Formulas for *n*-tet Phase Invariants and Embedded Seminvariants for All Space Groups

By Jerome Karle

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375, USA

Abstract

Formulas are presented for the calculation of the cosines of *n*-tet phase invariants and embedded seminvariants in all the space groups. They are shown to be of the form of a particular type of expected value formula that is derivable from the joint probability distribution. In the recent literature, formulas for phase invariants and seminvariants have been given in the form of conditional probability distributions. A detailed comparison of the relative merits of the two types of formulas, expected value and conditional distribution, has not yet been made. The variety of potential applications is quite vast and therefore it may require much effort to make evaluations of current theories. Should it seem worthwhile, the determinantal joint probability distributions employed in this paper could provide the basis for the derivation of additional conditional probability distributions. They are likely to be much more complex, however, than the expected value formulas. Some simple calculations with triplet and quartet invariants involving random structures in space group $P1$ show a considerable decrease in the reliability of the expected value formulas as the complexity of the structure increases. A comparable observation had been made in the past for conditional probability distributions for triplet phase invariants. Current theories present the possibility of obtaining information in special circumstances, for example, with respect to selected embedded seminvariants. How extensive and how useful such information might be, particularly with respect to the truly difficult structures that occur among the essentially equal-atom, noncentrosymmetric crystals with 100 or more nonhydrogen atoms in the asymmetric unit, remains to be seen.

Introduction

In an effort to improve procedures for phase determination, the mathematics of the joint probability distribution has been used to investigate the higher-order phase invariants, linear combinations of phases whose subscripts add up to zero. This interest in the higher-order phase invariants is understandable in view of the dramatic advances in structure determination that have accrued over the past thirty years from use of the mathematical properties of the third order, or triplet, phase invariant, $\varphi_{h_1} + \varphi_{h_2} + \varphi_{h_3}$, where $h_1 + h_2 + h_3 = 0$. Early in the development of phase determining formulas it was recognized that the higher-order phase invariants might be helpful in phase determination. For example, the special quartet formula, $\sum_3$ (Hauptman & Karle, 1953) was found to be useful in several early structural studies and was used extensively in the initial stages of the determination of the structure of p,p'-dimethoxybenzophenone (Karle, Hauptman, Karle & Wing, 1958).

Because of the tedious nature of the derivations, only a limited number of conditional probability distributions for individual higher-order invariants have so far been obtained. Associated with derivations of the joint probability distributions is 'neighborhood' theory (Hauptman, 1977*a,b*) or 'representation' theory (Giacovazzo, 1977, 1980*a*). These theories concern the types of structure factors to be combined in the joint distributions in order to achieve the desired conditional probability formulas for *n*-tet phase invariants and embedded seminvariants. Numerous individual applications of probability theory are required to derive new formulas of interest in the many space groups.

A mathematical approach that provides general formulas for computing the values of *n*-tet invariants and embedded seminvariants in all the space groups involves a type of expected value formula that will be the concern of this paper. The expected value formulas are, in the first instance, suggested by insights that derive from earlier experiences. They have also been obtained from determinantal joint probability distributions (Karle, 1982*a,b*). These formulas obviate the necessity for developing a separate neighborhood or representation theory since the importantly related structure factors are already contained within them. Methods for using the new formulas and some illustrative calculations will be presented.

Phase invariant formulas

The formulas for the computation of *n*-tet invariants and embedded seminvariants are of two kinds which

are termed here 'special phase invariant formulas' and 'general phase invariant formulas'. They are defined as follows:

Special n-tet invariant formulas

$$\cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \cdots + \varphi_{\mathbf{h}_n})$$

$$\simeq \langle(|E_{\mathbf{h}_i}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{h}_j}|^2 - 1)\ldots$$

$$\times (|E_{\mathbf{h}_i + \mathbf{h}_j + \ldots + \mathbf{h}_v}|^2 - 1)\rangle_{i,j,\ldots,v}$$

$$\times (S_n\langle|\text{Num.}|\rangle_{i,j,\ldots,v})^{-1}, \tag{1}$$

where $\mathbf{h}_1 + \mathbf{h}_2 + \cdots + \mathbf{h}_n = 0$, $n = 4, 5, 6, \ldots$; $i, j, \ldots$, $v = 1, \ldots, n$ except that no two of the subscripts can have the same integer value and the number of subscripts $i, j, \ldots, v$ is $n = 1$. If the nth subscript is w, then because of the relationship among the indices, $\mathbf{h}$, we have $\mathbf{h}_i + \mathbf{h}_j + \cdots + \mathbf{h}_v = -\mathbf{h}_w$. Duplication in computing the average terms over $i, j, \ldots, v$ in (1) is avoided by imposition of the rule $i < w$. The $\varphi_{\mathbf{h}}$ are phases of index $\mathbf{h}$, the E are normalized structure factors, S is a scaling function that can be a constant equal to unity and $\langle|\text{Num.}|\rangle$ means the average of the absolute values of the terms entering the average function in the numerator.

General n-tet phase invariant formulas

$$\cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \cdots + \varphi_{\mathbf{h}_n})$$

$$\simeq \langle(|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{h}_j + \mathbf{k}}|^2 - 1)\ldots$$

$$\times (|E_{\mathbf{h}_i + \mathbf{h}_j + \ldots + \mathbf{h}_v + \mathbf{k}}|^2 - 1)\rangle_{\mathbf{k},i,j,\ldots,v}$$

$$\times (S_{\mathbf{k},n}\langle|\text{Num.}|\rangle_{\mathbf{k},i,j,\ldots,v})^{-1}. \tag{2}$$

The elements of (2) are the same as those of (1) except for the extra factor $(|E_{\mathbf{k}}|^2 - 1)$ and the occurrence of the vector $\mathbf{k}$ in the index definitions of the various E. In addition, the integer values of n start with 3 instead of 4. The average may be taken over selected values for $\mathbf{k}$ as well as the $i, j, \ldots, v$. It is seen that (2) reduces to (1) if $\mathbf{k} \equiv (0,0,0)$. For a given $\mathbf{k} \not\equiv (0,0,0)$, there is no duplication of terms in taking the average over all possible combinations of values for the $i, j, \ldots, v$.

Rationale

In this part, the origins of formulas (1) and (2) are outlined as an extrapolation and formal extension of earlier results.

We recall the formulas,

$$|E_{\mathbf{h}}|^2 - 1 \simeq G_2\langle|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}+\mathbf{k}}|^2 - 1)\rangle_{\mathbf{k}} \tag{3}$$

and

$$|E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3}| \cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_3})$$

$$\simeq G_3\langle(|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{h}_j + \mathbf{k}}|^2 - 1)\rangle_{\mathbf{k},i,j}, \tag{4}$$

where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 = 0$ and $i, j = 1, 2, 3$; $i \neq j$. In (4), for a specific $\mathbf{k}$, the average is taken over the permitted combinations of i and j. If $\mathbf{k}$ varies over all possible values, taking the average over i and j yields only duplication, so that one selection of values for i and j suffices. Formulas (3) and (4) have been derived by algebraic and probability methods. Scaled versions of (3) and (4) that exhibit forms for the scaling functions G_2 (Hauptman, 1964; Karle, 1966) and G_3 (Karle, 1970) have been published. Inspection of (3) and (4) suggests an extrapolation to obtain the quartet formula

$$|E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{\mathbf{h}_3} E_{\mathbf{h}_4}| \cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_3} + \varphi_{\mathbf{h}_4})$$

$$\simeq G_4\langle(|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{k}}|^2 - 1)(|E_{\mathbf{h}_i + \mathbf{h}_j + \mathbf{k}}|^2 - 1)$$

$$\times (|E_{\mathbf{h}_i + \mathbf{h}_j + \mathbf{h}_k + \mathbf{k}}|^2 - 1)\rangle_{\mathbf{k},i,j,k}, \tag{5}$$

where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$; $i, j, k = 1, 2, 3, 4$ except that i, j and k must assume different integer values, *i.e.* $i \neq j$, k; $j \neq k$. It is quite apparent how the extrapolation demonstrated by (5) may be extended further to define n-tets of higher and higher order. Except for an alternative manner for presenting the scaling, the extrapolation generates the general formulas given by (2).

As indicated above, the special formulas (1) may be obtained from the general ones by setting $\mathbf{k} \equiv (0,0,0)$. From the manner in which general formulas are derived from the joint probability distribution, it is not proper to set $\mathbf{k} \equiv (0,0,0)$. Use of the latter setting is, however, an example of a formalism that can be useful despite its lack of rigor. Additional support for formulas (1) and also (2) derives from showing that both the special and general formulas can arise from appropriately formed determinantal joint probability distributions (Karle, 1982*a*,*b*).

Computation of formulas (1) and (2)

For the derivations of the theoretical results, formulas (1) and (2), the type of atomic arrangement that is most appropriate is a random one. Experience has shown that such formulas are also applicable to crystals in which there are regularly ordered arrangements of atoms despite the random basis for the derivation. In particular, crystals that have some structure factors of unusually large magnitude because of special regularities are generally particularly suited for obtaining information from (1) and (2). Calculations show and the probability theory indicates that the most reliable computations of the invariants are those that are associated with the largest products of $|E|$ corresponding to the phases forming the invariants. The values of the invariants that are computed most reliably are those close to zero. At times it is also possible to accept some calculations that indicate that the value of the invariant is close to π. Relatively fewer invariants

will be computed reliably, however, as the number of atoms in the unit cell increases. This is consistent with observations made from computations of the values of invariants from conditional joint probability distributions.

In the next part, calculations will be discussed and illustrations of the manner for calculating formulas (1) and (2) for various types of invariants will be presented.

Triplet invariants

The general formula (2) for the triplet invariant is comparable to a scaled version of the $B_{3,0}$ formula (Karle, 1970), which can be used to screen out triplets whose values deviate significantly from zero. Applications of this method are described by Karle, Gibson & Karle (1970) and Flippen (1972). This type of formula rapidly deteriorates in its reliability as the complexity of a structure increases. This is also true for the corresponding conditional probability distribution (Karle, 1980).

The test of this formula was an ideal one in which data were calculated for random structures in space group $P1$ for 20, 50, 100 and 200 atom structures. The number of independent data computed in each case, as seen in Table 1, corresponded approximately to the amount in a Cu sphere of reflection, except in the second example, for which calculations were based on four times the Cu sphere of data. Formula (2) for the triplet invariant was computed by setting the scaling factor, $S_{k,3}$, equal to unity and allowing k to range through the complete data set including Friedel mates. Under this circumstance any one selection of values for i and j suffices. Additional selections merely produce a duplication of terms contributing to the averages of (2). In addition, terms were accepted for the average only if at least two of the three factors in a term were formed from $|E|$ values that exceeded the value of b given in the third column of Table 1. In all cases 150 triplet invariants were computed, ones with the largest values for $|E_{h_1} E_{h_2} E_{h_3}|$. Column 5 shows how many of the 150 triplets were computed by formula (2) to have values equal to or exceeding the value in column 4. Negative terms of large magnitude were not considered since they have already been shown to be unreliable (Karle, 1970). The three lowest values for the cosines of the triplet invariants included in the sets defined by column 4 are given in column 6. In examining columns 5 and 6, it is seen that, although a Cu sphere of data is maintained for the calculations, the amount of reliable information decreases dramatically with complexity. Comparison of the second example with the first shows the benefit that derived from having much more data than can normally be measured. As noted above, experience with the conditional probability distribution (Karle, 1980) has shown that this alternative probabilistic form does not overcome the problem of deteriorating reliability with complexity shown by formula (2).

The question arises concerning whether the triplet form of formula (2) should ever be considered if the unit cell contains approximately 100 or more atoms in the unit cell. The answer is that this formula may provide useful information in some circumstances when there are subsets of the data that have unusually large normalized structure factor magnitudes. It will also be noted later on that worthwhile information may be obtained for phase seminvariants on occasion, particularly when use can be made of embeddings. Although calculations of this sort have been made as long as twenty years ago, there is a lack of broad experience on which to draw conclusions.

Quartet invariants

Calculations of a sampling of quartet invariants were also made from the same data sets that were employed for the triplet invariants. Before discussing the details of the calculations, some remarks concerning the calculation of formulas (1) and (2) for quartet invariants are appropriate. In general, the number of contributing terms to the averages in (1) is $n!$, where n is the order of the n-tet. In this case, we have 24 terms. There is, however, a duplication of terms, so that there are 12 distinct ones. The latter are obtained by including only terms in the averages of (1) that satisfy $i < l$, where l is defined by $h_i + h_j + h_k = -h_l$.

Table 1. *Calculations of the cosines of triplet phase invariants*

Results of calculations of formula (2) (with $S_{k,3} = 1$) for triplet invariants for random structures of increasing complexity in space group $P1$. In each case, 150 invariants associated with the largest values for $|E_{h_1} E_{h_2} E_{h_3}|$, where $h_1 + h_2 + h_3 = 0$, were computed. The values for b in column 3 signify the minimum value for two of the three factors in each term in the averages of formula (2). The calculations show a decrease in acceptability with an increase in the complexity of the structures.

Number of atoms	Number of independent data	b	Minimum value accepted for (2)	Number of calculations above the minimum	Three lowest values for cosine of triplet
20	1182	1·4	0·50	50	0·72, 0·74, 0·75
			0·40	70	0·59, 0·61, 0·70
			0·30	86	0·59, 0·61, 0·70
20	4850	1·8	0·50	119	0·67, 0·74, 0·76
			0·40	140	0·67, 0·74, 0·76
50	3609	1·6	0·50	26	0·59, 0·70, 0·71
			0·45	36	−0·16, 0·59, 0·70
			0·40	48	−0·16, 0·59, 0·70
100	8124	1·8	0·40	27	0·26, 0·58, 0·81
			0·35	31	0·26, 0·58, 0·67
			0·30	34	0·21, 0·26, 0·58
200	15375	1·8	0·40	5	0·81, 0·88, 0·92
			0·35	13	−0·92, 0·78, 0·81
			0·30	15	−0·92, 0·65, 0·78
			0·25	20	−0·92, 0·32, 0·65

For each $\mathbf{k}$ in formula (2), there are in general $n!$ distinct contributing terms to the averages. In the computations of (2), $\mathbf{k}$ is permitted to range through the complete data set including Friedel mates. Under this circumstance, there are only three distinct sets of values for i, j and k. The subscripts for the four factors in (2) may be selected as follows:

$$
\begin{array}{llll}
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_4 + \mathbf{k} \quad (6) \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{k}.
\end{array}
$$

It is found that enhanced reliability is obtained when computing a quartet invariant from (2) if it is required that at least three of the four factors in the terms contributing to the average be large. Instead of specifying a particular lower limit for the three factors, the averages are taken at three different limits, *i.e.* 1·4, 1·6 and 1·8 for the lowest values of three of the $|E|$ occurring among the four factors forming the terms in (2) for the case of quartet invariants. The limits 1·4, 1·6 and 1·8 are called the $|E|$ limits. At each of the three limits, there will be three different averages given by (6). One of the criteria for the acceptability of the calculation is the internal consistency of the three averages at each of the three limits. An optimal circumstance occurs when all nine averages are large and in agreement. When contradictory signs are obtained for some of the averages, the calculation is rejected.

Results of computations of formula (2) for small samplings of the quartet invariants are shown in Table 2 ($S_{k,4}$ is set equal to 1). The data sets are the same as those used for the triplet phase invariants. As seen in

columns 3, 5 and 7, the samplings cover a range of known values for the cosine of the quartet represented by cos Φ. The number of calculations accepted is given in columns 4, 6 and 8. Except for the calculation represented by the entry in column 4 for the 100 atom case, none of the accepted calculations reversed the true sign of the invariant. Evidently, to avoid the error in the 100 atom case, more stringent acceptance criteria would be required.

The acceptance criteria applied here are that there must be no contradictions among the calculations described by (6), computed at the three different $|E|$ limits of 1·4, 1·6 and 1·8, and, in addition, an acceptance limit must be satisfied, namely the magnitudes of the results of the calculations of (2) at the $|E|$ limits of 1·4, 1·6 and 1·8 are required to equal or exceed the acceptance limit. Such limiting values are specified in column 9 of Table 2 and have been selected here for illustrative purposes. Some of them may very well not be safe, particularly values as low as 0·30. Therefore, in some of the cases, higher values may be required to ensure that errors in selection would be avoided. Experience should facilitate the establishment of reliable acceptance limits.

Table 2 indicates that the general formula (2) applied to quartet invariants for structures having random arrangements of atoms gives a rapidly decreasing amount of information as the complexity of the structures increases. This is comparable to the results for triplet invariants given in Table 1. In all cases except that given by the second row in Table 2, the quartet invariants were samples from those associated with the largest values for $|E_{h_1}E_{h_2}E_{h_3}E_{h_4}|$, where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$. The quartets computed for the second row of

Table 2. *Calculations of the cosines of quartet phase invariants*

Results of calculations of formula (2) (with $S_{k,4} = 1$) for quartet invariants for the same random structures as used for the calculations in Table 1. In each case, except for the second row, the quartet invariants computed were among those that had the largest values for $|E_{h_1}E_{h_2}E_{h_3}E_{h_4}|$ where $\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 = 0$. The calculations were performed at three different minimum values of three of the four factors in each term in the averages of formula (2), *i.e.* three of the $|E|$ in the factors had to exceed 1·8 or 1·6 or 1·4. Acceptances were based on the internal consistency of the calculations and their equaling or exceeding in value the acceptance limit in column 9. As in Table 1, the calculations show a decrease in acceptability with an increase in the complexity of the structures.

Number of atoms	Number of independent data	Number of quartets computed (cos $\Phi > 0.70$)	Number from column 3 accepted	Number of quartets computed ($-0.40 <$ cos $\Phi < 0.40$)	Number from column 5 accepted	Number of quartets computed (cos $\Phi > -0.65$)	Number from column 7 accepted	Acceptance limit
20	1182	12	12	4	1*	4	1	0·70
20†	4850	8	4			9	6	0·40
50	3609	15	2	2	0	2	0	0·60
100	8124	14	1‡	7	0	1	0	0·50
200	15375	16	0	5	0	2	0	0·30

* Accepted as if it were a large negative value (actual value −0·33).

† Direct comparison with the first row is not appropriate since quartets with much smaller values for $|E_{h_1}E_{h_2}E_{h_3}E_{h_4}|$ were computed here.

‡ Incorrect, since strong negative value indicated.

Table 2 had, on the average, much lower values for the products of the four $|E|$ than the quartets computed for the first row. Despite the fact that the acceptance limit was lowered (except for row 2) as the complexity of the structure increased, it is clearly seen from Table 2 that the number of acceptances decreased greatly. The number of reliable values for the quartets that could be computed has not been determined. That would require a rather extensive computation. Also undetermined is how the number of reliable values might vary when data for actual structures containing varying degrees and types of structural regularity are used for the computations.

The question arises concerning what value to give an invariant for which an acceptable value has been calculated. It could be assigned the magnitude of unity with the computed sign attached, or, perhaps, a worthwhile improvement in accuracy would derive from assigning a magnitude somewhat less than unity, based on the magnitude of $|E_{h_1} E_{h_2} E_{h_3} E_{h_4}|$. Depending upon the application, values of ± 1 may be the more convenient to use.

The same quartets that were computed for rows 1 and 2 in Table 2 for 20 atoms were computed by use of the special formula (1) in which S_4 was set equal to unity. In formula (1) for quartets there are, at most, three different factors of the type $(|E_{h_i + h_j}|^2 - 1)$ and an acceptance criterion can be based on their value. If it is required that all three such factors have $|E| > 1 \cdot 3$ and two have $|E| > 1 \cdot 8$, two of the 12 quartets having large positive cosines are so acceptable. To accept a quartet as having a cosine of large negative value, it is required that the three factors $(|E_{h_i + h_j}|^2 - 1)$ have $|E| < 0 \cdot 7$ and two have $|E| < 0 \cdot 5$. None of the quartets having cosines of large negative value that occur in the first row of Table 2 have calculations with formula (1) that are acceptable. In fact, even the single requirement that two factors have $|E| < 0 \cdot 5$ is not satisfied. The only acceptances that could be made were those for cosines of large positive value. So far as the quartets used in the calculations of row 2 in Table 2 by use of formula (2) are concerned, application of the acceptance criteria to calculations of the same quartets with formula (1) led to the acceptance of only one as a quartet with a large positive value. It is possible to conclude that if the acceptance criteria are reasonable, for the 20 atom structure considered here, the general formula (2) applied to quartets provides more calculations that are acceptable than does the special formula (1). A simplification of special formula (1) can be made in applications where all the $|E|$ values associated with the invariant being computed are large and roughly of the same value. In that case the first and last factors in the terms contributing to the averages in formula (1) can be omitted, since they are always comprised of the $|E|$ associated with the invariant of interest.

Quintet invariants

Some of the characteristics of formulas (1) and (2) applied to the calculation of quintet phase invariants will now be considered, although examples of such calculations will not be presented here. For the special formula (1) applied to quintets, there will be four factors in a term in the average. There should be, if all terms in the average can be formed from the available data set, 60 different terms contributing to the average (since the number of distinct terms is $n!/2$, where $n = 5$ for the quintet). The latter are obtained by including only terms in the averages of (1) that satisfy $i < m$, where m is defined by $\mathbf{h}_i + \mathbf{h}_j + \mathbf{h}_k + \mathbf{h}_l = -\mathbf{h}_m$.

In the application of formula (2) to quintet invariants, there are $n! = 5!$ contributing terms to the average for each $\mathbf{k}$. One way to perform the computations is to permit $\mathbf{k}$ to range through the complete data set including Friedel mates. Under this circumstance, there are only 12 distinct sets of values for i, j, k and l. The subscripts for the five factors in (2) may be selected as follows:

$$
\begin{array}{lllll}
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_5 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_4 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_4 + \mathbf{h}_3 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_4 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_4 + \mathbf{h}_5 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_5 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_5 + \mathbf{h}_3 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_5 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_5 + \mathbf{h}_4 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_2 + \mathbf{h}_4 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_2 + \mathbf{h}_5 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_4 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_4 + \mathbf{h}_2 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_5 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_3 + \mathbf{h}_5 + \mathbf{h}_2 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_4 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_4 + \mathbf{h}_2 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_4 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{k} \\
\mathbf{k} & \mathbf{h}_1 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_4 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_4 + \mathbf{h}_3 + \mathbf{k} & \mathbf{h}_1 + \mathbf{h}_4 + \mathbf{h}_3 + \mathbf{h}_2 + \mathbf{k}.
\end{array}
\tag{7}
$$

It is conjectured that, in general, the number of distinct sets of values for the subscripts $i, j, k, \ldots, v$, when $\mathbf{k}$ is permitted to range through the complete data set is $(n - 1)!/2$, where n is the order of the n-tet.

Embedded seminvariants

Seminvariant phases or sums of phases are associated with planes of reflection that have specific mathematical characteristics for their reciprocal vectors that depend upon the space group and other features such as the trigonometric form of the structure factor. The characteristics of the reciprocal vectors are defined in terms of seminvariant vectors which have been identified for centrosymmetric crystals (Hauptman & Karle, 1953, 1959) and for noncentrosymmetric crystals (Hauptman & Karle, 1956; Karle & Hauptman, 1961b). These results have been summarized (Karle, 1974; see also Lessinger & Wondratschek, 1975 and Hövmoller, 1978).

Invariants can often be composed of the sum of two or more seminvariants. A seminvariant contained within an invariant is called an embedded seminvariant. The virtue of considering embedded seminvariants derives from the fact that there may be the possibility of

computing a value for it in many alternative ways that might enhance the reliability of the determined value. The concept of an embedded seminvariant dates back to the $B_{3,0}$ formula for the computation of triplet phase invariants (Hauptman & Karle, 1958; Karle & Hauptman, 1958), although the specific terminology is more recent (Hauptman, 1978). In fact, embedded seminvariants are contained in the inequalities and the $\sum_1$ and $\sum_3$ formulas (Hauptman & Karle, 1953), for example. The triplet invariants provide simple illustrations of the concepts involved.

A technique suitable for triplet invariants is to choose two phases that are related by space-group symmetry in such a way that, although their individual values may not be known, the sum of their phases would be a known value. These two phases would form a seminvariant and, therefore, the third phase in the triplet invariant, whose indices cancel the sum of the other two, is also a seminvariant. For example, in space group $P2_1$ (*b* axis unique), the sum of two phases given by $\varphi_{hkl} + \varphi_{\bar{h}k\bar{l}}$ is equal to zero or π, depending upon the parities of the indices which are known. The third phase is $\varphi_{2h,0,2l}$, a semivariant phase. The computation of the value of the cosine of the triplet invariant is immediately interpretable in terms of the value of $\cos \varphi_{2h,0,2l}$ since the value of the sum of the other two phases is zero or π. A matter of particular interest is the fact that for a specific $\varphi_{2h,0,2l}$ there are several possible combinations of $\varphi_{\bar{h}k\bar{l}} + \varphi_{\bar{h}\bar{k}\bar{l}}$ since the value of *k* can vary. The details for applying the formula for triplet invariants to the calculation of seminvariant phases in all centrosymmetric space groups has been published in a series of papers in *Acta Crystallographica* [starting with Karle & Hauptman (1959), and concluding with Karle & Hauptman (1961a)]. Similar relations are readily determined for noncentrosymmetric space groups. In general, an invariant is decomposed into desired seminvariants in such a way that information is available for one part and thus the evaluation of the invariant permits the evaluation of the remaining seminvariant part. Formulas (1) and (2) can be used generally for such purposes in all the space groups. Examples of recent theoretical investigations of embedded seminvariants in the context of conditional distributions may be seen for the three-phase structure seminvariant in $P\bar{1}$ (Hauptman, 1980) and for one-phase structure seminvariants in space groups up to orthorhombic (Giacovazzo, 1980b).

Concluding remarks

Expected value formulas for phase invariants and embedded seminvariants of all orders and for all space groups have been obtained from use of determinantal probability distributions. It has been seen in rather limited calculations with triplet and quartet invariants

that the reliability of the calculations decreases rapidly with complexity. The question arises concerning practical expectations for the structures that are now regarded as difficult. These usually occur among noncentrosymmetric, essentially equal-atom structures with 100 or more atoms in the asymmetric unit. The suggestion has already been made by workers with conditional joint probability distributions that thousands of computations should be made and only a small number of the most reliable ones be accepted. How extensive must the calculations be? How much information would be so gleaned and how useful would it be? In comparison with current calculations in structure determination, they would probably be some orders of magnitude more extensive. How much information would thus become available remains to be seen.

Focusing the calculations on the embedded seminvariants has been shown to have advantages because there are often many different ways to evaluate the same invariant and also at times the need arises only to distinguish between a limited number of quite different values such as zero or π.

It has been pointed out, in connection with formulas (1) and (2), that the number of contributing terms to the formulas increases factorially with the order of the invariant. This raises the question for the future of whether, in a finite data set, it is possible to provide such an increase in contributors or whether most of the increase is lost because needed factors occur beyond the measured range. There is also a great increase in the number of phase invariants, as the order of the invariant increases, whose cosines can be calculated. This also plays a role in extending the magnitude of the calculations.

The calculations in this paper were based on atomic arrangements whose coordinates were found from a random number generator. It is possible that formulas (1) and (2) might show a great variation in the nature of the results with actual crystal data depending upon the regularities in the structures and the possibility of having special subsets of normalized structure factors of large magnitude.

No general study has been made comparing the usefulness of the expected value formulas (1) and (2) in this paper with corresponding conditional probability formulas. It was pointed out that, for the quartet, quite comparable results would be obtained from both types of formula, the expected value and the conditional joint probability distribution. It was also pointed out that the same terms as occur in the conditional distributions that have been published for some of the more elementary space groups also occur in the expected value formulas, mainly the special formula. The conditional joint probability distributions could be obtained in a step in the derivation that precedes the obtaining of the expected value formulas. If it should

turn out that the conditional distributions have some advantages, they could be derived from the determinantal joint probability distributions. However, they would be rather more complicated than the expected value formulas of this paper and, because of this, I believe that the expected value formulas might well be tried first.

My thanks to Stephen A. Brenner who developed the programs and made the calculations for this paper.

References

FLIPPEN, J. L. (1972). Acta Cryst. B28, 3618–3624.
GIACOVAZZO, C. (1977). Acta Cryst. A33, 933–944.
GIACOVAZZO, C. (1980a). Acta Cryst. A36, 704–711.
GIACOVAZZO, C. (1980b). Acta Cryst. A36, 711–715.
HAUPTMAN, H. (1964). Acta Cryst. 17, 1421–1433.
HAUPTMAN, H. (1977a). Acta Cryst. A33, 553–555.
HAUPTMAN, H. (1977b). Acta Cryst. A33, 556–564.
HAUPTMAN, H. (1978). Acta Cryst. A34, 525–528.
HAUPTMAN, H. (1980). Acta Cryst. A36, 624–632.
HAUPTMAN, H. & KARLE, J. (1953). Am. Crystallogr. Assoc., Monogr. No. 3. Pittsburgh: Polycrystal Book Service.
HAUPTMAN, H. & KARLE, J. (1956). Acta Cryst. 9, 45–55.
HAUPTMAN, H. & KARLE, J. (1958). Acta Cryst. 11, 149–157.
HAUPTMAN, H. & KARLE, J. (1959). Acta Cryst. 12, 93–97.
HOVMÖLLER, S. (1978). Acta Cryst. A34, 826–827.
KARLE, I. L., GIBSON, J. W. & KARLE, J. (1970). J. Am. Chem. Soc. 92, 3755–3760.
KARLE, I. L., HAUPTMAN, H., KARLE, J. & WING, A. B. (1958). Acta Cryst. 11, 257–263.
KARLE, J. (1966). Acta Cryst. 20, 881–886.
KARLE, J. (1970). Acta Cryst. B26, 1614–1617.
KARLE, J. (1974). In International Tables for X-ray Crystallography, Vol. IV. Birmingham: Kynoch Press.
KARLE, J. (1980). Proc. Natl Acad. Sci. USA, 77, 5–9,
KARLE, J. (1982a). Proc. Natl Acad. Sci. USA, 79, 1337–1339.
KARLE, J. (1982b). Proc. Natl Acad. Sci. USA, 79. 2125–2127.
KARLE, J. & HAUPTMAN, H. (1958). Acta Cryst. 11, 264–269.
KARLE, J. & HAUPTMAN, H. (1959). Acta Cryst. 12, 404–410.
KARLE, J. & HAUPTMAN, H. (1961a). Acta Cryst. 14, 105–110.
KARLE, J. & HAUPTMAN, H. (1961b). Acta Cryst. 14, 217–223.
LESSINGER, L. & WONDRATSCHEK, H. (1975). Acta Cryst. A31, 521.

X-ray Diffraction Studies on the Arrangement of Atoms in Carbon Blacks. I. Results of Radial Distribution Analysis

BY S. K. SANYAL, B. K. SAMANTARAY AND G. B. MITRA

Indian Institute of Technology, Kharagpur – 721302, India

Abstract

Radial distribution analysis studies by X-ray diffraction of carbon blacks obtained by pyrolysis of benzene C_6H_6, toluene, C_7H_8, and sucrose, $C_{12}H_{22}O_{11}$, and samples of ultracarbon of the medicinal variety have been carried out. Bonding distances, coordination numbers and the layer dimensions have been determined. It has been observed that the carbon blacks have turbostratic structures. The C—C distances in the carbon layers lie in the range of that for aromatic rings and double bonds. Ultracarbon has the C—C bonding distance corresponding to graphite.

1. Introduction

Because of their increasing uses, various allotropic modifications of carbon are being investigated more and more regarding their structure–property relations. Atomic arrangements in carbon in its various forms – especially the amorphous forms – are becoming more and more fascinating to study.

X-ray diffraction patterns of all types of amorphous carbons have one feature in common: they all have two broad peaks in the approximate neighbourhood of diffraction angles corresponding to 002 and 100 peaks of graphite. This led early workers like Asahara (1922), Mahadevan (1929), Blayden, Gibson & Riley (1944), Siever (1952), Mitra (1953), Nelson (1954) and Hirsch (1954) to assume that amorphous carbon consisted of very small particles of graphite so disordered that only the 002 and 100 reflections remained intact while other peaks cancelled out. Some additional peaks were observed by some of these workers, but the model essentially remained the same.