Theoretical values of $(P_z^{\nu}/P_x^{\nu})_c$ for K=1, 3 and 5 are 1.00, 0.76 and 0.41. The average experimental values of $(P_z^{\nu}/P_x^{\nu})_{exp}$ for K=1 and 3 are 1.03 (for $i \ge 0.75$ mA) and 0.61 (for $1 \ge 0.25$ mA).

For K=5 the comparison of the calculated value with the average experimental one is not possible, because the experimental value of $(P_z^{\nu}/P_x^{\nu})_{exp}$ depends on the high-frequency exciting current i, which dependence was not observed for K=1 and 3. Thus in the case of K=5 it is only possible to compare individual experimental quantities at low values of *i*. The increase of the ratio $(P_z^{\nu}/P_x^{\nu})_{exp}$ (for K=5) versus i is shown in Fig. 6. The increase is brought about by relatively high acceleration of the moving planes, in which case the assumption that each of n(t) 'crystalline lavers' diffracts totally is no longer justified. It is seen from Fig. 6 that for *i* ranging from 0 to 2 mA the change of $(P_z^{\nu}/P_x^{\nu})_{exp}$ is $\simeq 20\%$. The authors realize that the theory presented in this paper is applicable only to qualitative estimation of neutron diffraction by vibrating single crystals and to explanation of some phenomena observed.

Similarly, in case of K=3 and K=5 it is possible to introduce a factor y'_k analogous to the extinction factor y_{ext} (Zachariasen, 1967) and to express the integrated intensity P'_k in the form

$$P'_{k} = P^{v}_{kin} y'_{k} \tag{19}$$

where

$$y'_{k} \simeq \frac{\sin \frac{K\omega T_{z} \operatorname{tg} \theta_{B}}{2|V_{ny}|}}{\frac{K\omega T_{z} \operatorname{tg} \theta_{B}}{2|V_{ny}|}}.$$
(20)

On the basis of our experimental data and from our approximate theory we can make the following statement: For $A(\mu) = 1$, the integrated intensity difference between Bragg and Laue diffraction disappears when the displacement of the diffracting planes is accelerated in the direction of the reciprocal-lattice vector.

The authors wish to thank Miss B. Hašková, Mr J. Vávra and Mr P. Zeman for their valuable help throughout the measurements.

References

- ANTONINI, M., CORCHIA, M., NICOTERA, E. & RUSTICHELLI, F. (1972). Nucl. Instrum. Meth. 104, 147–152.
- BURAS, B. & GIEBULTOWICZ, T. (1972). Acta Cryst. A28, 151-153.
- BURAS, B., GIEBULTOWICZ, T., MINOR, W. & RAJCA, A. (1972). Phys. Stat. Sol. (a), 9, 423–433.
- CHALUPA, B., MICHALEC, R., PETRŽÍLKA, V., TICHÝ, J. & ZELENKA, J. (1968). *Phys. Stat. Sol.* **29**, K51-K54.
- KLEIN, A. G., PRAGER, P., WAGENFELD, H., ELLIS, P. E. & SABINE, T. M. (1967). Appl. Phys. Lett. 10, 293–295.
- MICHALEC, R., CHALUPA, B., SEDLÁKOVÁ, L., ČECH, J., MIKULA, P., PETRŽÍLKA, V. & ZELENKA, J. (1974). J. Appl. Cryst. In preparation.
- MICHALEC, R., SEDLÁKOVÁ, L., ČECH, J. & PETRŽÍLKA, V. (1971). Phys. Lett. 37 A, 403–405.
- MICHALEC, R., VAVŘÍN, J., CHALUPA, B. & VÁVRA, J. (1967). Report ÚJV 1562, Nuclear Research Institute, Prague-Řež.
- MIKULA, P., MICHALEC, R., SEDLÁKOVÁ, L., ČECH, J., CHALUPA, B. & PETRŽÍLKA, V. (1973). *Phys. Stat. Sol.* (a), 17, 163–168.
- MOYER, M. W. & PARKINSON, T. F. (1967). Nucl. Instrum. Meth. 53, 299–304.
- ZACHARIASEN, W. M. (1967). Acta Cryst. 23, 558-564.

Acta Cryst. (1974). A30, 564

Crystal Structure Determination by Simultaneous use of Cosine Invariant Computation and the Multisolution Method

BY BERNARD BUSETTA AND GÉRARD COMBERTON

Laboratoire de Cristallographie et de Physique Cristalline, Université de Bordeaux I, 351 Cours de la Libération, 33405 Talence, France

(Received 16 January 1974; accepted 18 March 1974)

An algorithm is given for the rapid computation of the cosine invariants, $\cos(\varphi_{-H1} + \varphi_{H2} + \varphi_{H1-H2})$, and the results are compared with the actual values for three structures. A weighting scheme is derived which enables this information to be incorporated directly into the multisolution tangent method of phase determination. Details are given of the determination of four unknown structures by this method.

Introduction

Three different direct methods have been proposed for solving crystal structures.

(1) Methods based on the zero value of the mean sine invariant are all derived from use of the \sum_{2}

formula (Karle & Hauptman, 1953) and the tangent formula (Karle & Hauptman, 1958):

$$\operatorname{tg} \varphi_{H_{1}} = \frac{\sum_{H_{2}}^{H_{2}} A_{H_{1}, H_{2}} \sin (\varphi_{H_{2}} + \varphi_{H_{1} - H_{2}})}{\sum_{H_{2}}^{H_{2}} A_{H_{1}, H_{2}} \cos (\varphi_{H_{2}} + \varphi_{H_{1} - H_{2}})},$$

where

$$A_{H_{1},H_{2}} = 2 \frac{\sigma_{3}}{\sigma_{2}^{3/2}} |E_{H_{1}}E_{H_{2}}E_{H_{1}-H_{2}}|$$

is the weight associated with the relation

$$\sin(\varphi_{-H_1} + \varphi_{H_2} + \varphi_{H_1 - H_2}) = 0$$

The simplicity of the use of these two formulae explains why they remain so popular. The multisolution method (Germain, Main & Woolfson, 1970) permits a nearly complete automation of the process of structure determination (Declerq, Germain, Main & Woolfson, 1973) as is shown by the numerous successes of MULTAN.

(2) Methods based on the cosine invariant computation (Hauptman, Fisher, Hancock & Norton, 1969; Messager & Tsoucaris, 1972) aim at using the exact value of the invariant phase rather than equating it implicitly to zero as is done in the first methods. They also have the great advantage of using all the structure factors in the phase generation, and not only the largest. However, such a computation, even with the use of the approximate triple product formula (Hauptman et al., 1969) is very time-consuming and cannot be achieved on small computers. Thus, its use has been limited to the determination of a small number of phases (up to one hundred) which are afterwards used in the normal tangent formula refinement. This explains the difficulties of a complete automation of this process of structure determination and few structures have been elucidated in this way.

(3) Methods based on the Karle-Hauptman determinant (Tsoucaris, 1970) are undoubtedly the most accurate for phase determination. However they need longer computing times than the two preceding methods. Moreover it seems difficult to reconcile them with the multisolution as well as the cosine invariant computation.

It would be interesting to combine the accuracy of the cosine invariant computation with the simplicity



Fig. 1. Variation of x_1 as a function of f (see text).

which characterizes the multisolution method. For that, we must be able to use information provided by the cosine invariant computation during the whole process of phase determination and not merely use it to generate starting phases. As it seems difficult to use the least-squares refinement proposed by Hauptman *et al.* (1969) for the whole process, we must find a means of bringing the computed cosine invariants directly into the phase refinement by the tangent formula. However the improvement resulting from the combination of these two techniques only becomes interesting if it does not involve too large an increase in computing time. This is why we sought, first, the most suitable way of computing cosine invariants on relatively small computers.

Fast computation of cosine invariants by the triple product formula (Hauptman *et al.*, 1969)

The computation of the triple product formula:

$$|E_{H_1}E_{H_2}E_{H_3}| \cos (\varphi_{H_1} + \varphi_{H_2} + \varphi_{H_3}) = \langle \varepsilon_K \varepsilon_{K+H_1} \varepsilon_{K-H_3} \rangle_K + R_3$$

where $H_1 + H_2 + H_3 = 0$,
 $\varepsilon_K = |\sqrt{|E_K|} - \langle |\sqrt{|E|} \rangle_k$,
 R_3 depends only on $|E_{H_1}|$, $|E_{H_2}|$, $|E_{H_3}|$

and K is an empirical scale factor, requires a search of a great number of sets of three reflexions K, $K+H_1$, $K-H_3$. It is obvious that such a search by computing and comparing the three indices h, k, l, for each reflexion is time-consuming. It is better to associate each reflexion with a position in the central store of the computer, as used in a search for the \sum_2 relations of Karle and Hauptman (see, for example, Beurskens, 1963).

In our case, each reflexion H(h,k,l) is associated with the address:

$$A_{H} = h + k(2h_{Max} + 1) + l(2k_{Max} + 1) (2h_{Max} + 1)$$

for the core position containing the ε_H value which then can be called directly during the computation.

When we search the \sum_2 relations for an *H* reflexion (that is to say the pairs of reflexions *K* and *K*-*H*), we may reduce the computing time by half by using the symmetry of this operation with respect to the *H*/2 point.*

^{*} In fact, the notation H/2 corresponds to only half the reciprocal-space vector when the three indices of H are even. However when this is not the case, we may nevertheless define a reflexion H' which will play in the core the same role as H/2 in the computation above. If l is odd, H' is $-h_{max}, -k_{max}, (l+1)/2$ because, to all the reflexions of the (l+1)/2 plane, the vectorial operation (-H) makes a reflexion of the plane -(l-1)/2 correspond. We obtain the same pairs of reflexions as with the planes (l-1)/2 and -(l+1)/2 after using Friedel's law. The same considerations show that: if l is even and k odd, H' is $-h_{max}, (k+1)/2, (l)/2$ and if l and k are even and h odd, H' is (h+1)/2, (k)/2, (l)/2.

Indeed, the symmetric of K with respect to H/2 is:

$$H/2 - (K - H/2) = H - K$$

and the pairs K, K-H and H-K, -K built on these two symmetrical points are identical from Friedel's law. Then we have:

$$\sum_{K} E_{K}E_{K-H} = 2 \sum_{K>H/2} E_{K}E_{K-H}.$$

all space

In the same way, we may decrease the number of operations needed in the triple product evaluation by using the symmetry of this formula with respect to the point $H_3/2$. The triple product built from the point H_3-K , symmetric in K with respect to $H_3/2$ may be written:

$$\mathcal{E}_{H_3-K}\mathcal{E}_{H_1+H_3-K}\mathcal{E}_K$$

or by using Friedel's law and the equality $H_1 + H_2 + H_3 = 0$

$$\varepsilon_K \varepsilon_{K+H2} \varepsilon_{K-H3}$$
.

Finally, the triple product formula may be computed for half of reciprocal space:

$$\langle \varepsilon_{\mathbf{K}}\varepsilon_{\mathbf{K}-H_3}\varepsilon_{\mathbf{K}+H_1}\rangle_{\mathbf{K}} = \langle \varepsilon_{\mathbf{K}}\varepsilon_{\mathbf{K}-H_3}(\varepsilon_{\mathbf{K}+H_1}+\varepsilon_{\mathbf{K}+H_2})\rangle_{\mathbf{K}>H_3/2}$$

which decreases the computing time by about one third. During the computation we obtain directly the operations on the values whose core addresses are A_{K} , A_{K-H_3} , A_{K+H_1} , A_{K+H_2} . A program written with this scheme for a 360-44

A program written with this scheme for a 360-44 IBM (128 Kbytes) computer allows the determination

of 2000 cosine invariants with 2500 to 3000 K contributors in each in about half an hour.

Before we describe how to incorporate the results of the cosine invariants computation in the normal direct method calculations, we shall compare, for known structures, the computed cosine invariants with their real values, to justify our approach.

Experimental study

We have tested three different structures solved by normal direct methods, oestrone I $(P2_12_12_1, Z=4)$ and oestrone III $(P2_1, Z=4)$ (Busetta, Courseille & Hospital, 1973) and hydrated androstanolone $(P2_1, Z=2)$ (Busetta, Courseille, Fornies-Marquina & Hospital, 1972). The results presented here relate to the first structure, but similar results were obtained for the other two.

The different cosine invariants were computed from the triple products, after determination of a semiempirical scale factor as proposed by Hauptman *et al.* (1969).

Our main purpose was to discover what improvement was brought to the knowledge of cosine invariants by the computation of the triple product, compared with the value +1.0 with which they are identified in the use of the tangent formula. By comparing cosine invariants computed by the triple product formula with their real values determined when the structure was solved, we see that:

(a) The cosine invariants computed near + or -1.0 fit quite well with their real values.

Table 1. Mean deviation of cosine invariants with whole set of normalized observed structure factors

 	A #11	~ ~	*^
 			••
			L.7

	$ heta_{comp}$	< 60° > 120°			$60^\circ < \theta_{co}$	_{omp} < 120°
$\langle A \rangle$	Standard deviations			Standard deviations		
	Quantity	Triple product	Tangent	Quantity	Triple product	Tangent
1·428 < A < 1·529	59	0.56	0.34	32	0.75	0.49
1·530 <i><A<</i> 1·641	60	0·26	0.34	27	0.77	0.44
1·642 <i><A<</i> 1·799	59	0.32	0.30	32	0.87	0.48
1·800 < <i>A</i> < 1·992	61	0·24	0.33	32	0.81	0.48
1·999 <i>< A</i> < 2·487	95	0.25	0.32	33	0.74	0.43
2·490 < A < 4·298	40	0.19	0.21	14	0.67	0.25

Table 2. Mean deviation of cosine invariants with structure factors corresponding to $\sin \theta < 0.7$

	Invariants					
$\langle A \rangle$	< 60° $\theta_{comp} > 120^{\circ}$ Standard deviations			$60^\circ < \theta_{comp} < 120^\circ$ Standard deviations		
	Quantity	Triple product	Tangent	Quantity	Triple product	Tangent
·428 < A < 1,529	63	0.51	0.42	28	0.76	0 ·48
1.530 < A < 1,641	60	0.34	0.34	27	0.70	0.30
1.642< <i>A</i> <1,799	58	0.39	0.35	33	0.80	0.30
·800 < A < 1,992	59	0.26	0.40	34	0.80	0.40
·999 < <i>A</i> < 2,487	94	0.30	0.30	34	0.60	0.53
2 ·490 <i>< A <</i> 4,298	43	0. 28	Q·22	10	Q·7Q	0.50

(b) On the contrary cosine invariants computed near 0 are generally quite wrong.

In Tables 1 and 2 we present the mean deviation of the real cosine invariants from the computed values and from the +1.0 determination used with the tangent formula. We consider two classes:

(a) The set of cosine invariants corresponding to phases less than 60° or greater than 120° .

(b) Those corresponding to a phase between 60° and 120° .

The comparison first uses the whole set of normalized observed structure factors (Table 1) and then only the structure factors corresponding to $\sin \theta < 0.7$ (Table 2). The comparison clearly shows the dependence of the computation of the cosine invariants by the triple-product formula on the decrease in resolution.

Table 1 also shows that, for the first class of invariants, the use of the triple-product formula brings about a real improvement compared with the tangent formula.

On the other hand, for the second class, the computed cosines appear quite wrong and it seems better to use the tangent formula. However we note, for this second class, that the real cosine invariants deviate much more from the implicit value +1.0 than do those of the first class. The use of the triple-product formula is thus distinguishing between the different structure invariants used in the phase determination.

Finally, if the computation of cosine invariants by the triple-product formula does not seem sufficiently reliable to justify such a time-consuming technique as the least-squares refinement proposed by Hauptman *et al.* (1969), nevertheless the possibility of using those results during the whole generation in a fast technique like symbolic addition or multisolution would be interesting. The weighting scheme we now propose permits this.

ì

Proposed weighting scheme

The structure invariants $H_1 + H_2 + H_3$ are ordered in groups of 100 elements in decreasing values of A_{H_1, H_2} . Within each group, we **may** consider to a first approximation the A_{H_1, H_2} values as constant and equal to \overline{A} . Then in each group, the structure invariants are ordered, according to increasing triple product. f is the rank of a given invariant obtained.

If the cosine invariants follow the theoretical distribution defined by Cochran (1955), the probability that a cosine invariant is less than x, is:

$$F(x) = \frac{1}{\pi I_0(A)} \int_{-1}^{x} \frac{\exp{(At)}}{\sqrt{1-t}} dt .$$

For any group of N=100 elements, the theoretical cumulative frequency corresponding to the value x is f=NF(x). Reciprocally, the invariant of rank f should correspond to a cosine equal to $x_1=F^{-1}(f/N)$. The variation of x_1 as a function of f for A=2.0 is shown in Fig. 1,

If, on the contrary, the triple-product values bore no relation to the corresponding cosine invariants the most probable value for each of them would remain the mean value defined by Germain, Main & Woolfson (1970):

$$x_2 = \frac{I_1(A)}{I_0(A)} = G(A)$$
.

In fact, if, as is shown by the experimental study, the partition achieved by the triple-product computation is not perfect, we may still expect a certain selection among the cosine invariants in which case they are no longer randomly distributed. For an invariant of rank f, the most probable cosine value may be expressed as:

$$x = \gamma x_1 + (1 - \gamma) x_2,$$

where γ lies between 0 and 1 and represents the confidence we have in the ability of the triple product computation to sort the cosine invariants. We think that $\gamma = 0.5$ is a realistic estimate of the capability of such a computation.

The weight A_{H_1, H_2} normally used in the tangent formula corresponds to a value x_2 taken as the most probable cosine invariant. If we obtain, with the tripleproduct formula, x instead of x_2 as the most probable value, it seems logical to associate it reciprocally with a weight A'_{H_1, H_2} defined by:

$$A'_{H_1,H_2} = G^{-1}(x).$$

The cosine invariants for which x_1 is close to 1 and thus $x < x_2$ will have a new weight A'_{H_1, H_2} greater than the initial A_{H_1, H_2} value. On the other hand, if x_1 is close to -1, the weight A'_{H_1, H_2} will be lower than A_{H_1, H_2} .

The proposed weighting scheme permits use of all the information provided by the cosine-invariant computation directly in the tangent-formula refinement, as well as in the symbolic-addition procedure. Furthermore it eliminates the sign indetermination which follows the cosine invariant inversion which is the principal difficulty of the process described by Hauptman *et al.* (1969).

As the immediate consequence of this weighting scheme is to weaken the disastrous influence of the negative cosine invariants during the phase determination, we may expect an improvement of the generation process and a greater chance of success. Moreover our weighting scheme should influence the choice of the origin reflexions obtained by the convergence method (Germain *et al.*, 1970) thus giving a better starting set.

The use of this new weighting scheme also gives an improvement in the final phases obtained by refinement with the tangent formula. For example, with androstanolone, starting with the 'fixing origin and symbol' reflexions corresponding to the correct solution and refining them by the tangent formula with the normal weighting scheme and with our new one, the mean deviation between the computed and the real phases is $\langle \Delta \Phi \rangle = 19.7^{\circ}$ and 15.9° respectively.

Application to the determination of some crystal structures

The determination process described above was developed in the set of programs 'Methodes directes' used at Bordeaux (*MEDIBOR*). Parts of it are written in assembler language, mainly the search of the \sum_2 relations and the computation of the triple products (we have two versions: one for an IBM 360-44 and one for a CII IRIS 80 computer). This program has been used successfully in the determination of four crystal structures:

(1) For stilbamidine diisethionate

 $(\frac{1}{2}C_{16}N_4H_{16}.SO_3CH_2CH_2OH.H_2O, P\bar{1}, Z=2)$ (Courseille, Busetta, Comberton & Hospital, 1971), the Fourier synthesis obtained with signs determined by *MEDIBOR* showed not only the atoms of the two molecular groups but also the unexpected water molecule.

(2) For androstenedione $(C_{19}O_2H_{20}, P2_12_12_1, Z=4)$ (Busetta, Comberton, Courseille & Hospital, 1972) we obtained a very good determination as is shown by the small mean deviation of the computed from the real phases $(\langle \Delta \Phi \rangle = 11.8^{\circ})$ as well as by the absence of spurious peaks at the centre of the hydrocarbon rings which we have observed for other steroids determined by the multisolution (*MULTAN*) method.

These two structures would undoubtedly have been solved by the multisolution method, although the phase determination would have been less accurate. However, for the following two structures no satisfactory solution was obtained by the *MULTAN* program.

(3) For a dimer related to diquinacrine

 $(C_{37}O_2N_3Cl_2H.2CHCl_3, P\overline{1}, Z=2)$ (Courseille, Busetta & Hospital, to be published) none of the 128 solutions seems better than the others as far as the usual figures of merit are concerned, and the five or six solutions we used for a Fourier synthesis did not yield a suitable trial structure. On the other hand one of the 64 solutions obtained with *MEDIBOR* gave very good values for the figures of merit $\sum \alpha$ and Ψ_0 . The Fourier map computed with this solution showed 36 of the 44 atoms α of the dimer molecule.

(4) For ethylenediamine (ED) ditartrate

 $(C_4O_6H_6.\frac{1}{2}C_2N_2H_6, P4_12_12, Z=8)$ (Perez, 1973), the presence of two molecular islands (Busetta, 1973) and also the particular location of the ED ion with respect to the 4₁ axis made phase generation much more difficult. None of the solutions provided by *MULTAN* gave an *ABSFOM* figure of merit greater than 0.80, which suggests a wrong determination. Of the different solutions provided by a cosine-invariant computation, only one appeared self-consistent (*ABSFOM*=0.975). It permitted location of the tartrate ion. In order to find the half ED ion, a new generation of phases was initiated with the phases given by the tartrate ion and with a convenient variable symbolic phase for defining the ED island. This process, which illustrates an application of the method described elsewhere (Busetta 1973) for determining structures with molecular islands, provided, by tangent-formula refinement, a set of phases for which the mean deviation to the real final phases reached the very low value $\langle \Delta \Phi \rangle = 6^{\circ}$ for reflexions with E > 1.40.

These results are sufficient to encourage us to attempt the determination of crystal structures containing molecules with as many as 100 atoms.

The authors are indebted to Dr M. Hospital for stimulating discussions and for his critical reading of the manuscript.

References

- BEURSKENS, P.T. (1963). Technical Report on Sign Correlation by the Sayre equation, The Crystallography Laboratory, University of Pittsburgh, Pa.
- BUSETTA, B. (1973). Thèse doctorat ès-Sciences Bordeaux.
- BUSETTA, B., COMBERTON, G., COURSEILLE, C. & HOSPITAL, M. (1972). Cryst. Struct. Commun. 1, 129–133.
- BUSETTA, B., COURSEILLE, C., FORNIES-MARQUINA, J. M. & HOSPITAL, M. (1972). Cryst. Struct. Commun. 1, 43-46.
- BUSETTA, B., COURSEILLE, C. & HOSPITAL, M. (1973). Acta Cryst. B29, 298-313.
- COCHRAN, W. (1955). Acta Cryst. 8, 473-478.
- COURSEILLE, C., BUSETTA, B., COMBERTON, G. & HOSPITAL, M. (1971). C. R. Acad. Sci. Paris, 272, 1115-1118.
- DECLERC, J. P., GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1973). Acta Cryst. A 29, 231-234.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). Acta Cryst. B26, 274–285.
- HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. A. (1969). Acta Cryst. B24, 811-814.
- KARLE, J. & HAUPTMAN, H. (1953). Solution of the Phase Problem. I. The Centrosymmetric Crystal. ACA monograph No. 3.
- KARLE, J. & HAUPTMAN, H. (1958). Acta Cryst. 11, 264-269.
- MESSAGER, J. C. & TSOUCARIS, G. (1972). Acta Cryst. A 28, 482-485.
- PEREZ, S. (1973). Thèse 3ème cycle Bordeaux.
- TSOUCARIS, G. (1970). Acta Cryst. A 26, 492-499