  to change its value but it will not introduce any systematic influence for the new value to be greater or less than that input.

The technique works extremely well. The recalculated values of the extra phases, inserted with arbitrary

values of two cycles, come out in the range 1.985–2.015. What is more significant is that the refined values of the other phases end up within a few degrees of these values obtained by solving a subset of equations which exclude phases for which there are no estimates. Table 4 gives an example of this process and the results found may be compared with those in Table 3.

Table 4. Phases found for 31 lithocholic acid reflexions in the presence of 10 reflexions for which there are no phase estimates

Code	True phase or 720°	Calculated phase	Error
39	210	234	24
5	338	333	5
6	95	99	4
7	333	349	16
8	208	185	23
9	303	259	44
10	90	90	0
14	358	326	32
11	100	65	35
12	79	64	15
25	63	80	17
23	95	66	29
20	221	251	30
56	198	191	7
28	94	115	21
31	38	103	65
26	2	8	6
27	302	264	38
47	224	221	3
85	240	287	47
66	189	175	14
89	131	135	4
94	319	318	1
95	80	80	0
77	144	142	2
18	180	180	0
36	256	226	30
79	23	28	5
74	258	285	27
35	187	137	50
87	180	180	0
83	720	715	
21	720	719	
19	720	716	
81	720	721	
88	720	725	
67	720	722	
32	720	719	
29	720	718	
73	720	718	
49	720	723	

This idea is very important for the practical implementation of *MAGLIN*. Now one needs only to invert a few matrices and typically these may be of order 25, 40, 60, 90, 135, 200, 300; the total inversion time is just 43% greater than that for inverting the largest matrix alone.

It turns out in practice that it is unnecessary even to invert a large-order matrix at all. As the process of

phase development proceeds certain phases stabilize to the extent that they change only by fractions of a degree in going from one stage to the next. The stable phases are completely predictable as they can be recognized from the magnitudes of the elements of the matrix  $L$  (equation 12). With such phases kept constant the order of the matrix to be inverted is much reduced. It should be noted that the stability criterion does not depend on the particular set of phases being refined but only on the elements of the matrix, which are common to all sets. It is confidently expected that the inversion of a matrix of order greater than 150 or so will only rarely be required and that matrix-inversion time may not be a significant limiting factor in the application of the *MAGLIN* method.

### Conclusions

Although *MAGLIN* is being presented as a successor to *MULTAN* it is by no means certain that it will be a replacement – at least for some time. A survey of literature suggests that some 30% or more of structures are being solved by *MULTAN* at present and such a well-established method will not be lightly set aside until the evidence is overwhelming that the replacement is really better.

It will take some time for *MAGLIN* to evolve to the same state of development and reliability as *MULTAN* even though it will call on many existing *MULTAN* components such as those for normalizing data, fixing the origin and enantiomorph and finding  $\Sigma_2$  relationships. However, on the basis of the present evidence one may make some predictions about *MAGLIN* capability.

By the use of magic determinants it should be possible to find starting sets with at least 30 reflexions and with mean errors less than about  $40^\circ$ . Experience with *MULTAN* suggests that this size of starting set will be able to cope comfortably with structures with 100 or so atoms in the asymmetric unit. This is about twice the size of structure which can be handled by *MULTAN* in a fairly routine way. On the other hand *MULTAN* has solved structures with about 100 atoms in the asymmetric unit in favourable circumstances.

On the basis of this comparison, and a feeling for

the power of the least-squares refinement procedure, it is predicted that *MAGLIN* should be able routinely to solve structures with 100 atoms in the asymmetric unit and occasionally to solve structures of twice this complexity.

Trials of the various elements which have come together to form *MAGLIN* have been going on for about two years and have involved the collaboration of the author with a number of other researchers. The description of the individual parts, with more detail and theoretical background than could be given here, will be presented in separate papers in the near future. However it seemed sensible to first present a general overview of *MAGLIN* before publication of detailed accounts of its various components. It is hoped thereby that the detailed work will be better understood and appreciated in the context of what is presented here. In the meantime the author wishes to acknowledge the contribution of the following who are contributing to the *MAGLIN* project: Dr J. P. Declercq, Dr G. Germain, Dr P. Main, Dr P. Mondal and Mr. D. Taylor.

All the separate parts of this work and also the overall *MAGLIN* development programme have been generously supported by the Science Research Council. Thanks are also due to the Computer Laboratories of the University of York and the University of Leeds and to the Centre de Calcul of the University of Louvain-la-Neuve, Belgium.

### References

- ARORA, S. K. & BATES, R. B. (1977). In preparation.
- DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* A **31**, 367–372.
- HAUPTMAN, H. (1975a). *Acta Cryst.* A **31**, 671–679.
- HAUPTMAN, H. (1975b). *Acta Cryst.* A **31**, 680–687.
- KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181–187.
- LESSINGER, L. (1976). *Acta Cryst.* A **32**, 538–550.
- RANGO, C. DE (1969). Thesis, Paris.
- TAYLOR, D. J. & WOOLFSON, M. M. (1975). *Acta Cryst.* A **31**, S 16.
- TSOUCARIS, G. (1970). *Acta Cryst.* A **26**, 492–499.
- WHITE, P. & WOOLFSON, M. M. (1975). *Acta Cryst.* A **31**, 53–56.