

- IUPAC (1990). *Nomenclature of Inorganic Chemistry*. Oxford: Blackwell.
- JONES, P. M., RACKHAM, G. M. & STEEDS, J. W. (1977). *Proc. R. Soc. London Ser. A*, **354**, 197-222.
- KAINUMA, Y. (1955). *Acta Cryst.* **8**, 247-257.
- KELLY, P. M., JOSTSONS, A., BLAKE, R. G. & NAPIER, J. G. (1975). *Phys. Status Solidi A*, **31**, 771-780.
- KIKUCHI, S. (1928). *Proc. Jpn Acad. Sci.* **4**, 271-278.
- KIRKLAND, E. J. (1990). *Ultramicroscopy*, **32**, 349-364.
- KIRKLAND, E. J., LOANE, R. F. & SILCOX, J. (1987). *Ultramicroscopy*, **23**, 77-96.
- KIRKLAND, E. J., LOANE, R. F., XU, P. & SILCOX, J. (1989). *Computer Simulation of Electron Microscopy Diffraction and Images*, edited by W. KRAKOW & M. O'KEEFE, pp. 13-31. New York: Minerals, Metals and Materials Society.
- KOMATSU, K. & TERAMOTO, K. (1966). *J. Phys. Soc. Jpn*, **6**, 1152-1159.
- LANGMORE, J. P., WALL, J. & ISAACSON, M. S. (1973). *Optik (Stuttgart)*, **38**, 335-350.
- LOANE, R. F., KIRKLAND, E. J. & SILCOX, J. (1988). *Acta Cryst.* **A44**, 912-927.
- MARADUDIN, A. A., MONTROLL, E. W., WEISS, G. H. & IPATOVA, I. P. (1971). *Theory of Lattice Dynamics in the Harmonic Approximation*, pp. 304-309. New York: Academic Press.
- MITRA, S. S. & MASSA, N. E. (1982). *Handbook on Semiconductors*, Vol. 1, edited by T. S. MOS & W. PAUL, pp. 81-192. Amsterdam: North Holland.
- MOTT, N. F. (1930). *Proc. R. Soc. London. Ser. A*, **127**, 658-665.
- PENNYCOOK, S. J. (1989). *Ultramicroscopy*, **30**, 58-69.
- PENNYCOOK, S. J. & JESSON, D. E. (1990). *Phys. Rev. Lett.* **64**, 938-941.
- PENNYCOOK, S. J., JESSON, D. E. & CHISHOLM, M. F. (1990). *Proc. 6th Conf. on the Microscopy of Semiconducting Materials*, pp. 51-58. Oxford Univ. Press.
- PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes*, pp. 191-225. Cambridge Univ. Press.
- REIF, F. (1965). *Fundamentals of Statistical and Thermal Physics*, pp. 331-346. New York: McGraw-Hill.
- REIMER, L. (1984a). *Transmission Electron Microscopy*, pp. 160-166. Berlin: Springer Verlag.
- REIMER, L. (1984b). *Transmission Electron Microscopy*, pp. 433-439. Berlin: Springer Verlag.
- REIMER, L., FROMM, I. & NAUNDORF, I. (1990). *Ultramicroscopy*, **32**, 80-91.
- SAKURAI, J. J. (1985). *Modern Quantum Mechanics*, edited by S. F. TUAN, pp. 89-94, 453. Menlo Park, California: Benjamin/Cummings.
- SARIKAYA, M. & REZ, P. (1982). *40th Annual Proc. Electron Microscopy Society of America*, edited by G. W. BAILEY, pp. 486-487. Baton Rouge: Claitors.
- SHIN, D. H., KIRKLAND, E. J. & SILCOX, J. (1989). *Appl. Phys. Lett.* **55**, 2456-2458.
- SINHA, S. K. (1973). *Crit. Rev. Solid State Sci.* **3**, 273-334.
- SPENCE, J. C. H., ZUO, J. M. & LYNCH, J. (1989). *Ultramicroscopy*, **31**, 233-239.
- STEEDS, J. W. (1983). *Quantitative Electron Microscopy*, edited by J. N. CHAPMAN & A. J. CRAVEN, pp. 49-96. Glasgow: Scottish Univ. Summer School in Physics.
- TAKAGI, S. (1958). *J. Phys. Soc. Jpn*, **13**, 278-296.
- TANAKA, M., SAITO, R. & SEKII, H. (1983). *Acta Cryst.* **A39**, 357-368.
- VINCENT, R., BIRD, D. M. & STEEDS, J. W. (1984). *Philos. Mag.* **A50**, 745-786.
- WANG, Z. L. & COWLEY, J. M. (1989). *Ultramicroscopy*, **31**, 437-453.
- WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal Vibrations in Crystallography*, pp. 34-78. Cambridge Univ. Press.
- XU, P., KIRKLAND, E. J., SILCOX, J. & KEYSE, R. (1990). *Ultramicroscopy*, **32**, 93-102.
- ZUO, J. M., SPENCE, J. C. H. & O'KEEFE, M. (1988). *Phys. Rev. Lett.* **61**, 353-356.
- ZUO, J. M., SPENCE, J. C. H. & HØIER, R. (1989). *Phys. Rev. Lett.* **62**, 547-550.

Acta Cryst. (1991). **A47**, 278-281

On the Application of Phase Relationships to Complex Structures. XXIX. Choosing the Large E_s

BY G. GERMAIN AND V. NASTOPOULOS*

*Unité de Chimie Physique Moléculaire et de Cristallographie, Université de Louvain,
1348 Louvain-la-Neuve, Belgium*

AND M. M. WOOLFSON

Department of Physics, University of York, York YO1 5DD, England

(Received 8 August 1987; accepted 16 January 1991)

Abstract

The set of large E_s through which a structure is solved by direct methods is usually chosen by a convergence or convergence-divergence process. This process aims to give a strong phase-extension pathway starting from a small set of E_s whose phases are known or allocated in some way. Sometimes sets of reflexions

thus obtained are poorly conditioned and under tangent-formula refinement even initially correct phases will degenerate to randomness. A simple new algorithm has been developed which improves the conditioning of the complete set of reflexions and their relationships and is more appropriate to current trends to start refinement from a complete set of random phases. A particular feature of this algorithm is that it maximizes the minimum number of relationships for any reflexion.

* Now at University of Patras, Patras, Greece.

Introduction

In earlier forms of direct methods, such as the symbolic addition method introduced by Karle & Karle (1963) and the first version of *MULTAN* (Germain, Main & Woolfson, 1970), phase sets were generated by a bootstrap operation beginning with a fairly small starting set of reflexions. The starting-set phases usually contained some which could be allocated specifically to fix the origin and enantiomorph and others were given either symbols or permuted numerical phases.

In order to ensure a strong development of phases using the tangent formula, Germain *et al.* (1970) designed an algorithm called *CONVERGENCE* which aimed to give a phase-extension pathway without weak links, *i.e.* steps where a new phase was indicated with a very large variance. The *CONVERGENCE* process is fairly efficient but by no means perfect. Lessinger (1976) investigated a number of *MULTAN* failures (structures which were eventually solved in some other way) and found that correct phases were not stable under tangent-formula refinement. If correct phases degenerate to random values under the given refinement process then they are hardly likely to be found in the first place. Other workers extended and improved on the *CONVERGENCE* idea by adding to it a divergence procedure and one can find examples of these in the *SIMPEL* package (Schenk, 1973) and *SHELX* (Sheldrick, 1975).

Over the past decade or so, a new philosophy of phase determination has been introduced in which the need to start from a small base has been removed. Steps in this direction were the introduction of *YZARC* (Baggio, Woolfson, Declercq & Germain, 1978), *MAGEX* (Hull, Viterbo, Woolfson & Zhang, 1981) and culminated in *RANTAN* (Yao, 1981). In *RANTAN* all reflexions are allocated initial phases, mostly random, and all relationships are brought into play, using the tangent formula, from the very beginning of the procedure. The same feature occurs in the more recently introduced Sayre-equation tangent formula method, *SAYTAN* (Debaerdemaeker, Tate & Woolfson, 1985, 1988).

With these new approaches the whole *raison d'être* of the *CONVERGENCE* procedure, creating a reliable phase-extension pathway, is invalidated. For this reason we have considered alternatives for selecting the set of reflexions in terms of which the structure is to be solved.

The aim

We go back to the work of Lessinger (1976) who found that the reason for a set of correct phases deteriorating under tangent formula refinement was that a few reflexions were poorly linked, *via* suitable phase relationships, with the remainder. Even correct

phases gave relatively large errors for the phase estimates for these reflexions when substituted in the tangent formula. These incorrect phases fed back to give errors in other phases in such a way that for each iteration the phase errors became larger until eventually the initially correct phases were randomized. To counteract this process as much as possible the *CONVERGENCE* process was modified in *MULTAN*; 50 extra reflexions were included in the system to generate \sum_2 relationships and then in the *CONVERGENCE* process the 50 reflexions which were the first to be eliminated, which were probably (but not certainly) the least well connected to the system as a whole, were then removed.

As previously stated, for methods like *RANTAN* which used all relationships immediately, the *CONVERGENCE* criterion of having a strong phase-extension pathway was not relevant. What was relevant was that no phases should be poorly linked to the whole system because correct phases might then be unstable under the refinement process. The requirement for avoiding this situation is that even the least well linked reflexion should give a phase estimate with the smallest possible variance. This is equivalent to having as large a value as possible of

$$\alpha(\mathbf{h}) = \left| \sum_{\mathbf{k}} \kappa(\mathbf{h}, \mathbf{k}) \exp \{ \varphi(\mathbf{h}) - \varphi(\mathbf{k}) - \varphi(\mathbf{h} - \mathbf{k}) \} \right| \quad (1)$$

where

$$\kappa(\mathbf{h}, \mathbf{k}) = 2(\sigma_3/\sigma_2^{3/2}) |E(\mathbf{h})E(\mathbf{k})E(\mathbf{h} - \mathbf{k})|$$

and $\sigma_n = \sum_{i=1}^N z_i^n$. There are N atoms in the unit cell and the atomic number of the i th atom is z_i . While a knowledge of phases is required to calculate $\alpha(\mathbf{h})$, it can be estimated from the magnitudes of normalized structure factors with

$$\alpha_{\text{est}}(\mathbf{h}) = \left\{ \sum_{\mathbf{k}} \kappa(\mathbf{h}, \mathbf{k})^2 + \sum_{\substack{\mathbf{k} \neq \mathbf{k}' \\ \mathbf{k} \neq \mathbf{k}'}} \kappa(\mathbf{h}, \mathbf{k}) \kappa(\mathbf{h}, \mathbf{k}') \right. \\ \left. \times \frac{I_1\{\kappa(\mathbf{h}, \mathbf{k})\}}{I_0\{\kappa(\mathbf{h}, \mathbf{k})\}} \frac{I_1\{\kappa(\mathbf{h}, \mathbf{k}')\}}{I_0\{\kappa(\mathbf{h}, \mathbf{k}')\}} \right\}^{1/2} \quad (2)$$

where $I_0(x)$ and $I_1(x)$ are modified Bessel functions.

For a complete set of reflexions and a given number of relationships it is better to have a uniform number of relationships per reflexion than to have large differences so that some reflexions have a large value of α while others have much smaller values. The reason for this can be seen in the way that the variance of a phase estimate depends on the value of α , which was given by Karle & Karle (1966) and is shown in Fig. 1. If a reflexion has a value of α given by the point P in the figure then the change in α produced by having one more or one less relationship will change the variance very little. On the other hand, the change of α produced by the addition of one relationship for a reflexion with α corresponding to point Q may significantly reduce the variance.

While the values of α_{est} are the actual quantities of interest in deciding on the reliability or otherwise or phase indications we have found in practice that simply using the number of relationships for a reflexion is just as effective and makes the algorithms we propose much simpler to implement.

On the basis of the above considerations we conclude that for a given total number of large-magnitude reflexions they should be chosen:

- (1) to have as many relationships as possible;
- (2) to have as little variation in the number of relationships per reflexion as possible;
- (3) to have the minimum number of relationships for a reflexion as large as possible, consistent with (1) and (2).

Thus, in an ideal case, if there were 2000 different phase relationships (each involving three phases) linking 200 reflexions then one would like to have exactly 30 relationships for each reflexion.

Methods of choosing reflexions

To achieve the conditions set out in the previous section the set of reflexions was chosen to maximize some function and four different functions were tried:

$$Q_1 = \exp(\bar{n}) \sum_{\mathbf{h}} |E(\mathbf{h})| \exp[-\gamma|1 - n(\mathbf{h})/\bar{n}|] \quad (3)$$

$$Q_2 = \bar{n}^\beta \sum_{\mathbf{h}} |E(\mathbf{h})| \{1 - \exp[-n(\mathbf{h})/\bar{n}]\} \quad (4)$$

$$Q_3 = \sum_{\mathbf{h}} \log \{|E(\mathbf{h})|n(\mathbf{h})\} \quad (5)$$

$$Q_4 = n_T, \quad (6)$$

where \bar{n} = average number of relationships per reflexion, $n(\mathbf{h})$ = number of relationships for the reflexion of index \mathbf{h} , n_T = total number of relationships and β and γ are two numerical parameters.

Two different methods (*M1* and *M2* below) were developed to maximize the selected function for a predetermined number of reflexions. In each method

the starting point is an initial number of reflexions N_I and the relationships linking them and the aim is to finish with N_F reflexions having the largest value of Q_i .

When we have our final N_F reflexions we judge the effectiveness of the combination of method and optimized function by the three quantities n_T , previously given, n_{min} , the minimum number of relationships for any reflexion, and n_{max} , the maximum number of relationships for any reflexion. From the point of view of the conditioning of the final set of reflexions, undoubtedly the most important of the three quantities is n_{min} , which should be as large as possible.

Method M1

(i) We start with N_I reflexions and eliminate them one by one by the *CONVERGENCE* procedure until 40 remain.

(ii) Reflexions are added one by one to the starting set of 40 under the condition that the reflexion added should give the largest value of Q_i . This is continued until there are $40 + N_F$ reflexions.

(iii) Reflexions are eliminated one by one under the condition that the reflexion removed should leave the largest value of Q_i . This is continued until N_F reflexions remain.

In this final step, reflexions can be eliminated two or three at a time without making much difference to the process.

Method M2

(i) We start with N_I reflexions and eliminate them one by one with the *CONVERGENCE* process until 40 remain.

(ii) Reflexions are added one by one to the starting set of 40 under the condition that the one added should give the largest value of Q_i . This is continued until there are N_F reflexions.

(iii) The reflexions are now considered in two groups - group *A* containing the N_F reflexions found in (ii) and group *B* containing the residual $N_I - N_F$ reflexions. Phase relationships are found linking three *A*-group reflexions (*AAA*) and one *B* and two *A* reflexions (*BAA*).

(iv) The reflexion with the minimum number of relationships, $n_{min}(A)$, from group *A* is found and that with the maximum number, $n_{max}(B)$, from group *B*. If $n_{min}(A) < n_{max}(B)$ then the two reflexions are interchanged in groups.

Steps (iii) and (iv) are repeated until $n_{min}(A) \geq n_{max}(B)$.

There is a third method (*M3*) which does not use the Q_i functions but merely seeks to maximize the value of n_{min} .

(i) We start with N_I reflexions and reduce them by convergence until N_F remain.

Steps (ii) and (iii) are then steps (iii) and (iv) of *M2*.

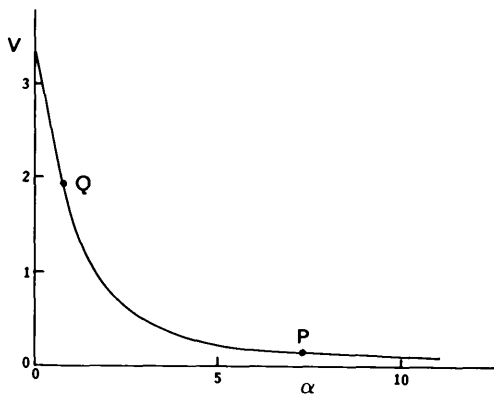


Fig. 1. The variance, V , of a phase estimate given by the tangent formula as a function of α .

Table 1. A comparison of methods M1, M2 and M3 with CONVERGENCE for a selection of structures

The three columns under each structure are n_T , n_{\min} and n_{\max} . For all cases $N_F = 200$.

		Cortisone			Roma1			Roma2			Roma4			Nancy1			
M1	Q ₁	$\gamma = 1$	3797	33	135	2332	22	70	1488	9	51	1490	13	56	1221	9	56
		$\gamma = 2$	3797	33	137	2335	23	75	1510	9	49	1489	13	55			
		$\gamma = 3$	3797	33	137	2334	23	75	1510	9	49	1490	13	53			
		$\gamma = 4$	3786	33	134	2334	23	75	1508	9	48	1492	13	54			
Q ₂	$\beta = 1$	3778	28	134	2314	20	75	1491	9	48	1484	11	54	1205	8	56	
	$\beta = 2$	3786	26	133	2334	22	70	1497	9	49	1491	13	54				
	$\beta = 3$	3791	33	133	2329	22	74	1497	9	49	1492	13	55				
Q ₃		3771	31	131	2304	20	73	1473	10	49	1481	12	56	1214	9	54	
		3792	33	134	2335	23	71	1485	10	48	1487	13	52	1218	10	53	
M ₂	Q ₁	$\gamma = 1$	3802	33	138						1492	13	56				
		$\beta = 1$	3802	33	138						1494	13	57				
	Q ₃		3798	32	135						1483	13	54				
			3796	32	137						1492	13	54				
M ₃		3798	33	135	2328	22	76	1512	9	48	1476	13	57	1262	11	50	
	CONVERGENCE	3554	16	128	2174	16	73	1428	8	47	1413	7	59	1191	8	48	

Results

Table 1 shows the results obtained for a selection of structures, the details of which are shown in Table 2. The first thing to notice is that all three methods M1, M2 and M3 are superior to CONVERGENCE and sometimes markedly so in the critical number n_{\min} . It should also be mentioned that M1 and M2 are very expensive in computer time; this explains the gaps in Table 1 and also the fact that the methods were based on the numbers of relationships rather than trying to use values of α , which would have made the process even longer. Whilst methods M1 and M2 were being examined - with the idea that if they were effective then some method of speeding them up would be found - the much simpler M3 process occurred to us. This is very quick and easy (1 to 2 min on a VAX 8650) and is as effective as the other methods, or nearly so.

Comparisons have been made on the conditioning of sets of reflexions found by M3 and CONVERGENCE by refining correct phases to self-consistency by the tangent formula. In all cases the M3-derived sets are better than those from CONVERGENCE, sometimes very much better. As an example, for cortisone the correct phases for 150 reflexions from CONVERGENCE refined to a r.m.s. error of 32.0° while correct phases for 150 reflexions from M3 refined to a r.m.s. error of 27.1°.

Concluding remarks

It must be stressed that the rationale for the approach described here is to optimize the characteristics of a whole set of reflexions with all corresponding relationships. In early direct methods the critical stage was considered to be the early phase-extension process; emphasis was directed towards obtaining new phase estimates as close to being correct as possible from the starting set of phases which was the closest to being correct. This was the aim of the CONVERGENCE process and the various con-

Table 2. Details of the structures used in the tests

	Formula	Space group	Z	Number of independent non-hydrogen atoms
Cortisone	C ₂₁ H ₂₀ O ₅	P2 ₁ 2 ₁ 2 ₁	4	26
Roma1	C ₄₀ H ₄₇ Br ₂ N ₈ O ₁₀ S ₂	P4 ₁ 2 ₁ 2	8	62
Roma2	C ₂₁ H ₃₄ O ₅	C2	8	52
Roma4	C ₄₂ H ₅₀ ClNO ₉	P2 ₁ 2 ₁ 2 ₁	4	53
Nancy1	C ₃₇ H ₄₈ NO ₁₂ PF ₆	P2 ₁	2	57

vergence-divergence procedures of other workers. The realization that taking random starting phases, even for all the reflexions in the system, is an effective way to solve structures has pointed to phase refinement rather than phase extension as the critical process. This points to a new criterion for selecting the set of large E_s - that correct phases must not grossly degenerate under the refinement process - which is the topic of this paper.

Our gratitude is due to the Science and Engineering Research Council and to the European Economic Community for their support of this work.

References

- BAGGIO, R., WOOLFSON, M. M., DECLERCQ, J.-P. & GERMAIN, G. (1978). *Acta Cryst.* **A34**, 883-892.
- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286-290.
- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353-357.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274-285.
- HULL, S. E., VITERBO, D., WOOLFSON, M. M. & ZHANG, S.-H. (1981). *Acta Cryst.* **A37**, 566-572.
- KARLE, J. & KARLE, I. L. (1963). *Acta Cryst.* **16**, 969-975.
- KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849-859.
- LESSINGER, L. (1976). *Acta Cryst.* **A32**, 538-550.
- SCHENK, H. (1973). *Acta Cryst.* **A29**, 77-82.
- SHELDRICK, G. M. (1975). *SHELX*. Program for crystal structure determination. Univ. of Cambridge, England.
- YAO, J.-X. (1981). *Acta Cryst.* **A37**, 642-644.