DIRECT METHODS AND ANOMALOUS DISPERSION

by

HERBERT A. HAUPTMAN

Medical Foundation of Buffalo, Inc. 73 High Street, BUFFALO, N. Y. 14 203

1. INTRODUCTION

The electron density function, $\varrho(\mathbf{r})$, in a crystal determines its diffraction pattern, i.e. both the magnitudes and phases of its x-ray diffraction maxima, and conversely. If, however, as is always the case, only magnitudes are available from the diffraction experiment, then the density function $\rho(\mathbf{r})$ cannot be recovered. If one invokes prior structural knowledge, usually that the crystal is composed of discrete atoms of known atomic numbers, then the observed magnitudes are, in general, sufficient to determine the positions of the atoms, i.e. the crystal structure.

It should be noted here that the recognition that observed diffraction data are in general sufficient to determine crystal structures uniquely was an important milestone in the development of the direct methods of crystal structure determination. The erroneous contrary view, that crystal structures could not, even in principle, be deduced from diffraction intensities, had long been held by the crystallographic community prior to c. 1950 and constituted a psychological barrier which first had to be removed before real progress could be made.

2. THE TRADITIONAL DIRECT METHODS

2.1. The phase problem.

Denote by $\phi_{\mathbf{H}}$ the phase of the structure factor $F_{\mathbf{H}}$:

$$F_{\mathbf{H}} = |F_{\mathbf{H}}| \exp(i\phi_{\mathbf{H}}), \tag{1}$$

where $\mathbf{H}$ is a reciprocal lattice vector (having three integer components) which labels the corresponding diffraction maximum. Then the relationship between the structure factors $F_{\mathbf{H}}$ and the electron density function $\varrho(\mathbf{r})$ is given by

$$F_{\mathbf{H}} = \int_{V} \rho(\mathbf{r}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}) dV \tag{2}$$

and

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} F_{\mathbf{H}} \exp(-2\pi i \mathbf{H} \cdot \mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} |F_{\mathbf{H}}| \exp i(\phi_{\mathbf{H}} - 2\pi \mathbf{H} \cdot \mathbf{r}) \tag{3}$$

in which V represents the unit cell or its volume. Thus the structure factors F_H determine $\rho(\mathbf{r})$. The x-ray diffraction experiment yields only the magnitudes $|F_H|$ of a finite number of structure factors, but the values of the phases ϕ_H, which are also needed if one is to determine $\rho(\mathbf{r})$ from (3), cannot be determined experimentally. If arbitrary values for the phases ϕ_H are specified in Eq. (3), then density functions $\varrho(\mathbf{r})$ are defined which, when substituted into (2) yield structure factors F_H the magnitudes of which agree with the observed magnitudes $|F_H|$. It follows that the diffraction experiment does not determine $\rho(\mathbf{r})$. It was this argument which led crystallographers, prior to 1950, to the erroneous conclusion that diffraction intensities could not, even in principle, determine crystal structures uniquely. What had been overlooked was the fact that the phases ϕ_H could not be arbitrarily specified if (3) is to yield density functions characteristic of real crystals.

   Crystals are composed of discrete atoms. One exploits this prior structural knowledge by replacing the real crystal, with continuous electron density $\rho(\mathbf{r})$, by an ideal one, the unit cell of which consists of N discrete, non-vibrating, point atoms located at the maxima of $\rho(\mathbf{r})$. Then the structure factor F_H is replaced by the normalized structure factor E_H and (1) to (3) are replaced by

$$E_H = |E_H| \exp(i\phi_H), \tag{4}$$

$$E_H = \frac{1}{\sigma_2^{1/2}} \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{H}\cdot\mathbf{r}_j), \tag{5}$$

$$\left\langle E_H \exp(-2\pi i \mathbf{H}\cdot\mathbf{r}) \right\rangle_H = \frac{1}{\sigma_2^{1/2}} \left\langle \sum_{j=1}^{N} f_j \exp\left[2\pi i \mathbf{H}\cdot(\mathbf{r}_j - \mathbf{r})\right] \right\rangle_H$$

$$= \frac{f_j}{\sigma_2^{1/2}} \text{ if } \mathbf{r} = \mathbf{r}_j$$

$$= 0 \text{ if } \mathbf{r} \neq \mathbf{r}_j \tag{6}$$

where f_j is the zero-angle atomic scattering factor, $\mathbf{r}_j$ is the position vector of the atom labelled j, and

$$\sigma_n = \sum_{j=1}^{N} f_j^n, \ n = 1, 2, 3, \ldots \tag{7}$$

In the x-ray diffraction case the f_j are equal to the atomic numbers Z_j and are presumed to be known. From (6) it follows that the normalized structure factors E_H determine the atomic position vectors $\mathbf{r}_j$, j = 1,2,..., N, i.e. the crystal structure.

   In practice a finite number of magnitudes $|E_H|$ of normalized structure factors E_H are obtainable (at least approximately) from the observed magnitudes $|F_H|$ while the phases ϕ_H, as defined by (4) and (5), cannot be determined experimentally. Since one now requires only the 3N components of the N position vectors $\mathbf{r}$, rather than the much more complicated electron density

function $\varrho(\mathbf{r})$, it turns out that, in general, the known magnitudes are more than sufficient. This is most readily seen by equating the magnitudes of both sides of (5) in order to obtain a system of equations in which the only unknowns are the 3N components of the position vectors $\mathbf{r_j}$. Since the number of such equations, equal to the number of reciprocal lattice vectors $\mathbf{H}$ for which magnitudes $|E_\mathbf{H}|$ are available, usually greatly exceeds the number, 3N, of unknowns, this system is redundant. Thus observed diffraction intensities usually over-determine the crystal structure, i.e. the positions of the atoms in the unit cell. In short, by merely replacing the integral of Eq. (2) by the summation of Eq. (5), i.e. taking Eq. (5) as the starting point of our investigation rather than Eq. (2), one has transformed the problem from an unsolvable one to one which is solvable, at least in principle.

In summary then, the intensities (or magnitudes $|E_\mathbf{H}|$) of a sufficient number of x-ray diffraction maxima determine a crystal structure. The available intensities usually exceed the number of parameters needed to describe the structure. From these intensities a set of numbers $|E_\mathbf{H}|$ can be derived, one corresponding to each intensity. However, the elucidation of the crystal structure requires also a knowledge of the complex numbers $E_H = |E_\mathbf{H}| \exp(i\phi_\mathbf{H})$, the normalized structure factors, of which only the magnitudes $|E_\mathbf{H}|$ can be determined from experiment. Thus a "phase" $\phi_\mathbf{H}$, unobtainable from the diffraction experiment, must be assigned to each $|E_\mathbf{H}|$, and the problem of determining the phases when only the magnitudes $|E_\mathbf{H}|$ are known is called the "phase problem". Owing to the known atomicity of crystal structures and the redundancy of observed magnitudes $|E_\mathbf{H}|$, the phase problem is solvable in principle.

2.2. The structure invariants

Equation (6) implies that the normalized structure factors E_H determine the crystal structure. However (5) does not imply that, conversely, the crystal structure determines the values of the normalized structure factors E_H since the position vectors $\mathbf{r_j}$ depend not only on the structure but on the choice of origin as well. It turns out nevertheless that the magnitudes $|F_\mathbf{H}|$ of the normalized structure factors are in fact uniquely determined by the crystal structure and are independent of the choice of origin but that the values of the phases $\phi_\mathbf{H}$ depend also on the choice of origin. Although the values of the individual phases depend on the structure and the choice of origin, there exist certain linear combinations of the phases, the so-called structure invariants, whose values are determined by the structure alone and are independent of the choice of origin.

It follows readily from Eq. (5) that the linear combination of three phases

$$\psi_3 = \phi_\mathbf{H} + \phi_\mathbf{K} + \phi_\mathbf{L} \tag{8}$$

is a structure invariant (triplet) provided that

$$\mathbf{H} + \mathbf{K} + \mathbf{L} = 0; \tag{9}$$

the linear combination of four phases

$$\psi_4 = \phi_H + \phi_K + \phi_L + \phi_M \tag{10}$$

is a structure invariant (quartet) provided that

$$H + K + L + M = 0 ; \tag{11}$$

etc.

2.3. The structure seminvariants

If a crystal possesses elements of symmetry then the origin may not be chosen arbitrarily if the simplifications permitted by the space group symmetries are to be realized. For example, if a crystal has a centre of symmetry it is natural to place the origin at such a centre while if a two-fold screw axis, but no other symmetry element is present, the origin would normally be situated on this symmetry axis. In such cases the permissible origins are greatly restricted and it is therefore plausible to assume that many linear combinations of the phases will remain unchanged in value when the origin is shifted only in the restricted ways allowed by the space group symmetries. One is thus led to the notion of the structure seminvariant, those linear combinations of the phases whose values are independent of the choice of permissible origin.

If the only symmetry element is a centre of symmetry, for example (space group $P\bar{1}$), then it turns out (again from Eq. (5)) that a single phase ϕ_H is a structure seminvariant provided that the three components of the reciprocal lattice vector H are even integers; the linear combination of two phases $\phi_H + \phi_K$ is a structure seminvariant provided that the three components of $H + K$ are even integers; etc.

If the only symmetry element is a two-fold rotation axis (or twofold screw axis) then one finds from Eq. (5) that the single phase ϕ_{hkl} is a structure seminvariant provided that h and 1 are even integers and k = 0; the linear combination of two phases

$$\phi_{h_1k_1l_1} + \phi_{h_2k_2l_2}$$

is a structure seminvariant provided that $h_1 + h_2$ and $l_1 + l_2$ are even and $k_1 + k_2 = 0$; etc.

The structure invariants and seminvariants have been tabulated for all the space groups (Hauptman and Karle 1953, 1956, 1959; Karle and Hauptman 1961; Lessinger and Wondratschek 1975). In general the collection of structure invariants is a subset of the collection of structure seminvariants. If no element of symmetry is present, that is the space group is Pl, then the two classes coincide.

2.3.1. Origin and enantiomorph specification

The theory of the structure seminvariants leads in a natural way to space group dependent recipes for origin and enantiomorph (i.e. the handedness, right or left) specification.

In general the theory identifies an appropriate set of phases whose values are to be specified in order to fix the origin uniquely. For example, in space group Pl (no elements of symmetry) the values of any three phases

$$\phi_{h_1k_1l_1}, \ \phi_{h_2k_2l_2}, \ \phi_{h_3k_3l_3}, \tag{12}$$

for which the determinant $\triangle$ satisfies

$$\Delta = \begin{vmatrix} h_1k_1l_1 \\ h_2k_2l_2 \\ h_3k_3l_3 \end{vmatrix} = \pm 1, \tag{13}$$

may be specified arbitrarily, thus fixing the origin uniquely. Once this is done then the value of any other phase is uniquely determined by the structure alone. For enantiomorph specification it is sufficient to specify arbitrarily the sign of any enantiomorph sensitive structure invariant, i.e. one whose value is different from 0 or π. (See Hauptman 1972, pages 28-52, for further details.)

In the space group PI one again specifies arbitrarily the value (0 or π) of three phases (12), but now the condition is that the determinant Δ [defined by (13)] be odd. Similar recipes for all the space groups are now known and are to be found in the literature cited.

2.4. *The fundamental principle of direct methods*

It is known that the values of a sufficiently extensive set of cosine seminvariants (the cosines of the structure seminvariants) lead unambiguously to the values of the individual phases (Hauptman 1972). Magnitudes $|E|$ are capable of yielding estimates of the cosine seminvariants only or, equivalently, the magnitudes of the structure seminvariants; the signs of the structure seminvariants are ambiguous because the two enantiomorphous structures permitted by the observed magnitudes $|E|$ correspond to two values of each structure seminvariant differing only in sign. However, once the enantiomorph has been selected by specifying arbitrarily the sign of a particular enantiomorph sensitive structure seminvariant (i.e. one different from 0 or π), then the magnitudes $|E|$ determine both signs and magnitudes of the structure seminvariants consistent with the chosen enantiomorph. Thus, for fixed enantiomorph, the observed magnitudes $|E|$ determine unique values for the structure seminvariants; the latter, in turn, lead to unique values of the individual phases. In short, the structure seminvariants serve to link the observed magnitudes $|E|$ with the desired phases ϕ (the fundamental principle of direct methods). It is this property of the structure seminvariants which accounts for their importance and which justifies the stress placed on them here.

By the term "direct methods" is meant that class of methods which exploits relationships among the structure factors in order to go directly from the observed magnitudes $|E|$ to the needed phases ϕ.

2.5. *The neighborhood principle*

It has long been known that, for fixed enantiomorph, the value of any structure seminvariant ψ is, in general, uniquely determined by the magnitudes $|E|$ of the

normalized structure factors. In recent years it has become clear that, for fixed enantiomorph, there corresponds to ψ one or more small sets of magnitudes $|E|$, the neighborhoods of ψ, on which, in favorable cases, the value of ψ most sensitively depends; that is to say that, in favorable cases, ψ is primarily determined by the values of $|E|$ in any of its neighborhoods and is relatively insensitive to the values of the great bulk of remaining magnitudes. The conditional probability distribution of ψ, assuming as known the magnitudes $|E|$ in any of its neighborhoods, yields an estimate for ψ which is particularly good in the favorable case that the variance of the distribution happens to be small [the neighborhood principle (Hauptman, 1975a,b)].

The study of appropriate probability distributions (compare § 2.7) leads directly to the definition of the neighborhoods of the structure invariants. Definitions are given here only for the triplet ψ_3 and the quartet ψ_4, but recipes for defining the neighborhoods of all the structure invariants are now known (Hauptman 1977a,b, Fortier & Hauptman, 1977).

2.5.1. The first neighborhood of the triplet ψ_3
Let **H**, **K**, **L** be three reciprocal lattice vectors which satisfy Eq. (9). Then ψ_3, Eq. (8), is a structure invariant and its first neighborhood is defined to consist of the three magnitudes:

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{\mathbf{L}}|. \tag{14}$$

2.5.2. Neighborhoods of the quartet ψ_4
2.5.2.1. The first neighborhood
Let **H**, **K**, **L**, **M** be four reciprocal lattice vectors which satisfy Eq. (11). Then ψ_4, Eq. (10), is a structure invariant and its first neighborhood is defined to consist of the four magnitudes:

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |E_{\mathbf{L}}|, |E_{\mathbf{M}}|. \tag{15}$$

The four magnitudes (15) are said to be the main terms of the quartet ψ_4.

2.5.2.2. The second neighborhood
The second neighborhood of the quartet ψ_4 is defined to consist of the four magnitudes (15) plus the three additional magnitudes:

$$|E_{\mathbf{H+K}}|, |E_{\mathbf{K+L}}|, |E_{\mathbf{L+H}}|, \tag{16}$$

i.e. seven magnitudes $|E|$ in all. The three magnitudes (16) are said to be the cross-terms of the quartet ψ_4.

2.6. The extension concept
By embedding the structure seminvariant **T** and its symmetry related variants in suitable structure invariants Q one obtains the extensions Q of the seminvariant T. Owing to the space group dependent relations among the phases, T is related in a known way to its extensions. In this way the theory of the structure

seminvariants is reduced to that of the structure invariants. In particular, the neighborhoods of T are defined in terms of the neighborhoods of its extensions. The procedure will be illustrated in some detail only for the two-phase structure seminvariant in the space group PI which serves as the prototype for the structure seminvariants in general, in all space groups, noncentrosymmetric as well as centrosymmetric.

2.6.1. The two-phase structure seminvariant in PI

It has already been seen (§ 2.3) that the linear combination of two phases

$$T = \phi_{\mathbf{H}} + \phi_{\mathbf{K}} \tag{17}$$

is a structure seminvariant in $P\bar{I}$ if and only if the three components of the reciprocal lattice vector $\mathbf{H} + \mathbf{K}$ are all even. Then the components of each of the four reciprocal lattice vectors $\frac{1}{2} (\pm \mathbf{H} \pm \mathbf{K})$ are all integers. Note also that in this space group the structure factors are real and all phases are O or π.

2.6.2. The extensions of T

One embeds the two-phase structure seminvariant T (17) and its symmetry related variant

$$T_1 = \phi_{-\mathbf{H}} + \phi_{\mathbf{K}} \tag{18}$$

in the respective quartets

$$Q = T + \phi_{-\frac{1}{2} (\mathbf{H} + \mathbf{K})} + \phi_{-\frac{1}{2} (\mathbf{H} + \mathbf{K})} , \tag{19}$$

$$Q_1 = T_1 + \phi_{-\frac{1}{2} (-\mathbf{H} + \mathbf{K})} + \phi_{-\frac{1}{2} (-\mathbf{H} + \mathbf{K})} \tag{20}$$

In view of (17) and (18) and the space group-dependent relationships among the phases it is readily verified that Q and Q_1 are in fact (special) four-phase structure invariants (quartets) and

$$T = T_1 = Q = Q_1. \tag{21}$$

The quartets Q and Q_1 are said to be the extensions of the seminvariant T. In this way the theory of the two-phase structure seminvariant T is reduced to that of the quartets. In particular, the neighborhoods of T are defined in terms of the neighborhoods of the quartet.

2.6.3. The first neighborhoods of the extensions

Since two of the phases of the quartet Q (19) are identical, only three of the four main terms are distinct. The first neighborhood of Q is accordingly defined to consist of the three magnitudes

$$\left[\text{since }\left|E_{-\frac{1}{2}(\mathbf{H} + \mathbf{K})}\right| = \left|E_{\frac{1}{2}(\mathbf{H} + \mathbf{K})}\right|\right]:$$

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, \left|E_{\frac{1}{2}(\mathbf{H} + \mathbf{K})}\right|. \tag{22}$$

In a similar way the first neighborhood of the extension Q_1, (20), is defined to consist of the three magnitudes

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, \left|E_{\frac{1}{2}(\mathbf{H} - \mathbf{K})}\right|. \tag{23}$$

2.6.4. The first neighborhood of T

The first neighborhood of the two-phase structure seminvariant T is defined to consist of the set-theoretic union of the first neighborhoods of its extensions, i.e., in view of (22) and (23), of the four magnitudes

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, \left|E_{\frac{1}{2}(\mathbf{H} + \mathbf{K})}\right|, \left|E_{\frac{1}{2}(\mathbf{H} - \mathbf{K})}\right|. \tag{24}$$

2.7. The solution strategy

One starts with the system of equations (5). By equating real and imaginary parts of (5) one obtains two equations for each reciprocal lattice vector H. The magnitudes $|E_H|$ and the atomic scattering factors f_j are presumed to be known. The unknowns are the atomic position vectors r_j and the phases $\phi_{\mathbf{H}}$. Owing to the redundancy of the system (5), one naturally invokes probabilistic techniques in order to eliminate the unknown position vectors r_j, and in this way to obtain relationships among the unknown phases $\phi_{\mathbf{H}}$ having probabilistic validity.

Choose a finite number of reciprocal lattice vectors $\mathbf{H}$, $\mathbf{K}$,... in such a way that the linear combination of phases

$$\psi = \phi_{\mathbf{H}} + \phi_{\mathbf{K}} + \dots \tag{25}$$

is a structure invariant or seminvariant whose value we wish to estimate. Choose satellite reciprocal lattice vectors H′, K′, . . . in such a way that the collection of magnitudes

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|,\dots; |E_{\mathbf{H}}'|, |E_{\mathbf{K}}'|,\dots \tag{26}$$

constitutes a neighborhood of ψ. The atomic position vectors r_j are assumed to be the primitive random variables which are uniformly and independently distributed. Then the magnitudes $|E_H|$, $|E_K|$,...; $|E_H'|$, $|E_K'|$,...; and phases $\phi_{\mathbf{H}}$, $\phi_{\mathbf{K}}$,...; $\phi_{\mathbf{H}}'$, $\phi_{\mathbf{K}}'$,... of the complex, normalized structure factors E_H, E_K,...; E_H' E_K',..., as functions [(Eq. (5)] of the position vectors r_j, are themselves random variables, and their joint probability distribution P may be obtained. From the distribution P one derives the conditional joint probability distribution

$$P\ (\Phi_{\mathbf{H}},\ \Phi_{\mathbf{K}},...\ |\ |E_{\mathbf{H}}|,\ |E_{\mathbf{K}}|,...;\ |E_{\mathbf{H'}}|,\ |E_{\mathbf{K'}}|,...), \tag{27}$$

of the phases $\phi_{\mathbf{H}}, \phi_{\mathbf{K}},...,$ given the magnitudes $|E_H|,\ |E_K|,...;\ |E_{H'}|,\ |E_K|,...,$ by fixing the known magnitudes, integrating with respect to the unknown phases $\phi_{\mathbf{H}}', \phi_{\mathbf{K}}',...$ from O to 2π, and multiplying by a suitable normalizing parameter. The distribution (27) in turn leads directly to the conditional probability distribution

$$P(\Psi|\ |E_{\mathbf{H}}|,\ |E_{\mathbf{K}}|,...;\ |E_{\mathbf{H'}}|,\ |E_{\mathbf{K'}}|,...) \tag{28}$$

of the structure invariant or seminvariant ψ assuming as known the magnitudes (26) constituting a neighborhood of ψ. Finally, the distribution (28) yields an estimate for ψ which is particularly good in the favorable case that the variance of (28) happens to be small.

2.8. Estimating the triplet in PI

Let the three reciprocal lattice vectors $\mathbf{H}$, $\mathbf{K}$, and $\mathbf{L}$ satisfy (9). Refer to § 2.5.1 for the first neighborhood of the triplet ψ_3 (Eq. (8)] and to § 2.7 for the probabilistic background.

Suppose that R_1, R_2, and R_3 are three specified non-negative numbers. Denote by

$$P_{1/3} = P(\Psi|R_1, R_2, R_3)$$

the conditional probability distribution of the triplet ψ_3, given the three magnitudes in its first neighborhood:

$$|E_{\mathbf{H}}| = R_1,\ |E_{\mathbf{K}}| = R_2,\ |E_{\mathbf{L}}| = R_3. \tag{29}$$

Then, carrying out the program described in § 2.7, one finds (Cochran, 1955)

$$P_{1/3} = P(\Psi|R_1, R_2, R_3) \approx \frac{1}{2\pi I_0(A)}\ \exp\ (A \cos \Psi) \tag{30}$$

where

$$A = \frac{2\sigma_3}{\sigma_2^{3/2}}\ R_1\ R_2\ R_3, \tag{31}$$

I_O is the modified Bessel function, and σ_n is defined by (7). Since $A>O, P_{1/3}$ has a unique maximum at $\psi = 0$, and it is clear that the larger the value of A the smaller is the variance of the distribution. See Figure 1, where A = 2.316, Figure 2, where A = 0.731. Hence in the favorable case that A is large, say, for example, $A>3$, the distribution leads to a reliable estimate of the structure invariant ψ_3, zero in this case:

$$\psi_3 \approx O \text{ if A is large.} \tag{32}$$

Furthermore, the larger the value of A, the more likely is the probabilistic statement (32). It is remarkable how useful this relationship has proven to be in

the applications; and yet (32) is severely limited because it is capable of yielding only the zero estimate for ψ_3, and only those estimates are reliable for which A is large, the favorable cases.
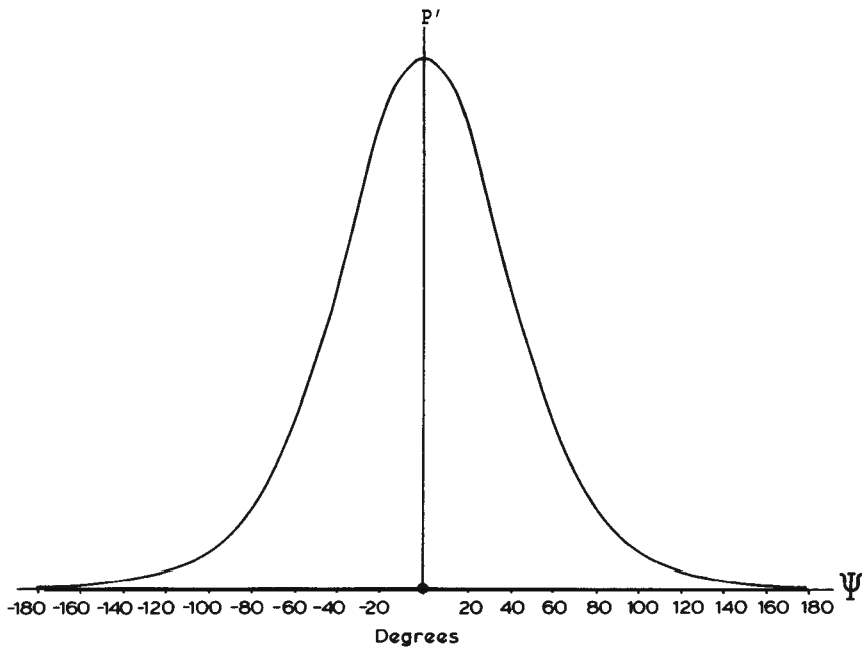

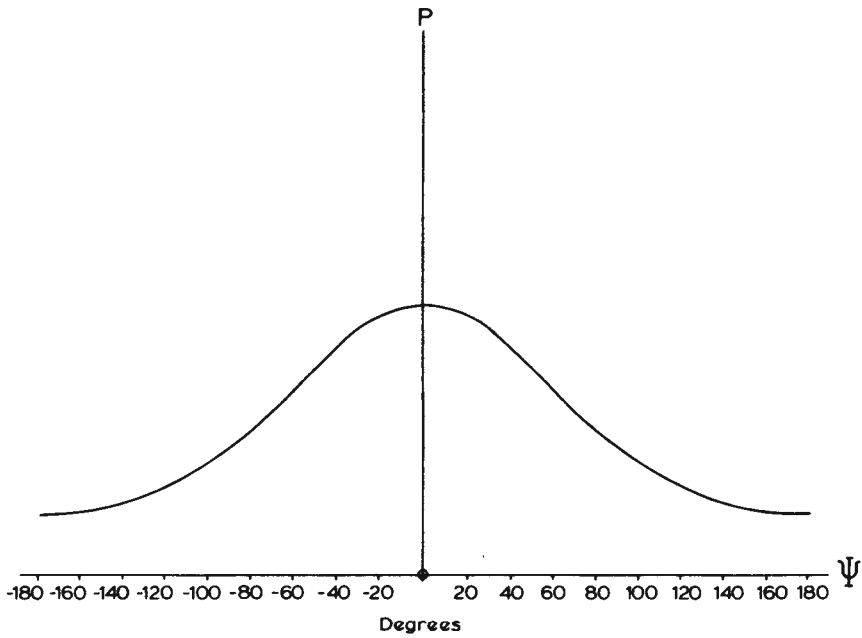*Figure 1.* The distribution $P_{1/3}$, equation (30), for A = 2.316

Figure 2. The distribution $P_{1/3}$, equation (30), for A = 0.731

It should be mentioned in passing that a distribution closely related to (30) leads directly to the so-called tangent formula (Karle & Hauptman, 1956) which is universaily used by direct methods practitioners:

$$\tan \phi_h = \frac{\left\langle |E_K E_{h-K}| \sin (\phi_K + \phi_{h-K}) \right\rangle_K}{\left\langle |E_K E_{h-K}| \cos (\phi_K + \phi_{h-K}) \right\rangle_K} \quad , \tag{33}$$

in which **h** is a fixed reciprocal lattice vector, the averages are taken over the same set of vectors **K** in reciprocal space, usually restricted to those vectors **K** for which $|E_K|$ and $|E_{h-K}|$ are both large, and the sign of sin ϕ_h (cos ϕ_h) is the same as the sign of the numerator (denominator) on the right hand side. The tangent formula is usually used to reline and extend a basis set of phases, presumed to be known.

2.9. Estimating the quartet in PI

Two conditional probability distributions are described, one assuming as known the four magnitudes $|E|$ in the first neighborhood of the quartet, the second assuming as known the seven magnitudes $|E|$ in its second neighborhood.

2.9.1. The first neighborhood

Suppose that **H**, **K**, **L**, and **M** are four reciprocal lattice vectors which satisfy (11). Refer to § 2.5.2.1 for the first neighborhood of the quartet ψ_4(10) and to § 2.7 for the probabilistic background. Suppose that R_1, R_2, R_3, and R_4 are four specified non-negative numbers. Denote by

$$P_{1/4} = P(\Psi | R_1, R_2, R_3, R_4)$$

the conditional probability distribution of the quartet ψ_4, given the four magnitudes in its first neighborhood:

$$|E_H| = R_1, |E_K| = R_2, |E_L| = R_3, |E_M| = R_4. \tag{34}$$

Then

$$P_{1/4} = P(\Psi | R_1, R_2, R_3, R_4) \approx \frac{1}{2\pi I_0(B)} \exp (B \cos \Psi) \tag{35}$$

where

$$B = \frac{2\sigma_4}{\sigma_2^2} R_1 R_2 R_3 R_4, \tag{36}$$

and σ_n is defined by (7). Thus $P_{1/4}$ is identical with P1/3, but B replaces A. Hence similar remarks apply to $P_{1/4}$. In particular, (35) always has a unique maximum at ψ = O so that the most probable value of ψ_4, given the four magnitudes (34) in its first neighborhood, is zero, and the larger the value of B the more likely it is that ψ_4 = O. Since B values, of order $1/N$, tend to be less than A values, of order $1/\sqrt{N}$, at least for large values of N, the estimate (zero)

of ψ_4 is in general less reliable than the estimate (zero) of ψ_3. Hence the goal of obtaining a reliable non-zero estimate for a structure invariant is not realized by (35). The decisive step in this direction is made next.

2.9.2. The second neighborhood

Employ the same notation as in § 2.9.1 but refer now to § 2.5.2.2 for the second neighborhood of the quartet ψ_4. Suppose that R_1, R_2, R_3, R_4, R_{12}, R_{23}, and R_{31}, are seven non-negative numbers. Denote by

$$P_{1/7} = P(\Psi | R_1, R_2, R_3, R_4; R_{12}, R_{23}, R_{31})$$

the conditional probability distribution of the quartet ψ_4, given the seven magnitudes in its second neighborhood:

$$|E_{\mathbf{H}}| = R_1, |E_{\mathbf{K}}| = R_2, |E_{\mathbf{L}}| = R_3, |E_{\mathbf{M}}| = R_4; \tag{37}$$

$$|E_{\mathbf{H+K}}| = R_{12}, |E_{\mathbf{K+L}}| = R_{23}, |E_{\mathbf{L+H}}| = R_{31}. \tag{38}$$

Then (Hauptman, 1975 a, b; 1976)

$$P_{1/7} \approx \frac{1}{L} \exp\left(-2B'\cos\Psi\right) I_0\left[\frac{2\sigma_3}{\sigma_2^{3/2}} R_{12}X_{12}\right] I_0\left[\frac{2\sigma_3}{\sigma_2^{3/2}} R_{23}X_{23}\right] \times$$

$$I_0\left[\frac{2\sigma_3}{\sigma_2^{3/2}} R_{31}X_{31}\right], \tag{39}$$

where

$$B' = \frac{1}{\sigma_2^3} \left(3\sigma_3^2 - \sigma_2\sigma_4\right) R_1R_2R_3R_4, \tag{40}$$

$$X_{12} = [R_1^2R_2^2 + R_3^2R_4^2 + 2R_1R_2R_3R_4\cos\Psi]^{1/2}, \tag{41}$$

$$X_{23} = [R_2^2R_3^2 + R_1^2R_4^2 + 2R_1R_2R_3R_4\cos\Psi]^{1/2}, \tag{42}$$

$$X_{31} = [R_3^2R_1^2 + R_2^2R_4^2 + 2R_1R_2R_3R_4\cos\Psi]^{1/2}, \tag{43}$$

$\sigma_{\mathbf{n}}$ is defined by (7), and L is a normalizing parameter, independent of ψ, which is not needed for the present purpose.

　　Figures 3-5 show the distribution (39) (solid line ▬) for typical values of the seven parameters (37) and (38). For comparison the distribution (35) (broken line ▬▬▬) is also shown. Since the magnitudes |E| have been obtained from a real structure with N = 29, comparison with the true value of the quartet is also possible. As already emphasized, the distribution (35) always has a unique maximum at ψ = O. The distribution (39), on the other hand,

may have a maximium at $\psi = O$, or π, or any value between these extremes, as shown by Figures 3-5. Roughly speaking, the maximum of (39) occurs at 0 or π according as the three parameters R_{12}, R_{23}, R_{31} are all large or all small, respectively. These figures also clearly show the improvement which may result when, in addition to the four magnitudes (37), the three magnitudes (38) are also assumed to be known. Finally, in the special case that

$$R_{12} \approx R_{23} \approx R_{31} \approx O \tag{44}$$

the distribution (39) reduces to

$$P_{1/7} \approx \frac{1}{L} \exp\left(-2B'\cos\Psi\right), \tag{45}$$

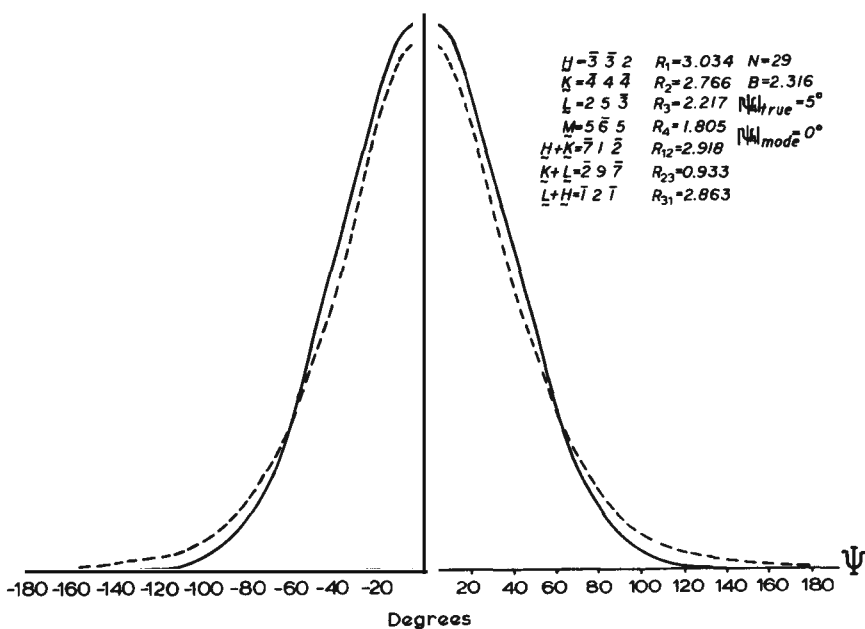|		
$H = \bar{3}\ \bar{3}\ 2$	$R_1 = 3.034$	$N = 29$
$K = \bar{4}\ 4\ \bar{4}$	$R_2 = 2.766$	$B = 2.316$
$L = \bar{2}\ 5\ \bar{3}$	$R_3 = 2.217$	$\Psi_{true} = -5°$
$M = 5\ \bar{6}\ 5$	$R_4 = 1.805$	$\Psi_{mode} = 0°$
$H+K = \bar{7}\ 1\ \bar{2}$	$R_{12} = 2.918$	
$K+L = \bar{2}\ 9\ \bar{7}$	$R_{23} = 0.933$	
$L+H = \bar{1}\ 2\ \bar{1}$	$R_{31} = 2.863$	

Figure 3. The distribution (39) (___) and (35) (---) for the values of the seven parameters (37) and (38) shown. The mode of (39) is O, of (35) always O.
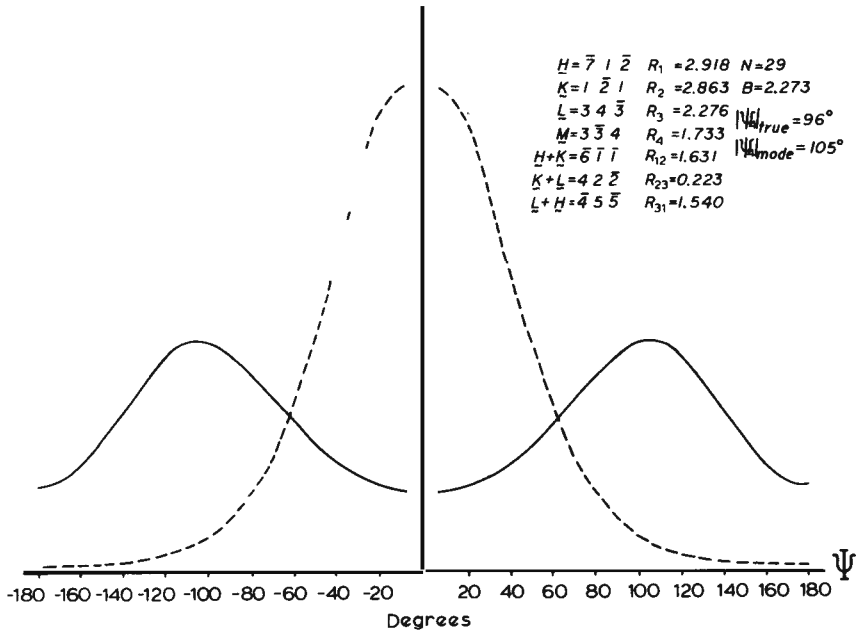
Figure 4. The distribution (39) (__) and (35) (---) for the values of the seven parameters (37) and (38) shown. The mode of (39) is 105°, of (35) always O.
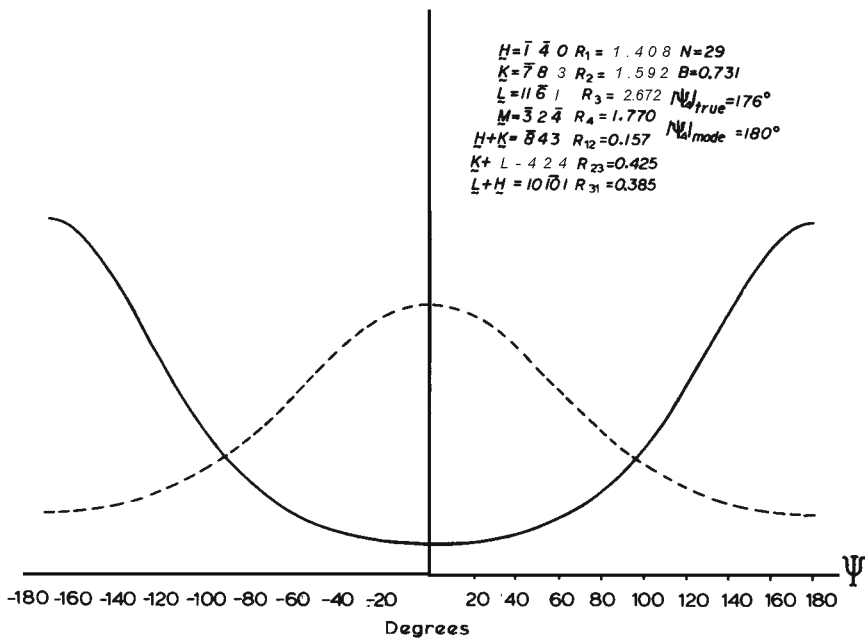

Figure 5. The distribution (39) (__) and (35) (---) for the values of the seven parameters (37) and (38) shown. The mode of (39) is 180°, of (35) always O.

which has a unique maximum at $\psi = \pi$ (Fig. 5).

2.10. Estimating the two-phase structure seminvariant in $P\bar{1}$

Suppose that **H** and **K** are two reciprocal lattice vectors such that the three components of **H** + **K** are even integers. Then the linear combination T of two phases (17) is a structure seminvariant. Refer to § 2.6.4 for the four magnitudes (24) in the first neighborhood of T and to § 2.7 for the probabilistic background. Suppose that R_1, R_2, r_{12}, and $r_{1\bar{2}}$ are four non-negative numbers. In this space group every phase is O or π so that T = 0 or π and the conditional probability distribution of T, assuming as known the four magnitudes in its first neighborhood, is discrete. Denote by $P_+(P_-)$ the conditional probability that T = O (π), given the four magnitudes in its first neighborhood:

$$|E_\mathbf{H}| = R_1, \ |E_\mathbf{K}| = R_2, \ \left|E_{\frac{1}{2}(\mathbf{H}+\mathbf{K})}\right| = r_{12}, \ \left|E_{\frac{1}{2}(\mathbf{H}-\mathbf{K})}\right| = r_{1\bar{2}}. \tag{46}$$

In the special case that all N atoms in the unit cell are identical, the solution strategy described in § 2.7 leads to (Green & Hauptman, 1976)

$$P_\pm \approx \frac{1}{M} \exp\left\{ \mp \frac{R_1 R_2 (r_{12}^2 + r_{1\bar{2}}^2)}{2N} \right\} \cosh\left\{ \frac{r_{12} r_{1\bar{2}} (R_1 \pm R_2)}{N^{1/2}} \right\} \tag{47}$$

where upper (lower) signs go together and

$$M = \exp\left\{ -\frac{R_1 R_2 (r_{12}^2 + r_{1\bar{2}}^2)}{2N} \right\} \cosh\left\{ \frac{r_{12} r_{1\bar{2}} (R_1 + R_2)}{N^{1/2}} \right\} +$$

$$\exp\left\{ +\frac{R_1 R_2 (r_{12}^2 + r_{1\bar{2}}^2)}{2N} \right\} \cosh\left\{ \frac{r_{12} r_{1\bar{2}} (R_1 - R_2)}{N^{1/2}} \right\}. \tag{48}$$

It is easily verified that, under the assumption that R_1 and R_2 are both large, P+>>1/2 if r_{12} and $r_{1\bar{2}}$ are both large, but P+<<1/2 if one of r_{12}, $r_{1\bar{2}}$ is large and the other is small. Hence T $\approx$ O or π respectively (the favorable cases of the neighborhood principle for T).

3. COMBINING DIRECT METHODS WITH ANOMALOUS DISPERSION

3.1. Introduction

The overview of the traditional direct methods described in §§ 1 and 2 is readily generalized to the case that the atomic scattering factors are arbitrary complex-valued functions of $(\sin \theta)/\lambda$, thus including the special case that one or more anomalous scatterers are present. Once again the neighborhood concept plays an essential role. Final results from the probabilistic theory of the two- and three-phase structure invariants are briefly summarized. In particu-

lar, the conditional probability distributions of the two- and three-phase structure invariants, given the magnitudes |E| in their first neighborhoods, are described. The distributions yield estimates for these invariants which are particularly good in those cases that the variances of the distributions happen to be small (the neighborhood principle). It is particularly noteworthy that these estimates are unique in the whole range from $-\pi$ to $+\pi$. An example shows that the method is capable of yielding unique estimates for tens of thousands of structure invariants with unprecedented accuracy, even in the macromolecular case. It thus appears that this fusion of the traditional techniques of direct methods with anomalous dispersion will facilitate the solution of those crystal structures which contain one or more anomalous scatterers.

Most crystal structures containing as many as 80-100 independent nonhydrogen atoms are more or less routinely solvable nowadays by direct methods. On the other hand, it has been known for a long time (Peerdeman & Bijvoet, 1956; Ramachandran & Raman, 1956; Okaya & Pepinsky, 1956) that the presence of one or more anomalous scatterers facilitates the solution of the phase problem; and some recent work (Kroon, Spek & Krabbendam, 1977; Heinerman, Krabbendam, Kroon & Spek, 1978), employing Bijvoet inequalities and the double Patterson function, leads in a similar way to estimates of the sines of the three-phase structure invariants. Again, some early work of Rossmann (1961), employing the difference synthesis $(|F_H| - |F_H|)^*$ in order to locate the anomalous scatterers and recently applied by Hendrickson and Teeter (1981) in their solution of the crambin structure, shows that the presence of anomalous scatterers facilitates the determination of crystal structures. This work strongly suggests that the ability to integrate the techniques of direct methods with anomalous dispersion would lead to improved methods for phase determination. The fusion of these techniques is described here. That the anticipated improvement is in fact realized is shown in Tables 1 and 2 and Fig. 6. Not only do the new formulas lead to improved estimates of the structure invariants but, more important still, because the distributions derived here are unimodal in the whole interval from $-\pi$ to $+\pi$, the twofold ambiguity inherent in all the earlier work is removed. It is believed that this resolution of the twofold ambiguity results from the ability now to make use of the individual magnitudes in the first neighborhood of the structure invariant and the avoidance of explicit dependence on the Bijvoet differences; the explicit use of the Bijvoet differences, as had been done in all previous work, leads apparently to a loss of information resulting in a twofold ambiguity in estimates of the structure invariants. It may be of some interest to observe that in the earlier work with anomalous dispersion only the sine of the invariant may be estimated; in the absence of anomalous scatterers only the cosine of the invariant may be estimated; as a result of the work described here both the sine and the cosine, that is to say the invariant itself, may be estimated. Since, in the presence of anomalous scatterers, the observed intensities are known to determine a unique enantiomorph, and therefore unique values for all the structure seminvariants, formulas of the kind described here should not be unexpected; nevertheless not even their existence appears to have been anticipated.

3.2. The normalized structure factors

In the presence of anomalous scatterers the normalized structure factor

$$E_{\mathbf{H}} = |E_{\mathbf{H}}|\exp(i\phi_{\mathbf{H}}) \tag{49}$$

is defined by

$$E_{\mathbf{H}} = \frac{1}{\alpha_{\mathbf{H}}^{1/2}} \sum_{j=1}^{N} f_{j\mathbf{H}}\exp(2\pi i\mathbf{H}\cdot\mathbf{r}_j) \tag{50}$$

$$= \frac{1}{\alpha_{\mathbf{H}}^{1/2}} \sum_{j=1}^{N} |f_{j\mathbf{H}}|\exp[i(\delta_{j\mathbf{H}} + 2\pi\mathbf{H}\cdot\mathbf{r}_j)] \tag{51}$$

where

$$f_{j\mathbf{H}} = |f_{j\mathbf{H}}|\exp(i\delta_{j\mathbf{H}}) \tag{52}$$

is the (in general complex) atomic scattering factor (a function of $|\mathbf{H}|$ as well as of j) of the atom labeled j, $\mathbf{r}_j$ is its position vector, N is the number of atoms in the unit cell, and

$$\alpha_{\mathbf{H}} = \sum_{j=1}^{N} |f_{j\mathbf{H}}|^2. \tag{53}$$

For a normal scatter, $\delta_{j\mathbf{H}} = 0$; for an atom which scatters anomalously, $\delta_{j\mathbf{H}} \neq 0$. Owing to the presence of the anomalous scatterers, the atomic scattering factors $f_{j\mathbf{H}}$, as functions of $\sin\theta/\lambda$, do not have the same shape for different atoms, even approximately. Hence the dependence of the $f_{j\mathbf{H}}$ on $|\mathbf{H}|$ cannot be ignored, in contrast to the usual practice when anomalous scatterers are not present. For this reason the subscript $\mathbf{H}$ is not suppressed in the symbols $f_{j\mathbf{H}}$ and $\alpha_{\mathbf{H}}$ [Eq. (53)].

The reciprocal-lattice vector $\mathbf{H}$ is assumed to be fixed, and the primitive random variables are taken to be the atomic position vectors $\mathbf{r}_j$ which are assumed to be uniformly and independently distributed. Then $E_{\mathbf{H}}$, as a function, (Eq. (51)], of the primitive random variables $\mathbf{r}_j$, is itself a random variable and, as it turns out,

$$\left\langle |E_{\mathbf{H}}|^2 \right\rangle_{\mathbf{r}_j} = 1. \tag{54}$$

3.3. The two-phase structure invariant

The two-phase structure invariant, which has no analogue when no anomalous scatterers are present, is defined by

$$\psi = \phi_{\mathbf{H}} + \phi_{\overline{\mathbf{H}}}. \tag{55}$$

3.3.1. The first neighborhood

The first neighborhood of the two-phase structure invariant ψ [Eq. (55)] is defined to consist of the two magnitudes

$$|E_H|, \; |E_{\overline{H}}|, \tag{55}$$

which, because of the breakdown of Friedel's Law, are in general distinct.

3.3.2. Estimating the two-phase structure invariant

Define C_H and S_H by means of

$$C_H = \frac{1}{\alpha_H} \sum_{j=1}^{N} |f_{jH}|^2 \cos 2 \, \delta_{jH} \tag{57}$$

$$S_H = \frac{1}{\alpha_H} \sum_{j=1}^{N} |f_{jH}|^2 \sin 2 \, \delta_{jH} \tag{58}$$

where f_{jH}, δ_{jH}, and α_H are defined in (52) and (53). Define X and ξ by means of

$$X \cos \xi = C_H, \; X \sin \xi = -S_H, \tag{59}$$

$$X = \left[C_H^2 + S_H^2 \right]^{1/2}, \; \tan \xi = - S_H/C_H \tag{60}$$

Suppose that R and $\bar{R}$ are fixed non-negative numbers. In view of (48) to (50) the two-phase structure invariant $\phi_H + \phi_{\overline{H}}$, as a function of the primitive random variables r_j, is itself a random variable. Denote by $P(\Psi|R,\bar{R})$ the conditional probability distribution of the two-phase structure invariant $\phi_H + \phi_{\overline{H}}$, given the two magnitudes in its first neighborhood:

$$|E_H| = R, \; |E_{\overline{H}}| = \bar{R}. \tag{61}$$

Then (Hauptman, 1982; Giacovazzo, 1983)

$$P(\Psi|R, \bar{R}) \approx \left[2\pi I_0 \left[\frac{2R\bar{R}X}{1-X^2} \right] \right]^{-1} \exp \left\{ \frac{2R\bar{R}X}{1-X^2} \cos (\Psi + \xi) \right\}, \tag{62}$$

where X and ξ, defined by (57) - (60) are seen to be functions of the (complex) atomic scattering factors f_{jH}, which are presumed to be known. It should be noted that the distribution (62) has the same form as (30) but is centered at $-\xi$ instead of O. Since (62) has a unique maximum at $\Psi = -\xi$, it follows that

$$\phi_H + \phi_{\overline{H}} \approx -\xi \tag{63}$$

provided that the variance of the distribution is small i.e. provided that

$$A = \frac{2R\bar{R}X}{1-X^2} \text{ is large.} \tag{64}$$

It should be noted that, while A depends on R, $\bar{R}$ and $|H|$, for a fixed chemical composition ξ depends only on $|H|$ (or $\sin\theta/\lambda$) and is independent of R and $\bar{R}$.

3.4. The three-phase structure invariant

It will be assumed throughout that **H, K,** and **L** are fixed reciprocal-lattice vectors satisfying

$$H + K + L = O . \tag{65}$$

Owing to the breakdown of Friedel's law there are, in sharp contrast to the case that no anomalous scatterers are present, eight distinct three-phase structure invariants:

$$\psi_0 = \phi_H + \phi_K + \phi_L, \tag{66}$$

$$\psi_1 = -\phi_{\bar{H}} + \phi_K + \phi_L, \tag{67}$$

$$\psi_2 = \phi_H - \phi_{\bar{K}} + \phi_L, \tag{68}$$

$$\psi_3 = \phi_H + \phi_K - \phi_{\bar{L}}, \tag{69}$$

$$\psi_{\bar{0}} = \phi_{\bar{H}} + \phi_{\bar{K}} + \phi_{\bar{L}}, \tag{70}$$

$$\psi_{\bar{1}} = -\phi_H + \phi_{\bar{K}} + \phi_{\bar{L}}, \tag{71}$$

$$\psi_{\bar{2}} = \phi_{\bar{H}} - \phi_K + \phi_{\bar{L}}, \tag{72}$$

$$\psi_{\bar{3}} = \phi_{\bar{H}} + \phi_{\bar{K}} - \phi_L. \tag{73}$$

3.4.1. The first neighborhood

The first neighborhood of each of the three-phase structure invariants (66)-(73) is defined to consist of the six magnitudes:

$$|E_H|, |E_K|, |E_L|, |E_{\bar{H}}|, |E_{\bar{K}}|, |E_{\bar{L}}| \tag{74}$$

which, again owing to the breakdown of Friedel's law, are not in general equal in pairs.

3.4.2. The probabilistic background

Fix the reciprocal-lattice vectors **H**, **K**, and **L**, subject to (65). Suppose that the six non-negative numbers R_1, R_2, R_3, $R_{\bar{1}}$, $R_{\bar{2}}$ and $R_{\bar{3}}$ are also specified. Define the N-fold Cartesian product W to consist of all ordered N-tuples $(r_1, r_2, ..., r_N)$, where $r_1, r_2, ..., r_N$ are atomic position vectors. Suppose that the primitive random variable is the N-tuple $(r_1, r_2, ..., r_N)$ which is assumed to be uniformly distributed over the subset of W defined by

$$\left|E_{\mathbf{H}}\right| = R_1, \left|E_{\mathbf{K}}\right| = R_2, \left|E_{\mathbf{L}}\right| = R_3, \tag{75}$$

$$\left|E_{\overline{\mathbf{H}}}\right| = R_{\bar{1}}, \left|E_{\overline{\mathbf{K}}}\right| = R_{\bar{2}}, \left|E_{\overline{\mathbf{L}}}\right| = R_{\bar{3}}, \tag{76}$$

where the normalized structure factors E are defined by (50). Then the eight structure invariants

$$\psi_j, \psi_{\bar{j}}, j = 0, 1, 2, 3, \tag{77}$$

(66)-(73), as functions of the primitive random variables $(r_1, r_2, ... r_N)$, are themselves random variables.

Our major goal is to determine the conditional probability distribution of each of the three-phase structure invariants (66)-(73), given the six magnitudes (75) and (76) in its first neighborhood, which, in the favorable case that the variance of the distribution happens to be small, yields a reliable estimate of the invariant (the neighborhood principle).

3.4.3. Estimating the three-phase structure invariant
Denote by

$$P_j(\Psi|R_1, R_2, R_3, R_{\bar{1}}, R_{\bar{2}}, R_{\bar{3}}) = P_j(\Psi),$$

$$j = 0, 1, 2, 3, \bar{0}, \bar{1}, \bar{2}, \bar{3}, \tag{78}$$

the conditional probability distribution of each ψ_j, assuming as known the six magnitudes (74) in its first neighborhood. Then the final formula, the major result of this article, is simply (Hauptman, 1982; Giacovazzo, 1983)

$$P_j(\Psi) \approx \frac{1}{K_j} \exp\left\{ A_j \cos (\Psi - \omega_j) \right\},$$

$$j = 0, 1, 2, 3, \bar{0}, \bar{1}, \bar{2}, \bar{3} \tag{79}$$

where the parameters K_j, A_j, and ω_j are expressible in terms of the complex scattering factors $f_{j\mathbf{H}}$, $f_{j\mathbf{K}}$, $f_{j\mathbf{L}}$, presumed to be known, and the observed magni-

tudes $|E_H|$, $|E_K|$, $|E_L|$, $|E_{\overline{H}}|$, $|E_{\overline{K}}|$, $|E_{\overline{L}}|$ in the first neighborhood of the invariant. Since the K_j's and A_j's are positive, the maximum of (79) occurs at $\Psi = \omega_j$. Hence when the variance of the distribution (79) is small, i.e. when A_j is large, one obtains the reliable estimate

$$\psi_j = \omega_j, j = 0, 1, 2, 3, \overline{0}, \overline{1}, \overline{2}, \overline{3}, \tag{80}$$

for the structure invariant ψ_j. It should be emphasized that the estimate (80) is unique in the whole range from $-\pi$ to $+\pi$. No prior knowledge of the positions of the anomalous scatterers is needed, nor is it required that the anomalous scatterers be identical.

3.4.4. The applications

Using the presumed known coordinates of the $PtCl_4^{2-}$ derivative of the protein Cytochrome c_{550} from *Paracoccus denitrificans* (Timkovich & Dickerson, 1976), molecular weight $M_r \simeq 14{,}500$, space group $P2_12_12_1$, some 8300 normalized structure factors E were calculated (to a resolution of 2.5Å). In addition to the anomalous scatterers Pt and Cl, this structure contains one Fe and six S atoms which also scatter anomalously at the wavelength used ($CuK\alpha$). Using the 4000 phases ϕ_{hkl} corresponding to the 4000 largest $|E_{hkl}|$'s with hkl $\neq$ O, the three-phase structure invariants ψ_j, j = 0, 1, 2, 3, $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, [(66)-(73)], were gene-rated and the parameters ω_j and A_j needed to define the distributions (79), were calculated. All calculations were done on the VAX 11/780 computer; double precision (approximately 15 significant digits) was used in order to eliminate round-off errors. The values of the A's were arranged in descending order and the first 2000, sampled at intervals of 100, were used in the construction of Table 1; the top 60,000 were used for Table 2.

Table 1 lists 21 values of A_j, sampled as shown from the top 2000, the corresponding estimates ω_j (in degrees) of the invariants ψ_j, the true values of the ψ_j, and the magnitude of the error, $|\omega_j - \psi_j|$. Also listed are the six magni-tudes |E| in the first neighborhood of the corresponding invariant.

Table 2 gives the average magnitude of the error,

$$<|\omega_j - \psi_j|>, \tag{81}$$

in the nine cumulative groups shown, for the 60,000 most reliable estimates ω_j of the invariants ψ_j.

Tables 1 and 2 show firstly that, owing to the unexpectedly large number of large values of A_j, our formulas yield reliable (and unique) estimates of tens of thousands of the three-phase structure invariants. Secondly, the invariants which are most reliably estimated lie anywhere in the range from -180" to +180°, and appear to be uniformly distributed in this range (Columns 9 and 10 of Table 1). Finally , in sharp contrast to the case that no anomalous scatterers are present, the most reliable estimates are not necessarily of invariants corre-sponding to the most intense reflections but of those corresponding instead to reflections of only moderate intensity (Columns 2-7 of Table 1).

Fig. 6 shows a scatter diagram of ω_j versus ψ_j for the $PtCl_4^{2-}$ derivative of Cytochrome c_{550}, using 201 invariants sampled at intervals of length ten from the top 2000, as well as the line $\omega_j = \psi_j$. Since the line falls evenly among the points, it appears that the ω_j are unbiased estimates of the invariants ψ_j.

3.4.5. Concluding remarks

In this article the goal of integrating the techniques of direct methods with anomalous dispersion is realized. Specifically, the conditional probability distribution of the three-phase structure invariant, assuming as known the six magnitudes in its first neighborhood, is obtained. In the favorable case that the variance of the distribution happens to be small, the distribution yields a reliable estimate of the invariant (the neighborhood principle). It is particularly noteworthy that, in strong contrast to all previous work, the estimate is unique in the whole interval $(-\pi, \pi)$ and that any estimate in this range is possible (even, for example, in thr vicinity of $\pm\pi/2$ or π). The first applications of this work using error-free diffraction data have been made, and these show that in a typical case some tens of thousands of three-phase structure invariants may be estimated with unprecedented accuracy, even for a macromolecular crystal structure. Some preliminary calculations on a number of structures, not detailed here, show that the accuracy of the estimates depends in some complicated way on the complexity of the crystal structure, the number of anomalous scatterers, the strength of the anomalous signal, and the range of $\sin\theta/\lambda$. With smaller structures the accuracy may be greatly increased, average errors of only three or four degrees for thousands of invariants being not uncommon.

It should be stated in conclusion that the availability of reliable estimates for large numbers of the three-phase structure invariants implies that the traditional machinery of direct methods, in particular the tangent formula [Eq. (33)], suitably modified to accommodate the non-zero estimates of the invariants, may be carried over without essential change to estimate the values of the individual phases and thus to facilitate structure determination via anomalous dispersion. In view of the calculations summarized in Tables 1 and 2 and Fig. 6, it seems likely that, in time, even macromolecules will prove to be solvable in this way. It is clear, too, that, owing to the ability to estimate both the sine and cosine invariants, that is to say both the signs and magnitudes of the invariants, the unique enantiomorph determined by the observed intensities is automatically obtained. In fact the first application of these techniques using experimental diffraction data has already facilitated the solution of the unknown macromolecular structure Cd, Zn Metallothionein [Furey, et al., 1986].

Table 1. Twenty-one estimates ω_j (in degrees) of the structure invariants ψ_j sampled from the top 2,000 for the Pt Cu$_4^{2-}$ derivative of Cytochrome c$_{550}$.

Serial No.	$\|E_H\|$	$\|E_{\bar{H}}\|$	$\|E_K\|$	$\|E_{\bar{K}}\|$	$\|E_L\|$	$\|E_{\bar{L}}\|$	A_j	Estimated value ω_j of ψ_j	True value of ψ_j	Mag. of the Error $\|\omega_j-\psi_j\|$
1	2.17	2.04	0.89	1.03	0.85	0.67	6.92	− 58°	− 88°	30°
100	1.91	2.06	1.61	1.49	0.85	0.67	5.62	148	130	18
200	1.91	2.06	1.96	2.06	1.41	1.57	4.83	− 79	−121	42
300	2.36	2.48	1.56	1.69	0.82	0.68	4.52	52	2	50
400	2.17	2.04	1.34	1.48	1.28	1.15	4.31	79	96	17
500	1.85	1.94	0.85	0.67	0.78	0.92	4.21	56	42	14
600	2.17	2.04	0.92	1.04	0.86	0.70	4.10	146	148	2
700	1.39	1.28	0.85	0.67	0.87	0.75	4.02	− 72	− 68	4.
800	1.41	1.57	1.61	1.49	0.71	0.85	3.93	70	50	20
900	1.88	1.98	1.28	1.15	0.85	0.67	3.87	104	96	8
1,000	1.29	1.43	0.79	0.71	0.85	0.67	3.80	− 88	−138	50
1,100	1.34	1.48	1.34	1.22	1.25	1.16	3.76	− 72	−126	54
1,200	1.56	1.69	1.41	1.57	0.98	0.90	3.72	73	78	5
1,300	1.98	2.07	2.08	1.94	1.08	1.21	3.68	−161	−124	37
1,400	1.56	1.67	1.41	1.57	1.24	1.33	3.63	− 72	− 3	69
1,500	2.38	2.50	1.91	2.06	0.74	0.64	3.59	84	77	7
1,600	1.91	2.06	1.34	1.22	0.72	0.83	3.55	− 64	− 94	30
1,700	1.91	2.06	2.02	2.12	2.15	2.24	3.51	− 64	− 72	8
1,800	2.38	2.50	1.61	1.49	0.78	0.90	3.46	78	82	4
1,900	2.38	2.50	1.63	1.70	1.81	1.93	3.43	63	123	60
2,000	0.85	0.67	0.97	0.83	1.02	1.09	3.42	− 96	−126	30

Table 2. Average magnitude of the error (in degrees) in the top 60,000 estimated values of the three-phase structure invariants, cumulated in the nine groups shown, for the Pt Ct_4^{2-} derivative of Cytochrome c_{550}.

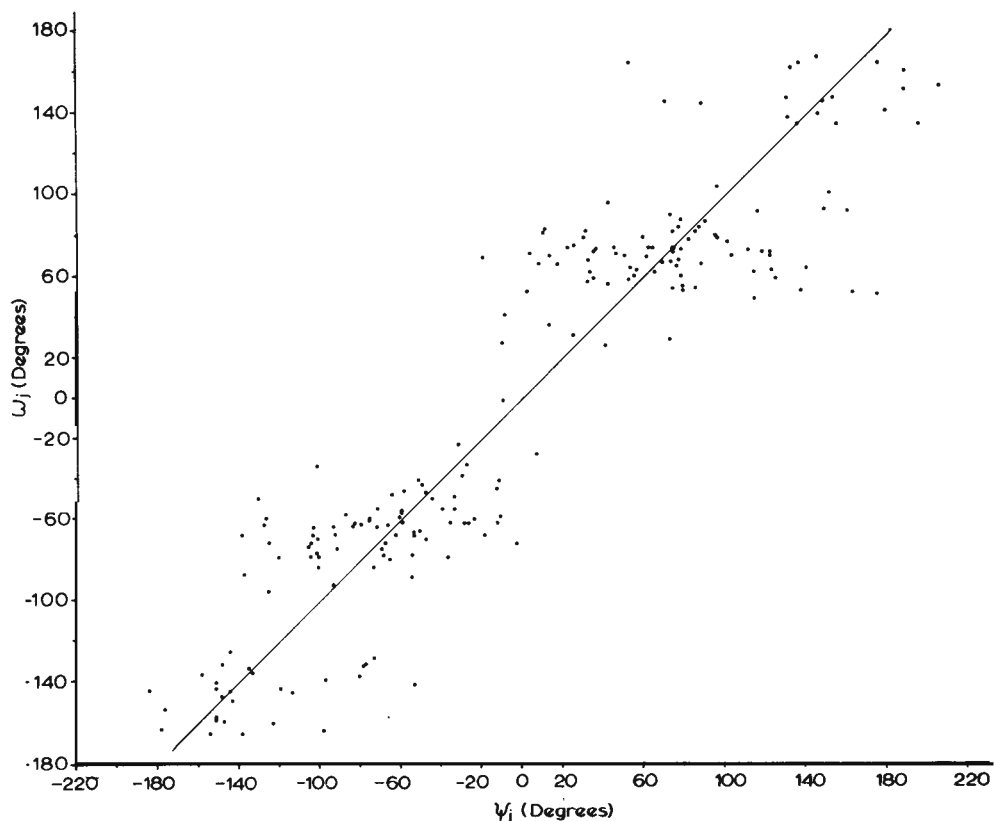| Group No.	No. in Group	Average Value of A	Average Mag. of Error
1	100	6.01	27.9°
2	500	4.90	29.3
3	1,000	4.44	28.8
4	2,000	4.02	28.0
5	5,000	3.49	31.4
6	10,000	3.09	33.8
7	20,000	2.71	36.1
8	40,000	2.35	38.6
9	60,000	2.15	39.8

Fig. 6. A scatter diagram of ω_j *versus* ψ_j, using 201 invariants sampled at intervals of length ten from the top 2000, for the $PtCt_4^{2-}$ derivative of Cytochrome c_{550}, as well as the line $\omega_j = \psi_j$.

REFERENCES

Cochran, W. (1955), Acta Cryst. 8, 473-478.

Green, E. and Hauptman, H. (1976), Acta Cryst. A32, 940-944.

Fortier, S. and Hauptman, H. (1977), A33, 694-696.

Furey, W. F., Robbins, A. H., Clancy, L. L., Winge, D. R., Wang, B. C., and Stout, C. D. (1986).
    Science 231, 704-710.

Giacovazzo, C. (1983), Acta Cryst. A39, 585-592.

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

- (1975a), Acta Cryst. A31, 671-679.

- (1975b), Acta Cryst. A31, 680-687.

- (1976), Acta Cryst. A32, 877-882.

- (1977a), Acta Cryst. A33, 553-555.

- (1977b), Acta Cryst. A33 , 568-571.

- (1982), Acta Cryst. A38 , 632-641.

- and Karle, J. (1953), Solution of the Phase Problem I. The Centrosymmetric Crystal. ACA
    Monograph No. 3. Polycrystal Book Service.

- and Karle, J. (1956), Acta Cyst. 9, 45-55.

- and Karle, J. (1959), Acta Cryst. 12, 93-97.

Heinerman, J. J. L., Krabbendam, H., Kroon, J., and Spek, A. L. (1978). Acta Cryst. A34, 447-450.

Hendrickson, W. A. and Teeter, M. M. (1981), Nature (London) 290, 107- 113.

Karle, J. and Hauptman, H. (1955), Acta Cryst. 9, 635-651.

- - (1961), Acta Cryst. 14, 217-223.

Kroon, J., Spek, A. L., and Krabbendam, H. (1977), Acta Cryst. A33,382-385.

Lessinger, L. and Wondratschek, H. (1975), Acta Cryst. A31, 521.

Okaya, Y. and Pepinsky, R. (1956), Phys. Rev. 103, 1645-1647.

Peerdeman, A. F. and Bijvoet, J. M. (1956), Proc. K. Ned. Akad. Wet. B59, 312- 313.

Ramachandran, G. N. and Raman, S. (1956), Curr. Sci. 25, 348-351.

Rossmann, M. G. (1961), Acta Cryst. 14, 383-388 .

Timkovich, R. and Dickerson, R. E. (1976), J. Biol. Chem. 251, 4033-4046 .