Acta Cryst. (1985). A41, 457-462

A New Method of Estimating a Triple Phase Invariant *via* its Quintet Extension: an Assessment

BY CHRISTOPHER J. GILMORE

Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

AND HERBERT HAUPTMAN

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(*Received* 10 *January* 1984; *accepted* 15 *April* 1985)

Abstract

The system of ten linear equations derived in the previous paper [Hauptman (1985), *Acta Cryst.* A41, 454-456, equations (3.21)-(3.30)] contains a triple phase cosine as an unknown parameter, which can be calculated *via* standard techniques. The viability of this set of equations and the accuracy with which cosine invariants may be calculated is assessed with reference to a number of variable parameters in the equations. The assessment is performed on both ideal and real data sets; in both cases the method is capable of identifying a well determined subset of invariants for use in direct methods.

1. Introduction

In the previous paper (Hauptman, 1985) a method of estimating the cosines of triple phase invariants *via* a system of linear equations described. In these equations the triplet cosine is one of ten unknown parameters. The purpose of this paper is to investigate the efficacy of the technique as a method of estimating these cosines. The method is first applied to idealized crystal structures and then to real crystallographic data sets. The following features of the method will be discussed in detail.

(1) The overall viability of the technique. There are already several methods of estimating triplet cosines from the quintet extension, these include the MDKS and TPROD formulae (Hauptman, 1972), the use of quadrupoles (Viterbo & Woolfson, 1973), the P_6 formula (Giacovazzo, 1976, 1977), the use of fourth-order Karle-Hauptman determinants (Messager & Tsoucaris, 1972; Karle, 1979, 1980) and more recently the P_{10} formula of Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori (1984), which exploits the second representation of the triplet and hence the space-group symmetry in a more systematic way. Many of these workers have used the MDKS formula as a comparison and accordingly this paper does likewise. In this way, it is possible to compare the above methods.

(2) The linear-equations approach (subsequently called LE) uses the second neighbourhood of the triplet. If the latter is defined as

$$\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}}=\varphi_3,$$

where $\mathbf{h}+\mathbf{k}+\mathbf{l}=0$, the first neighbourhood comprises the E magnitudes $|E_{\mathbf{h}}|$, $|E_{\mathbf{k}}|$ and $|E_{\mathbf{l}}|$; the second neighbourhood uses the seven additional magnitudes $|E_{\mathbf{h}\pm\mathbf{H}}|$, $|E_{\mathbf{k}\pm\mathbf{H}}|$, $|E_{\mathbf{l}\pm\mathbf{H}}|$ and $|E_{\mathbf{H}}|$, where $\mathbf{H}$ is a floating vector such that $|E_{\mathbf{H}}|$ is not necessarily large. In most of the previous utilizations of this neighbourhood $|E_{\mathbf{H}}|$ is constrained to be large, but the present approach does not require it, and so it is necessary to define an optimum range of $|E_{\mathbf{H}}|$.

(3) A similar problem arises concerning the magnitudes of the diagonal coefficients $B_1^2, B_2^2, \ldots, D_3^2$ of equations (3.21)-(3.30) in the previous paper. Thus in (3.21) B_1^2 is constrained such that $B_1^2>\alpha$, where α is 'large'; a similar constraint applies to B_2^2 in equation (3.22) *etc.* For practical purposes, it is necessary to define the optimum value of α. Similarly, terms D_1^2, D_2^2 and D_3^2 are required to be $<\beta$ where β is 'small' in equations (3.21)-(3.23) and an optimum value of β also needs to be defined.

(4) As formulated, there are ten linear equations with ten unknown parameters:

(i) three functions of $|G|$, *i.e.* $|G_{\mathbf{h}}|^2-1$, $|G_{\mathbf{k}}|^2-1$ and $|G_{\mathbf{l}}|^2-1$;

(ii) seven cosine invariants, one of which is the desired cosine $\cos\varphi_3$.

For equal-atom structures $G=E$ and hence all the terms in (i) may be considered known. The system of linear equations reduces to a problem in linear least squares (subsequently referred to as LS) with ten equations and seven unknown parameters, all of which are three-phase invariant cosines. It might be expected that this overdeterminacy would improve the accuracy of the cosine estimates, but this feature needs to be investigated quantitatively. An optimum weighting scheme for the LE system was also considered.

(5) The proper utilization of E magnitudes missing from the second neighbourhood is also important. If

0108-7673/85/050457-06$01.50

Table 1. *Deviations between observed and calculated triple cosines for all calculated cosines*

The constraints are as follows:
(*a*) The diagonal matrix elements B_1^2, B_2^2, ..., >1·0.
(*b*) No constraints on $|E_H|$.
(*c*) No missing $|E|$'s.
(*d*) Linear equations employed.

	CARD			PENT			TPALA			CROWN		
A limits	R.m.s. dev.*	Mean dev.†	Number	R.m.s. dev.*	Mean dev.†	Number	R.m.s. dev.*	Mean dev.†	Number	R.m.s. dev.*	Mean dev.†	Number
<1·0	0·56	0·35	9	*x*	*x*	0	1·08	0·42	259	1·03	0·35	47
1·0–1·1	0·93	0·30	46	*x*	*x*	0	1·03	0·42	349	1·11	0·50	84
1·1–1·2	0·92	0·42	111	*x*	*x*	0	1·02	0·42	414	1·12	0·53	118
1·2–1·4	0·92	0·43	263	1·14	0·61	47	1·00	0·45	814	1·15	0·46	211
1·4–1·6	0·87	0·53	351	1·10	0·65	230	0·98	0·45	554	1·08	0·48	175
1·6–1·8	0·90	0·62	333	0·97	0·53	291	1·02	0·54	404	0·95	0·37	91
1·8–2·0	0·92	0·65	275	1·02	0·57	204	1·00	0·45	240	0·93	0·48	59
2·0–2·5	0·88	0·67	541	0·95	0·58	246	0·99	0·49	212	1·00	0·34	38
2·5–3·0	0·88	0·72	373	0·95	0·66	66	0·92	0·58	62	0·73	0·35	12
>3·0	0·98	0·88	525	0·56	0·42	31	0·72	0·32	16	2·00	2·00	1

* Root-mean-square deviation.
† Mean deviation.

one or more of the vectors **h**, **k**, **l** have indices close to the maximum or mininum value for the data set in question, then there are very severe restrictions on the vector **H** if all the required E magnitudes associated with the six cross vectors $\mathbf{h}\pm\mathbf{H}$, $\mathbf{k}\pm\mathbf{H}$, $\mathbf{l}\pm\mathbf{H}$ are to be within the measured data set. This constraint coupled with the restrictions in (3) will limit the triplets that are accessible to the equation. However, if the missing cross terms are included with E magnitudes of unity the severity of the problem is reduced. The viability of this approach and the maximum permissible number of missing members in a given neighbourhood was thus investigated.

(6) The theory was derived for space group $P1$. The validity of the equations in other common space groups was tested, although, unlike the P_{10} formula (Cascanaro *et al.*, 1984), the space-group symmetry is not exploited fully.

Only the triplet $\cos\varphi_3$ was assessed. The remaining triplets in the equation system do not usually involve three large E magnitudes in the first neighbourhood. When, however, this happens, these invariants show the same characteristics as $\cos\varphi_3$ in the behaviour of their estimated cosines. This is to be expected since the invariants in the equations all have the same $2N^{-3/2}$ dependence (N is the number of atoms - assumed equal - in the unit cell), and similar coefficients in the equations and so should be determined with similar accuracy.

2. Experimental results for idealized data

As an initial test of viability, ideal data sets were generated for four structures using published coordinates and excluding hydrogen atoms. Data were generated to a θ limit corresponding to the Cu sphere. This corresponds exactly to the theoretical conditions under which the equations are valid, and is therefore a useful first test. The four structures were:

(1) (20*S*)-3β,14-dihydroxy-5β,14β-cardanolide, $C_{23}H_{36}O_4$ (Messerschmidt, Hohne & Lindig, 1981; Rohrer, Fullerton, Yoshioka, Kitatsuji, Ahmed & From, 1983). The space group is $P1$ with $Z=1$. This will be referred to by the acronym CARD.

(2) A pentapeptide Ac-aib-pro-aib-ala-aib-OBzl, $C_{29}H_{44}N_5O_7$ (Smith, Fitzgerald & Duax, 1981). The space group is $P1$ with $Z=2$. This is a difficult structure, originally solved by molecular replacement. The acronym is PENT.

(3) A peptide Boc-pro-aib-ala-aib-OBzyl, $C_{28}H_{42}N_5O_7$ (Smith, Pletnev, Duax, Balasubramanian, Bosshard, Czerwinski, Kendrick, Matthews & Marshall, 1981). This is another difficult structure in space group $P2_1$ with $Z=2$. The acronym is TPALA.

(4) A crown ether $C_{56}H_{68}O_{10}$ (Goldberg, 1980). This is a relatively large structure in space group $P2_12_12_1$ with $Z=4$. The acronym is CROWN.

Table 1 summarizes the results of the application of equations (3.21)-(3.30) from the previous paper. The results are tabulated for the optimum settings of the parameters (2)-(5) described in the *Introduction*. It can be seen that individual cosine estimates are unreliable with rather large positive mean deviations and r.m.s. deviations [deviations are defined as $(\cos\varphi)_{\text{calculated}}-(\cos\varphi)_{\text{true}}$ throughout this paper], but Table 2 shows that it is possible to identify the negative cosines with some confidence, and hence exclude them from the phasing process in an *a priori* direct-methods analysis. It was also found that both the LE and LS methods produced results that were largely consistent with each other, and that they were of similar accuracy. Both were relatively insensitive to the parameters (2)-(5) outlined in the *Introduction*.

Table 2. *The number of cosines that are negative and the number that were successfully identified*

The constraints are as defined in Table 1.

	CARD		PENT		TPALA		CROWN	
A limits	Number negative	Number found	Number negative	Number found	Number negative	Number found	Number negative	Number found
<1·0	2	2	0	0	72	53	13	9
1·0–1·1	13	10	0	0	74	43	17	13
1·1–1·2	24	18	0	0	81	46	26	18
1·2–1·4	54	39	6	4	128	100	38	20
1·4–1·6	49	42	24	17	79	51	35	22
1·6–1·8	35	30	29	21	32	22	14	9
1·8–2·0	11	9	16	9	40	21	10	8
2·0–2·5	9	9	10	8	19	10	4	2
2·5–3·0	2	2	3	3	3	2	0	0
>3·0	0	0	0	0	1	0	0	0
Total	199	161	88	62	528	348	157	101

Table 3. *The six structures selected for LE and LS analysis and their crystallographic details*

Compound name	Reference	Chemical formula	Space group	*Z*
TPALA	Smith *et al* (1981)	$C_{28}H_{42}N_4O_7$	$P2_1$	2
PGE2	DeTitta, Langs, Edmonds & Duax (1980)	$C_{20}H_{32}O_5$	$P1$	1
TVAL	Karle (1975); Smith *et al.* (1975)	$C_{54}H_{90}N_6O_{18}$	$P1$	2
SUOA	Oliver & Strickland (unpublished)	$C_{28}H_{38}O_{10}$	$P2_12_12_1$	4
MUNICH1	Szeimies-Seebach *et al.* (1978)	$C_{20}H_{16}$	$C2$	8
AZET	Colens, Declercq, Germain, Putzeys & Van Meerssche (1974)	$C_{21}H_{16}ClNO$	$Pca2_1$	8

The formulae work equally well in any space group tested. It was also found that, in the least-squares environment, the standard deviations for the cosines derived from the error matrix in the usual way were not a reliable guide to the accuracy of the cosine estimate. Also, when using LS it was possible to compare agreements between observed E magnitudes (and hence $|E|^2-1$ magnitudes) and the corresponding calculated $|G|^2-1$ terms from the equations. Typical R factors relating observed and calculated E magnitudes were about 0·2. However, many systems of equations were found where the observed and calculated E's were in poor agreement. Surprisingly, it was not possible to correlate triplet reliability with the occurrence of poor E agreement – many well estimated triplets had poorly estimated E magnitudes. This topic was not pursued further with real data sets.

3. Experimental results on real data sets

Having established the viability of the formulae (although in a somewhat limited sense) with idealized data, the techniques were then applied to real data sets. For this purpose the database of structures resistant to direct methods compiled by Sheldrick was used. Table 3 summarizes the relevant crystallographic details of the six structures selected for analysis from this database.

As expected, these tests were much less successful:

(i) the cosine estimates were highly sensitive to even small variations in the parameters (2)–(5);

(ii) most cosines were outside the region −1·0 to +1·0;

(iii) the LE and LS estimates often differed markedly.

To overcome this, two scaling parameters, C and K, were introduced to produce a scaled estimate $\cos\varphi_3'$ such that

$$\cos\varphi_3' = K(\cos\varphi_3 + C)$$

(K is >0, there is no restriction on C). These parameters were used to force the distribution of $\cos\varphi_3'$ to follow the Cochran distribution as far as possible (Cochran, 1955). This is by no means a unique requirement of this method – scaling is also required in the MDKS, TPROD and P_{10} formulae. The method of scaling closely follows that used in deriving the constants M and K in the MDKS formula:

(i) The triplets are sorted on A value, where

$$A = 2N^{-1/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{l}}|$$

and then divided into groups of 100–200. Under these circumstances each group contains cosines of approximately constant A.

(ii) For each group, the triplets are re-sorted in descending order of $\cos\varphi_3'$. C is then chosen so that the fraction of positive cosines in the group is given by

$$\int_0^1 e^{Ax}[\pi I_0(A)(1-x^2)^{1/2}]^{-1}\,dx.$$

(iii) Finally, K is selected such that

$$\langle\cos\varphi_3\rangle = I_1(A)/I_0(A),$$

where $I_0(x)$ and $I_1(x)$ are modified Bessel functions of the first kind.

Different scaling constants are used for LE and LS.

This method was applied to six structures in the Sheldrick database (Table 3). For each structure over 100 LE and LS analyses were performed in which the parameters (2)–(5) in § 1 were varied in a systematic way. From an analysis of the accuracy of the cosine estimates and the ability of the formula to predict negative cosines, the following general observations and rules may be formulated:

(1) The optimum range of $|E_H|$ is that which spans the large E magnitudes; there is no benefit (nor any loss) in including the low E_H magnitudes as well, and there is a considerable increase in computer time required. Typically a range $|E_H| > 1{\cdot}0$–$1{\cdot}3$ is suitable (just as in MDKS).

(2) The constraints α and β described in the *Introduction* should be around 1·0 and 0·5 respectively. Harder constraints than this reduce the number of accessible cosines without any improvement in either the accuracy or in the ability to identify negative cosines. Relaxing the constraints further, however, does worsen the reliability of the method.

(3) Both the LE and LS techniques give answers of similar accuracy. Usually the estimates agree, but where there is a marked disagreement, it is optimum to use the minimum cosine estimate when identifying negative cosines. If both estimates are greater than zero, then the largest cosine estimate is usually the most reliable. This is the preferred method of using the method in a structure-solving environment.

A LS weighting scheme whereby each equation was given a weight

$$w = 1/n^{1/2}$$

was found to be satisfactory, where n is the number of terms contributing to the averages in each equation. This scheme was marginally more accurate than unit weights especially when some equations had few contributors.

It was found that the cosine estimates retained their viability even when some equations had only one contributor, and that any requirement for a minimum number of contributors of ten or 20 served only to restrict the number of cosines accessible without any consequent improvement in their accuracy.

(4) A similar situation prevails concerning the maximum number of missing second neighbours permitted. The accuracy does not decline when missing second neighbours are allowed, although there is a corresponding increase in the computer time needed since more cross-term searches are successful. If a maximum of three such missing cross terms is allowed and their magnitudes are set to unity, more triplets are accessible to the technique and there is a marginal increase in accuracy. Beyond three missing neighbours, there is a rapid fall in accuracy, as well as a further increase in computer time.

In summary, for practical utilization of the LE and LS formulae, minimal constraints should be applied in setting up the equations:

(i) restrict $|E_H| > 1{\cdot}0$–$1{\cdot}3$;

(ii) set $\alpha = 1{\cdot}0$ and $\beta = 0{\cdot}5$;

(iii) use a combination of LE and LS;

(iv) use equation weights of $1/n^{1/2}$ with a minimum of one contributor to each equation;

(v) allow up to three missing second neighbours.

Table 4. *R.m.s. and mean deviations and percentage of negative triplet cosines detected for the six data sets listed in Table* 3

The results use the LS formula (the LE results are very similar) except for the detection of negative cosines, where a combination of LE and LS is used. The MDKS results are in parentheses.

Compound name	% of triplets accessible	R.m.s. deviation	Mean deviation	% of negative cosines found
TPALA	83 (100)	0·58 (0·60)	−0·01 (−0·05)	35 (21)
PGE2	78 (100)	0·44 (0·44)	−0·02 (−0.06)	27 (19)
TVAL	72 (100)	0·69 (0·69)	−0·02 (−0.03)	55 (26)
SUOA	72 (100)	0·69 (0·69)	−0·27 (−0.03)	42 (25)
MUNICH1	77 (100)	0·61 (0·62)	−0·10 (−0·12)	44 (30)
AZET	79 (100)	0·58 (0·64)	−0·12 (−0.16)	61 (40)

4. Results

Table 4 summarizes the results obtained for six structures whose triplet cosines were estimated *via* LE and LS using the constraints (i)–(v) above. Calculations using the MDKS formula are included for comparison. [There is a comparison of MDKS and TPROD in Hauptman (1972), and a comparison of MDKS and P_{10} and P_6 in Cascanaro *et al.* (1984), so that it is possible to extend this comparison to other triplet-estimating formulae.] Table 5 duplicates some of the contents of Table 4, but includes only those cosines that had estimates $> 0{\cdot}8$; these should be well behaved subsets with few anomalous triplets, although it is important to remember that the structures under consideration here are *difficult* structures for direct methods so that the results will be correspondingly better for easier structures. However, this technique is only likely to be used in situations of difficulty, so that it is sensible to assess the method in these cases.

Both the LE and LS techniques produce estimates of the cosine in the range −1·0 to +1·0 after scaling, although a substantial number of cosines are outside this range. (In calculating mean and root-mean-square deviations, cosines outside this range have been set to −1·0 and +1·0 respectively.) For reasons of space, Tables 4 and 5 do not show the variation of the cos φ_3 estimate as a function of A value. In all cases the accuracy decreases as A decreases. When A is less than unity no worthwhile estimates are available, and these are excluded from the analysis.

It can be seen from r.m.s. and mean deviation calculations that, just as with perfect data sets, individual cosine estimates are too unreliable for

Table 5. *R.m.s. and mean deviations for the six data sets in Table* 3

Only cosine estimates >0·8 are included; the final column is the number of negative cosines that were undetected. The MDKS results are in parentheses.

Compound name	Number of triplets	R.m.s. deviation	Mean deviation	Number of negative cosines undetected
TPALA	381 (413)	0·43 (0·48)	0·21 (0·27)	20 (28)
PGE2	414 (588)	0·31 (0·36)	0·15 (0·16)	13 (27)
TVAL	533 (510)	0·36 (0·46)	0·12 (0·16)	20 (30)
SUOA	521 (608)	0·59 (0·63)	0·32 (0·35)	62 (80)
MUNICH1	318 (322)	0·42 (0·50)	0·21 (0·26)	20 (24)
AZET	485 (559)	0·33 (0·45)	0·14 (0·21)	16 (34)

them to be used directly. This is perhaps not surprising; a similar situation is found with the MDKS and related formulae. The cosine terms in equations (3.21)–(3.30) are of order $2N^{-3/2}$ whereas the $|G|^2-1$ terms are of order N^{-1}; hence the former may be considered small correction terms to the latter and hence less reliably determined. Also, as N increases, the difference between these two sets of terms will become much greater and hence the reliability will fall. It can be seen, however, that the method is more accurate than the MDKS formula.

In direct methods, the triplets that cause most difficulty are usually those that have $\cos\varphi_3 < 0{\cdot}0$. The LE and LS method is quite efficient in identifying such relationships, and this is also shown in Tables 4 and 5. A comparison with the MDKS results shows once again that the LE–LS method is more efficient than MDKS. However, a number of cosines are also flagged as having negative cosines when their true value is >0·0. This is also a feature of related formulae. For this reason, triplets identified as having negative cosines should be excluded from the direct-methods analysis rather than setting $\varphi_3 = \pi$. The over-determinacy of the phasing procedure makes this a viable proposition.

It is equally useful in direct methods to have available a subset of triplets whose true cosines are close to unity for use in the initial stages of phasing. In Table 5 those cosines estimated to be >0·8 are extracted from the full set and analysed. It is particularly noteworthy that the number of negative triplets is greatly reduced and there is a fall in the r.m.s. deviation.

One deficiency of the present technique is the problem of inaccessible triplets – a problem that does not generally arise in other methods. This arises from a dual system of constraints being applied in the building of the equations; not only are constraints placed upon $|E_{\mathbf{H}}|$ as in other triplet cosine formulae, but there are additional constraints imposed on those terms that may contribute to the averages in the equations. In consequence, 15–30% of the triplets are rendered inaccessible.

The direct-methods program *MITHRIL* (Gilmore, 1984) provides facilities for estimating triplet cosines *via* LE and LS, and for filtering out those relationships identified as unreliable whilst optionally up-weighting those that are deemed to belong to that subset identified as well determined. Triplets that are inaccessible can be retained or down-weighted. The computer time required to calculate the $\cos\varphi_3$ estimates is similar to that required by the MDKS and related methods, but its method of building and inverting matrices increases the time required by about 50% over that needed by MDKS.

5. Concluding remarks

The system of linear equations or least squares provides a method of estimating triple phase cosines that is at least superior to the MDKS formula. Although unable to provide accurate individual cosine estimates, the subsets for which $\cos\varphi_3 > 0{\cdot}8$ and $\cos\varphi_3 < 0{\cdot}0$ are quite reliable and may be used as the basis of a filtering technique by which poor triplets are removed and reliable triplets upweighted in a direct-methods analysis.

CJG wishes to acknowledge a period of study leave granted by the University of Glasgow. The research described here was funded by Grant No. CHE-8203930 from the National Science Foundation and a grant from the James H. Cummings Foundation, Inc.

References

CASCANARO, G., GIACOVAZZO, C., CAMALLI, M., SPAGNA, R., BURLA, M. C., NUNZI, A. & POLIDORI, G. (1984). *Acta Cryst.* A**40**, 278–283.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.

COLENS, A., DECLERCQ, J.-P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.

DETITTA, G. T., LANGS, D. A., EDMONDS, J. W. & DUAX, W. L. (1980). *Acta Cryst.* B**36**, 638–645.

GIACOVAZZO, C. (1976). *Acta Cryst.* A**32**, 967–976.

GIACOVAZZO, C. (1977). *Acta Cryst.* A**33**, 527–531.

GILMORE, C. (1984). *J. Appl. Cryst.* **17**, 42–46.

GOLDBERG, I. (1980). *Acta Cryst.* B**36**, 2104–2108.

HAUPTMAN, H. (1972). *Crystal Structure Determination: The Role of the Cosine Seminvariants.* New York: Plenum Press.

HAUPTMAN, H. (1985). *Acta Cryst.* A**41**, 454–456.

KARLE, I. (1975). *J. Am. Chem. Soc.* **97**, 4379–4386.

KARLE, J. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 2089–2093.

KARLE, J. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 5–9.

MESSAGER, J. C. & TSOUCARIS, G. (1972). *Acta Cryst.* A**28**, 482–484.

MESSERSCHMIDT, A., HOHNE, E. & LINDIG, C. (1981). *Cryst. Struct. Commun.* **10**, 629–631.

ROHRER, D. C., FULLERTON, D. S., YOSHIOKA, K., KITATSUJI, E., AHMED, K. & FROM, A. H. L. (1983). *Acta Cryst.* B**39**, 272–280.

SMITH, G. D., DUAX, W. L., LANGS, D. A., DETITTA, G. T., EDMONDS, J. W., ROHRER, D. C. & WEEKS, C. M. (1975). *J. Am. Chem. Soc.* **97**, 7242–7247.

SMITH, G. D., FITZGERALD, P. M. D. & DUAX, W. L. (1981). Proc. Am. Crystallogr. Assoc. Winter Meet., Texas, Abstr. PB12.
SMITH, G. D., PLETNEV, V. Z., DUAX, W. L., BALASUBRAMANIAN, T. M., BOSSHARD, H. E., CZERWINSKI, E. W., KENDRICK, N. E., MATHEWS, F. S. & MARSHALL, G. R. (1981). *J. Am. Chem. Soc.* **103**, 1493-1501.
SZEIMIES-SEEBACH, U., HARNISCH, J., SZIEMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J.-P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848-850.
VITERBO, D. & WOOLFSON, M. M. (1973). *Acta Cryst.* A**29**, 205-208.

Acta Cryst. (1985). A**41**, 462-466

Structure-Factor Calculations in Refinement of a Modulated Crystal Structure

BY W. A. PACIOREK AND D. KUCHARCZYK

Institute for Low Temperature and Structure Research, Polish Academy of Sciences, pl. Katedralny 1, 50-950 *Wrocław, Poland*

(*Received* 8 *October* 1984; *accepted* 15 *April* 1985)

Abstract

Problems concerning structural analysis of a one-dimensionally modulated structure using its (3+1)-dimensional symmetry are discussed. Simple modifications of the common structure-factor formula for occupational, displacive with small amplitudes and mixed modulation are obtained. The integral form of the structure factor known from literature is critically considered and an analytical form for harmonic displacive (not necessarily rectilinear) modulation has been found. Analytical corrections to the temperature factors have been introduced and generalized to cover the phase relationships of elliptic modulating waves. The results of this paper have been used to prepare a set of programs to refine modulated structures.

1. Introduction

As a result of many important works by de Wolff, Janssen & Janner (1981, and references therein), progress in the symmetry description of modulated structures by higher-dimensional crystallographic symmetry has been obtained. An alternative approach, based on the concept of wreath product, has been presented by Litvin (1980) and Koptsik (1978).

In the works by Yamamoto (1982*a*, *b*, *c*) a $(3+d)$-dimensional crystallographic symmetry approach has been applied to the structure refinement and the structure-factor formula (SFF) suitable for this purpose has been presented.

The form of the Debye-Waller factor for modulated structure has been discussed in papers by Overhauser (1971), Axe (1980) and Adlhart (1982) and applied by Steurer & Adlhart (1983) in the refinement of α-bis(*N*-methylsalicylideneiminato)nickel(II).

Although the SFF presented by Yamamoto covers practically all kinds of modulation, the necessity of numerical integration makes application unlikely in many cases for which much simpler formulae would be sufficient. There are also some questions still open concerning temperature-factor corrections, especially in relation to the treatment proposed by Overhauser and Axe. As will be discussed later, the approaches of Yamamoto and Adlhart can lead to different results.

The present work deals with structure factors of special as well as general (elliptical displacive waves included) cases of modulations. The compact analytical expression convenient for computing in the case of occupational, displacive with small amplitudes, mixed and general types of modulation in the harmonic approximation will be presented and their limitations discussed.

An extension of the temperature-factor corrections proposed by Axe is proposed for the case of general harmonic modulation with displacive wave.

Our considerations are restricted to the single-**q** modulated structures described by (3+1)-dimensional crystallographic groups.

The notation used in this paper is mostly adopted from original papers by de Wolff and Yamamoto in order to make the comparison easier.

2. Single-q modulated structure description

In the modulated structure the positional, occupational and thermal parameters can be written as periodic functions of a continuous parameter [or parameters in multidimensional modulation, see Yamamoto (1982*a*) for that more general case and the notation] as follows:

$$x_i^\mu(\bar{x}_4^\mu) = \bar{x}_i^\mu + \sum_n u_{i,n}^\mu \varepsilon_n + \text{c.c}, \quad i = 1, 2, 3,$$
$$y^\mu(\bar{x}_4^\mu) = \sum_n y_n^\mu \varepsilon_n + \text{c.c}, \quad y^\mu = P^\mu, B^\mu \text{ or } B_{ij}^\mu, \tag{2.1}$$

0108-7673/85/050462-05$01.50