

## 2.2. DIRECT METHODS

which plotted at various  $s^2$  should be a straight line of which the slope ( $2B$ ) and intercept ( $\ln K$ ) on the logarithmic axis can be obtained by applying a linear least-squares procedure.

Very often molecular geometries produce perceptible departures from linearity in the logarithmic Wilson plot. However, the more extensive the available *a priori* information on the structure is, the closer, on the average, are the Wilson-plot curves to their least-squares straight lines.

Accurate estimates of  $B$  and  $K$  require good strategies (Rogers & Wilson, 1953) for:

(1) treatment of weak measured data. If weak data are set to zero, there will be bias in the statistics. Methods are, however, available (French & Wilson, 1978) that provide an *a posteriori* estimate of weak (even negative) intensities by means of Bayesian statistics.

(2) treatment of missing weak data (Rogers *et al.*, 1955; Vicković & Viterbo, 1979). All unobserved reflections may assume

$$\mu = |F_{o\min}|^2/3 \text{ for cs. space groups}$$

$$\mu = |F_{o\min}|^2/2 \text{ for ncs. space groups,}$$

where the subscript ‘*o min*’ refers to the minimum observed intensity.

Once  $K$  and  $B$  have been estimated,  $E_{\mathbf{h}}$  values can be obtained from experimental data by

$$|E_{\mathbf{h}}|^2 = \frac{KI_{\mathbf{h}}}{\langle |F_{\mathbf{h}}^o|^2 \rangle \exp(-2Bs^2)},$$

where  $\langle |F_{\mathbf{h}}^o|^2 \rangle$  is the expected value of  $|F_{\mathbf{h}}^o|^2$  for the reflection  $\mathbf{h}$  on the basis of the available *a priori* information.

#### 2.2.4.4. Probability distributions of normalized structure factors

Under some fairly general assumptions (see Chapter 2.1) probability distribution functions for the variable  $|E|$  for cs. and ncs. structures are (see Fig. 2.2.4.1)

$${}_1P(|E|) d|E| = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{E^2}{2}\right) d|E| \quad (2.2.4.4)$$

and

$${}_1P(|E|) d|E| = 2|E| \exp(-|E|^2) d|E|, \quad (2.2.4.5)$$

respectively. Corresponding cumulative functions are (see Fig. 2.2.4.2)

$${}_1N(|E|) = \sqrt{\frac{2}{\pi}} \int_0^{|E|} \exp\left(-\frac{t^2}{2}\right) dt = \operatorname{erf}\left(\frac{|E|}{\sqrt{2}}\right),$$

$${}_1N(|E|) = \int_0^{|E|} 2t \exp(-t^2) dt = 1 - \exp(-|E|^2).$$

Some moments of the distributions (2.2.4.4) and (2.2.4.5) are listed in Table 2.2.4.1. In the absence of other indications for a given crystal structure, a cs. or an ncs. space group will be preferred according to whether the statistical tests yield values closer to column 2 or to column 3 of Table 2.2.4.1.

For further details about the distribution of intensities see Chapter 2.1.

#### 2.2.5. Phase-determining formulae

From the earliest periods of X-ray structure analysis several authors (Ott, 1927; Banerjee, 1933; Avrami, 1938) have tried to determine atomic positions directly from diffraction intensities. Significant

Table 2.2.4.1. Moments of the distributions (2.2.4.4) and (2.2.4.5)

$R(E_s)$  is the percentage of n.s.f.’s with amplitude greater than the threshold  $E_s$ .

Criterion	Centrosymmetric distribution	Noncentrosymmetric distribution
$\langle  E  \rangle$	0.798	0.886
$\langle  E ^2 \rangle$	1.000	1.000
$\langle  E ^3 \rangle$	1.596	1.329
$\langle  E ^4 \rangle$	3.000	2.000
$\langle  E ^5 \rangle$	6.383	3.323
$\langle  E ^6 \rangle$	15.000	6.000
$\langle  E^2 - 1  \rangle$	0.968	0.736
$\langle (E^2 - 1)^2 \rangle$	2.000	1.000
$\langle (E^2 - 1)^3 \rangle$	8.000	2.000
$\langle  E^2 - 1 ^3 \rangle$	8.691	2.415
$R(1)$	0.320	0.368
$R(2)$	0.050	0.018
$R(3)$	0.003	0.0001

developments are the derivation of inequalities and the introduction of probabilistic techniques *via* the use of joint probability distribution methods (Hauptman & Karle, 1953).

#### 2.2.5.1. Inequalities among structure factors

An extensive system of inequalities exists for the coefficients of a Fourier series which represents a positive function. This can restrict the allowed values for the phases of the s.f.’s in terms of measured structure-factor magnitudes. Harker & Kasper (1948) derived two types of inequalities:

*Type 1.* A modulus is bound by a combination of structure factors:

$$|U_{\mathbf{h}}|^2 \leq \frac{1}{m} \sum_{s=1}^m a_s(-\mathbf{h}) U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)}, \quad (2.2.5.1)$$

where  $m$  is the order of the point group and  $a_s(-\mathbf{h}) = \exp(-2\pi i \mathbf{h} \cdot \mathbf{T}_s)$ .

Applied to low-order space groups, (2.2.5.1) gives

$$P1: |U_{h,k,l}|^2 \leq 1$$

$$P\bar{1}: U_{h,k,l}^2 \leq 0.5 + 0.5U_{2h,2k,2l}$$

$$P2_1: |U_{h,k,l}|^2 \leq 0.5 + 0.5(-1)^k U_{2h,0,2l}.$$

The meaning of each inequality is easily understandable: in  $P\bar{1}$ , for example,  $U_{2h,2k,2l}$  must be positive if  $|U_{h,k,l}|$  is large enough.

*Type 2.* The modulus of the sum or of the difference of two structure factors is bound by a combination of structure factors:

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \leq \frac{1}{m} \left\{ \sum_{s=1}^m a_s(-\mathbf{h}) U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)} + \sum_{s=1}^m a_s(-\mathbf{h}') U_{\mathbf{h}'(\mathbf{I}-\mathbf{R}_s)} \right. \\ \left. \pm 2 \operatorname{Re} \left[ \sum_{s=1}^m a_s(-\mathbf{h}') U_{\mathbf{h}-\mathbf{h}'\mathbf{R}_s} \right] \right\} \quad (2.2.5.2)$$

where  $\operatorname{Re}$  stands for ‘real part of’. Equation (2.2.5.2) applied to  $P1$  gives

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \leq 2 \pm 2|U_{\mathbf{h}-\mathbf{h}'}| \cos \varphi_{\mathbf{h}-\mathbf{h}'},$$

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

A variant of (2.2.5.2) valid for cs. space groups is

$$(U_{\mathbf{h}} \pm U_{\mathbf{h}'})^2 \leq (1 \pm U_{\mathbf{h}+\mathbf{h}'})(1 \pm U_{\mathbf{h}-\mathbf{h}'}).$$

After Harker & Kasper's contributions, several other inequalities were discovered (Gillis, 1948; Goedkoop, 1950; Okaya & Nitta, 1952; de Wolff & Bouman, 1954; Bouman, 1956; Oda *et al.*, 1961). The most general are the Karle-Hauptman inequalities (Karle & Hauptman, 1950):

$$D_m = \begin{vmatrix} U_0 & U_{-\mathbf{h}_1} & U_{-\mathbf{h}_2} & \dots & U_{-\mathbf{h}_n} \\ U_{\mathbf{h}_1} & U_0 & U_{\mathbf{h}_1-\mathbf{h}_2} & \dots & U_{\mathbf{h}_1-\mathbf{h}_n} \\ U_{\mathbf{h}_2} & U_{\mathbf{h}_2-\mathbf{h}_1} & U_0 & \dots & U_{\mathbf{h}_2-\mathbf{h}_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{\mathbf{h}_n} & U_{\mathbf{h}_n-\mathbf{h}_1} & U_{\mathbf{h}_n-\mathbf{h}_2} & \dots & U_0 \end{vmatrix} \geq 0. \quad (2.2.5.3)$$

The determinant can be of any order but the leading column (or row) must consist of  $U$ 's with different indices, although, within the column, symmetry-related  $U$ 's may occur. For  $n=2$  and  $\mathbf{h}_2 = 2\mathbf{h}_1 = 2\mathbf{h}$ , equation (2.2.5.3) reduces to

$$D_3 = \begin{vmatrix} U_0 & U_{-\mathbf{h}} & U_{-2\mathbf{h}} \\ U_{\mathbf{h}} & U_0 & U_{-\mathbf{h}} \\ U_{2\mathbf{h}} & U_{\mathbf{h}} & U_0 \end{vmatrix} \geq 0,$$

which, for cs. structures, gives the Harker & Kasper inequality

$$U_{\mathbf{h}}^2 \leq 0.5 + 0.5U_{2\mathbf{h}}.$$

For  $m=3$ , equation (2.2.5.3) becomes

$$D_3 = \begin{vmatrix} U_0 & U_{-\mathbf{h}} & U_{-\mathbf{k}} \\ U_{\mathbf{h}} & U_0 & U_{\mathbf{h}-\mathbf{k}} \\ U_{\mathbf{k}} & U_{\mathbf{k}-\mathbf{h}} & U_0 \end{vmatrix} \geq 0,$$

from which

$$1 - |U_{\mathbf{h}}|^2 - |U_{\mathbf{k}}|^2 - |U_{\mathbf{h}-\mathbf{k}}|^2 + 2|U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}| \cos \alpha_{\mathbf{h}, \mathbf{k}} \geq 0, \quad (2.2.5.4)$$

where

$$\alpha_{\mathbf{h}, \mathbf{k}} = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}.$$

If the moduli  $|U_{\mathbf{h}}|$ ,  $|U_{\mathbf{k}}|$ ,  $|U_{\mathbf{h}-\mathbf{k}}|$  are large enough, (2.2.5.4) is not satisfied for all values of  $\alpha_{\mathbf{h}, \mathbf{k}}$ . In cs. structures the eventual check that one of the two values of  $\alpha_{\mathbf{h}, \mathbf{k}}$  does not satisfy (2.2.5.4) brings about the unambiguous identification of the sign of the product  $U_{\mathbf{h}}U_{\mathbf{k}}U_{\mathbf{h}-\mathbf{k}}$ .

It was observed (Gillis, 1948) that 'there was a number of cases in which both signs satisfied the inequality, one of them by a comfortable margin and the other by only a relatively small margin. In almost all such cases it was the former sign which was the correct one. That suggests that the method may have some power in reserve in the sense that there are still fundamentally stronger inequalities to be discovered'. Today we identify this power in reserve in the use of probability theory.

### 2.2.5.2. Probabilistic phase relationships for structure invariants

For any space group (see Section 2.2.3) there are linear combinations of phases with cosines that are, in principle, fixed by the  $|E|$  magnitudes alone (s.i.'s) or by the  $|E|$  values and the trigonometric form of the structure factor (s.s.'s). This result greatly stimulated the calculation of conditional distribution functions

$$P(\Phi|\{R\}), \quad (2.2.5.5)$$

where  $R_{\mathbf{h}} = |E_{\mathbf{h}}|$ ,  $\Phi = \sum A_i \varphi_{\mathbf{h}_i}$  is an s.i. or an s.s. and  $\{R\}$  is a

suitable set of diffraction magnitudes. The method was first proposed by Hauptman & Karle (1953) and was developed further by several authors (Bertaut, 1955*a,b*, 1960; Klug, 1958; Naya *et al.*, 1964, 1965; Giacovazzo, 1980*a*). From a probabilistic point of view the crystallographic problem is clear: the joint distribution  $P(E_{\mathbf{h}_1}, \dots, E_{\mathbf{h}_n})$ , from which the conditional distributions (2.2.5.5) can be derived, involves a number of normalized structure factors each of which is a linear sum of random variables (the atomic contributions to the structure factors). So, for the probabilistic interpretation of the phase problem, the atomic positions and the reciprocal vectors may be considered as random variables. A further problem is that of identifying, for a given  $\Phi$ , a suitable set of magnitudes  $|E|$  on which  $\Phi$  primarily depends. The formulation of the *nested neighbourhood principle* first (Hauptman, 1975) fixed the idea of defining a sequence of sets of reflections each contained in the succeeding one and having the property that any s.i. or s.s. may be estimated *via* the magnitudes constituting the various neighbourhoods. A subsequent more general theory, the *representation method* (Giacovazzo, 1977*a*, 1980*b*), arranges for any  $\Phi$  the set of intensities in a sequence of subsets in order of their expected effectiveness (in the statistical sense) for the estimation of  $\Phi$ .

In the following sections the main formulae estimating low-order invariants and seminvariants or relating phases to other phases and diffraction magnitudes are given.

### 2.2.5.3. Triplet relationships

The basic formula for the estimation of the triplet phase  $\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}$  given the parameter  $G = 2\sigma_3\sigma_2^{-3/2} \times R_{\mathbf{h}}R_{\mathbf{k}}R_{\mathbf{h}-\mathbf{k}}$  is Cochran's (1955) formula

$$P(\Phi) = [2\pi I_0(G)]^{-1} \exp(G \cos \Phi), \quad (2.2.5.6)$$

where  $\sigma_n = \sum_{j=1}^N Z_j^n$ ,  $Z_j$  is the atomic number of the  $j$ th atom and  $I_n$  is the modified Bessel function of order  $n$ . In Fig. 2.2.5.1 the distribution  $P(\Phi)$  is shown for different values of  $G$ .

The conditional probability distribution for  $\varphi_{\mathbf{h}}$ , given a set of  $(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})$  and  $G_j = 2\sigma_3\sigma_2^{-3/2} R_{\mathbf{h}}R_{\mathbf{k}_j}R_{\mathbf{h}-\mathbf{k}_j}$ , is given (Karle & Hauptman, 1956; Karle & Karle, 1966) by

$$P(\varphi_{\mathbf{h}}) = [2\pi I_0(\alpha)]^{-1} \exp[\alpha \cos(\varphi_{\mathbf{h}} - \beta_{\mathbf{h}})], \quad (2.2.5.7)$$

where

$$\alpha^2 = \left[ \sum_{j=1}^r G_{\mathbf{h}, \mathbf{k}_j} \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}) \right]^2 + \left[ \sum_{j=1}^r G_{\mathbf{h}, \mathbf{k}_j} \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j}) \right]^2 \quad (2.2.5.8)$$

$$\tan \beta_{\mathbf{h}} = \frac{\sum_j G_{\mathbf{h}, \mathbf{k}_j} \sin(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}{\sum_j G_{\mathbf{h}, \mathbf{k}_j} \cos(\varphi_{\mathbf{k}_j} + \varphi_{\mathbf{h}-\mathbf{k}_j})}. \quad (2.2.5.9)$$

$\beta_{\mathbf{h}}$  is the most probable value for  $\varphi_{\mathbf{h}}$ . The variance of  $\varphi_{\mathbf{h}}$  may be obtained from (2.2.5.7) and is given by

$$V_{\mathbf{h}} = \frac{\pi^2}{3} + [I_0(\alpha)]^{-1} \sum_{n=1}^{\infty} \frac{I_{2n}(\alpha)}{n^2} - 4[I_0(\alpha)]^{-1} \sum_{n=0}^{\infty} \frac{I_{2n+1}(\alpha)}{(2n+1)^2}, \quad (2.2.5.10)$$

which is plotted in Fig. 2.2.5.2.

Equation (2.2.5.9) is the so-called *tangent formula*. According to (2.2.5.10), the larger is  $\alpha$  the more reliable is the relation  $\varphi_{\mathbf{h}} = \beta_{\mathbf{h}}$ . For an equal-atom structure  $\sigma_3\sigma_2^{-3/2} = N^{-1/2}$ .

## 2.2. DIRECT METHODS

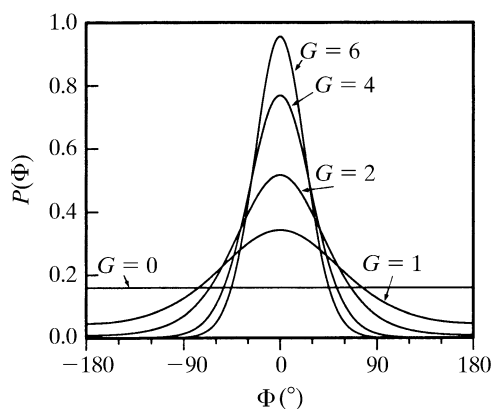


Fig. 2.2.5.1. Curves of (2.2.5.6) for some values of  $G = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|$ .

The basic conditional formula for sign determination of  $E_{\mathbf{h}}$  in cs. crystals is Cochran & Woolfson's (1955) formula

$$P^+ = \frac{1}{2} + \frac{1}{2} \tanh \left( \sigma_3 \sigma_2^{-3/2} |E_{\mathbf{h}}| \sum_{j=1}^r E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j} \right), \quad (2.2.5.11)$$

where  $P^+$  is the probability that  $E_{\mathbf{h}}$  is positive and  $\mathbf{k}$  ranges over the set of known values  $E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$ . The larger the absolute value of the argument of  $\tanh$ , the more reliable is the phase indication.

An auxiliary formula exploiting all the  $|E|$ 's in reciprocal space in order to estimate a single  $\Phi$  is the  $B_{3,0}$  formula (Hauptman & Karle, 1958; Karle & Hauptman, 1958) given by

$$\begin{aligned} & |E_{\mathbf{h}_1} E_{\mathbf{h}_2} E_{-\mathbf{h}_1-\mathbf{h}_2}| \cos(\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1+\mathbf{h}_2}) \\ & \simeq C \langle (|E_{\mathbf{k}}|^p - \overline{|E|^p}) (|E_{\mathbf{h}_1+\mathbf{k}}|^p - \overline{|E|^p}) (|E_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{k}}|^p - \overline{|E|^p}) \rangle_{\mathbf{k}} \\ & - \frac{2\sigma_6}{\sigma_4^{3/2}} + \frac{\sigma_8^{1/2}}{\sigma_4} (|E_{\mathbf{h}_1}|^2 + |E_{\mathbf{h}_2}|^2 + |E_{\mathbf{h}_1+\mathbf{h}_2}|^2) \dots, \quad (2.2.5.12) \end{aligned}$$

where  $C$  is a constant which differs for cs. and ncs. crystals,  $\overline{|E|^p}$  is the average value of  $|E|^p$  and  $p$  is normally chosen to be some small number. Several modifications of (2.2.5.12) have been proposed (Hauptman, 1964, 1970; Karle, 1970a; Giacovazzo, 1977b).

A recent formula (Cascarano, Giacovazzo, Camalli *et al.*, 1984) exploits information contained within the second representation of  $\Phi$ , that is to say, within the collection of special quintets (see Section 2.2.5.6):

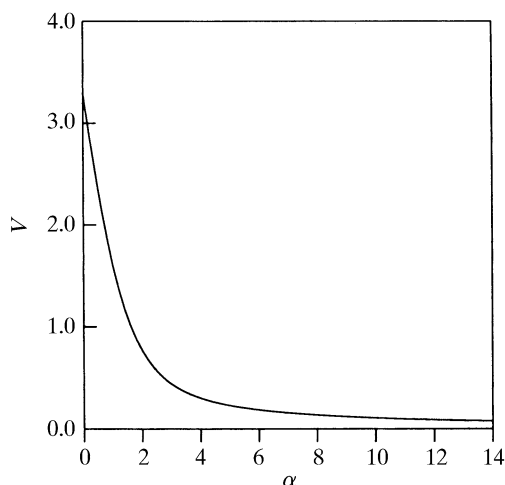


Fig. 2.2.5.2. Variance (in square radians) as a function of  $\alpha$ .

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} - \varphi_{\mathbf{h}_1+\mathbf{h}_2} + \varphi_{\mathbf{k}} - \varphi_{-\mathbf{k}},$$

where  $\mathbf{k}$  is a free vector. The formula retains the same algebraic form as (2.2.5.6), but

$$G = \frac{2R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3}}{\sqrt{N}} (1 + Q), \quad (2.2.5.13)$$

where  $[\mathbf{h}_3 = -(\mathbf{h}_1 + \mathbf{h}_2)]$ ,

$$Q = \sum_{\mathbf{k}} \frac{\sum_{i=1}^m A_{\mathbf{k},i} / N}{1 + (\varepsilon_{\mathbf{h}_1} \varepsilon_{\mathbf{h}_2} \varepsilon_{\mathbf{h}_3} + \sum_{i=1}^m B_{\mathbf{k},i}) / 2N},$$

$$\begin{aligned} A_{\mathbf{k},i} &= \varepsilon_{\mathbf{k}} [\varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i} (\varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i}) \\ &+ \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i} (\varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i}) \\ &+ \varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i} (\varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i})], \\ B_{\mathbf{k},i} &= \varepsilon_{\mathbf{h}_1} [\varepsilon_{\mathbf{k}} (\varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i}) \\ &+ \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i} \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} \varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}] \\ &+ \varepsilon_{\mathbf{h}_2} [\varepsilon_{\mathbf{k}} (\varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i}) \\ &+ \varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i} \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} \varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i}] \\ &+ \varepsilon_{\mathbf{h}_3} [\varepsilon_{\mathbf{k}} (\varepsilon_{\mathbf{h}_3+\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_3-\mathbf{k}\mathbf{R}_i}) \\ &+ \varepsilon_{\mathbf{h}_1+\mathbf{k}\mathbf{R}_i} \varepsilon_{\mathbf{h}_2-\mathbf{k}\mathbf{R}_i} + \varepsilon_{\mathbf{h}_1-\mathbf{k}\mathbf{R}_i} \varepsilon_{\mathbf{h}_2+\mathbf{k}\mathbf{R}_i}]; \end{aligned}$$

$\varepsilon = |E|^2 - 1$ ,  $(\varepsilon_{\mathbf{h}_1} \varepsilon_{\mathbf{h}_2} \varepsilon_{\mathbf{h}_3} + \sum_{i=1}^m B_{\mathbf{k},i})$  is assumed to be zero if it is experimentally negative. The prime to the summation warns the reader that precautions have to be taken in order to avoid duplications in the contributions.

$G$  may be positive or negative. In particular, if  $G < 0$  the triplet is estimated negative.

The accuracy with which the value of  $\Phi$  is estimated strongly depends on  $\varepsilon_{\mathbf{k}}$ . Thus, in practice, only a subset of reciprocal space (the reflections  $\mathbf{k}$  with large values of  $\varepsilon$ ) may be used for estimating  $\Phi$ .

(2.2.5.13) proved to be quite useful in practice. Positive triplet cosines are ranked in order of reliability by (2.2.5.13) markedly better than by Cochran's parameters. Negative estimated triplet cosines may be excluded from the phasing process and may be used as a figure of merit for finding the correct solution in a multisolution procedure.

### 2.2.5.4. Triplet relationships using structural information

A strength of direct methods is that no knowledge of structure is required for their application. However, when some *a priori* information is available, it should certainly be a weakness of the methods not to make use of this knowledge. The conditional distribution of  $\Phi$  given  $R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3}$  and the first three of the five kinds of *a priori* information described in Section 2.2.4.1 is (Main, 1976; Heinemann, 1977a)

$$P(\Phi) \simeq \frac{\exp\{2QR_1R_2R_3 \cos(\Phi - q)\}}{2\pi I_0(2QR_1R_2R_3)}, \quad (2.2.5.14)$$

where

$$Q \exp(iq) = \frac{\sum_{i=1}^p g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)}{\langle |F_{\mathbf{h}_1}|^2 \rangle^{1/2} \langle |F_{\mathbf{h}_2}|^2 \rangle^{1/2} \langle |F_{\mathbf{h}_3}|^2 \rangle^{1/2}}.$$

$\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$  stand for  $\mathbf{h}, -\mathbf{k}, -\mathbf{h} + \mathbf{k}$ , and  $R_1, R_2, R_3$  for  $R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h}-\mathbf{k}}$ . The quantities  $\langle |F_{\mathbf{h}_i}|^2 \rangle$  have been calculated in Section 2.2.4.1 according to different categories:  $g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)$  is a suitable average of the product of three scattering factors for the  $i$ th atomic group,  $p$  is the number of atomic groups in the cell including those related by symmetry elements. We have the following categories.

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

(a) *No structural information*

(2.2.5.14) then reduces to (2.2.5.6).

(b) *Randomly positioned and randomly oriented atomic groups*

Then

$$g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = \sum_{j,k,l} f_j f_k f_l \langle \exp[2\pi i(\mathbf{h}_1 \cdot \mathbf{r}_{kj} + \mathbf{h}_2 \cdot \mathbf{r}_{lj})] \rangle_R,$$

where  $\langle \dots \rangle_R$  means rotational average. The average of the exponential term extends over all orientations of the triangle formed by the atoms  $j, k$  and  $l$ , and is given (Hauptman, 1965) by

$$\begin{aligned} B(z, t) &= \langle \exp[2\pi i(\mathbf{h} \cdot \mathbf{r} + \mathbf{h}' \cdot \mathbf{r}')] \rangle \\ &= \left(\frac{\pi}{2z}\right)^{1/2} \sum_{n=0}^{\infty} \frac{t^{2n}}{(n!)^2} J_{(4n+1)/2}(z), \end{aligned}$$

where

$$z = 2\pi[q^2 r^2 + 2qrq'r' \cos \varphi_q \cos \varphi_r + q'^2 r'^2]^{1/2}$$

and

$$t = [2\pi^2 qrq'r' \sin \varphi_q \sin \varphi_r] / z;$$

$q, q', r$  and  $r'$  are the magnitudes of  $\mathbf{h}, \mathbf{h}', \mathbf{r}$  and  $\mathbf{r}'$ , respectively;  $\varphi_q$  and  $\varphi_r$  are the angles  $(\mathbf{h}, \mathbf{h}')$  and  $(\mathbf{r}, \mathbf{r}')$ , respectively.

(c) *Randomly positioned but correctly oriented atomic groups*

Then

$$\begin{aligned} g_i(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) &= \sum_{s=1}^m \sum_{j,k,l} f_j f_k f_l \\ &\times \exp[2\pi i(\mathbf{h}_1 \cdot \mathbf{R}_s \mathbf{r}_{kj} + \mathbf{h}_2 \cdot \mathbf{R}_s \mathbf{r}_{lk})], \end{aligned}$$

where the summations over  $j, k, l$  are taken over all the atoms in the  $i$ th group.

A modified expression for  $g_i$  has to be used in polar space groups for special triplets (Giacovazzo, 1988).

Translation functions [see Chapter 2.3; for an overview, see also Beurskens *et al.* (1987)] are also used to determine the position of a correctly oriented molecular fragment.

Such functions can work in direct space [expressed as Patterson convolutions (Buerger, 1959; Nordman, 1985) or electron-density convolutions (Rossmann *et al.*, 1964; Argos & Rossmann, 1980)] or in reciprocal space [expressed as correlation functions (Crowther & Blow, 1967; Karle, 1972; Langs, 1985) or residual functions (Rae, 1977)]. Both the probabilistic methods and the translation functions are quite efficient tools: the decision as to which one to use is often a personal choice.

(d) *Atomic groups correctly positioned*

Let  $p$  be the number of atoms with known position,  $q$  the number of atoms with unknown position,  $F_p$  and  $F_q$  the corresponding structure factors.

Tangent recycling methods (Karle, 1970b) may be used for recovering the complete crystal structure. The phase  $\varphi_{p, \mathbf{h}}$  is accepted in the starting set as a useful approximation of  $\varphi_{\mathbf{h}}$  if  $|F_{p, \mathbf{h}}| > \eta |F_{\mathbf{h}}|$ , where  $\eta$  is the fraction of the total scattering power contained in the fragment and where  $|F_{\mathbf{h}}|$  is associated with  $|E_{\mathbf{h}}| > 1.5$ .

Tangent recycling methods are applied (Beurskens *et al.*, 1979) with greater effectiveness to difference s.f.'s  $\Delta F = (|F| - |F_p|) \exp(i\varphi_p)$ . The weighted tangent formula uses  $\Delta F_{\mathbf{h}}$  values in order to convert them to more probable  $F_{q, \mathbf{h}}$  values.

From a probabilistic point of view (Giacovazzo, 1983a; Camalli *et al.*, 1985) the distribution of  $\varphi_{\mathbf{h}}$ , given  $E'_{p, \mathbf{h}}$  and some products  $(E'_{\mathbf{k}} - E'_{p, \mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}} - E'_{p, \mathbf{h}-\mathbf{k}})$ , is the von Mises function

$$P(\varphi_{\mathbf{h}} | \dots) = [2\pi I_0(\alpha)]^{-1} \exp[\alpha \cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})], \quad (2.2.5.15)$$

where  $\theta_{\mathbf{h}}$ , the most probable value of  $\varphi_{\mathbf{h}}$ , is given by

$$\begin{aligned} \tan \theta_{\mathbf{h}} &\simeq \alpha'_2 / \alpha'_1, \\ \alpha^2 &= \alpha_1'^2 + \alpha_2'^2 \end{aligned} \quad (2.2.5.16)$$

and

$$\begin{aligned} \alpha'_1 &= 2R'_{\mathbf{h}} \left\{ \mathcal{R} \left[ E'_{p, \mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} (E'_{\mathbf{k}} - E'_{p, \mathbf{k}}) \right. \right. \\ &\quad \left. \left. \times (E'_{\mathbf{h}-\mathbf{k}} - E'_{p, \mathbf{h}-\mathbf{k}}) \right] \right\} \\ \alpha'_2 &= 2R'_{\mathbf{h}} \left\{ \mathcal{I} \left[ E'_{p, \mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} (E'_{\mathbf{k}} - E'_{p, \mathbf{k}}) \right. \right. \\ &\quad \left. \left. \times (E'_{\mathbf{h}-\mathbf{k}} - E'_{p, \mathbf{h}-\mathbf{k}}) \right] \right\}. \end{aligned}$$

$\mathcal{R}$  and  $\mathcal{I}$  stand for 'real' and 'imaginary part of', respectively. Furthermore,  $E' = F / \sum_q^{1/2}$  is a pseudo-normalized s.f. If no pair  $(\varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}})$  is known, then

$$\begin{aligned} \alpha'_1 &= 2R'_{\mathbf{h}} R'_{p, \mathbf{h}} \cos \varphi_{p, \mathbf{h}} \\ \alpha'_2 &= 2R'_{\mathbf{h}} R'_{p, \mathbf{h}} \sin \varphi_{p, \mathbf{h}} \end{aligned}$$

and (2.2.5.15) reduces to Sim's (1959) equation

$$P(\varphi_{\mathbf{h}}) \simeq [2\pi I_0(G)]^{-1} \exp[G \cos(\varphi_{\mathbf{h}} - \varphi_{p, \mathbf{h}})], \quad (2.2.5.17)$$

where  $G = 2R'_{\mathbf{h}} R'_{p, \mathbf{h}}$ . In this case  $\varphi_{p, \mathbf{h}}$  is the most probable value of  $\varphi_{\mathbf{h}}$ .

(e) *Pseudotranslational symmetry is present*

Substructure and superstructure reflections are then described by different forms of the structure-factor equation (Böhme, 1982; Gramlich, 1984; Fan *et al.*, 1983), so that probabilistic formulae estimating triplet cosines derived on the assumption that atoms are uniformly dispersed in the unit cell cannot hold. In particular, the reliability of each triplet also depends on, besides  $R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h}-\mathbf{k}}$ , the actual  $\mathbf{h}, \mathbf{k}, \mathbf{h}-\mathbf{k}$  indices and on the nature of the pseudotranslation. It has been shown (Casarano *et al.*, 1985b; Casarano, Giacovazzo & Luić, 1987) that (2.2.5.7), (2.2.5.8), (2.2.5.9) still hold provided  $G_{\mathbf{h}, \mathbf{k}_j}$  is replaced by

$$G'_{\mathbf{h}, \mathbf{k}_j} = \frac{2R_{\mathbf{h}} R_{\mathbf{k}_j} R_{\mathbf{h}-\mathbf{k}_j}}{\sqrt{N_{\mathbf{h}, \mathbf{k}}}},$$

where factors  $E$  and  $n_i$  are defined according to Section 2.2.4.1,

$$N_{\mathbf{h}, \mathbf{k}} = \frac{(\zeta_{\mathbf{h}}[\sigma_2]_p + [\sigma_2]_q)(\zeta_{\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)(\zeta_{\mathbf{h}-\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)}{\{(\beta/m)[\sigma_3]_p(n_1^2 n_2^2 n_3^2 \dots) + [\sigma_3]_q\}^2},$$

and  $\beta$  is the number of times for which

$$\begin{aligned} \mathbf{hR}_s \cdot \mathbf{u}_1 &\equiv 0 \pmod{1} & \mathbf{hR}_s \cdot \mathbf{u}_2 &\equiv 0 \pmod{1} & \mathbf{hR}_s \cdot \mathbf{u}_3 &\equiv 0 \pmod{1} \dots \\ \mathbf{kR}_s \cdot \mathbf{u}_1 &\equiv 0 \pmod{1} & \mathbf{kR}_s \cdot \mathbf{u}_2 &\equiv 0 \pmod{1} & \mathbf{kR}_s \cdot \mathbf{u}_3 &\equiv 0 \pmod{1} \dots \\ (\mathbf{h}-\mathbf{k})\mathbf{R}_s \cdot \mathbf{u}_1 &\equiv 0 \pmod{1} & (\mathbf{h}-\mathbf{k})\mathbf{R}_s \cdot \mathbf{u}_2 &\equiv 0 \pmod{1} & & \\ & & (\mathbf{h}-\mathbf{k})\mathbf{R}_s \cdot \mathbf{u}_3 &\equiv 0 \pmod{1} & & \dots \end{aligned}$$

are simultaneously satisfied when  $s$  varies from 1 to  $m$ . The above formulae have been generalized (Casarano *et al.*, 1988b) to the case in which deviations both of replacive and of displacive type from ideal pseudo-translational symmetry occur.

### 2.2.5.5. Quartet phase relationships

In early papers (Hauptman & Karle, 1953; Simerska, 1956) the phase

## 2.2. DIRECT METHODS

$$\Phi = \varphi_h + \varphi_k + \varphi_l - \varphi_{h+k+l}$$

was always expected to be zero. Schenk (1973*a,b*) [see also Hauptman (1974)] suggested that  $\Phi$  primarily depends on the seven magnitudes:  $R_h, R_k, R_l, R_{h+k+l}$ , called *basis magnitudes*, and  $R_{h+k}, R_{h+l}, R_{k+l}$ , called *cross magnitudes*.

The conditional probability of  $\Phi$  in  $P_1$  given seven magnitudes ( $R_1 = R_h, \dots, R_4 = R_{h+k+l}, R_5 = R_{h+k}, R_6 = R_{h+l}, R_7 = R_{k+l}$ ) according to Hauptman (1975) is

$$P_7(\Phi) = \frac{1}{L} \exp(-2B \cos \Phi) I_0(2\sigma_3 \sigma_2^{-3/2} R_5 Y_5) \\ \times I_0(2\sigma_3 \sigma_2^{-3/2} R_6 Y_6) I_0(2\sigma_3 \sigma_2^{-3/2} R_7 Y_7),$$

where  $L$  is a suitable normalizing constant which can be derived numerically,

$$B = \sigma_2^{-3} (3\sigma_3^2 - \sigma_2 \sigma_4) R_1 R_2 R_3 R_4 \\ Y_5 = [R_1^2 R_2^2 + R_3^2 R_4^2 + 2R_1 R_2 R_3 R_4 \cos \Phi]^{1/2} \\ Y_6 = [R_3^2 R_1^2 + R_2^2 R_4^2 + 2R_1 R_2 R_3 R_4 \cos \Phi]^{1/2} \\ Y_7 = [R_2^2 R_3^2 + R_1^2 R_4^2 + 2R_1 R_2 R_3 R_4 \cos \Phi]^{1/2}.$$

For equal atoms  $\sigma_2^{-3} (3\sigma_3^2 - \sigma_2 \sigma_4) = 2/N$ . Denoting

$$C = R_1 R_2 R_3 R_4 / N, \\ Z_5 = 2Y_5 / \sqrt{N}, \quad Z_6 = 2Y_6 / \sqrt{N}, \quad Z_7 = 2Y_7 / \sqrt{N}$$

gives

$$P_7(\Phi) = \frac{1}{L} \exp(-4C \cos \Phi) \\ \times I_0(R_5 Z_5) I_0(R_6 Z_6) I_0(R_7 Z_7). \quad (2.2.5.18)$$

Fig. 2.2.5.3 shows the distribution (2.2.5.18) for three typical cases. It is clear from the figure that the cosine estimated near  $\pi$  or in the middle range will be in poorer agreement with the true values than the cosine near 0 because of the relatively larger values of the variance. In principle, however, the formula is able to estimate negative or enantiomorph-sensitive quartet cosines from the seven magnitudes.

In the cs. case (2.2.5.18) is replaced (Hauptman & Green, 1976) by

$$P^\pm \simeq \frac{1}{L} \exp(\mp 2C) \cosh(R_5 Z_5^\pm) \\ \times \cosh(R_6 Z_6^\pm) \cosh(R_7 Z_7^\pm), \quad (2.2.5.19) \quad \text{or}$$

where  $P^\pm$  is the probability that the sign of  $E_1 E_2 E_3 E_4$  is positive or negative, and

$$Z_5^\pm = \frac{1}{N^{1/2}} (R_1 R_2 \pm R_3 R_4), \\ Z_6^\pm = \frac{1}{N^{1/2}} (R_1 R_3 \pm R_2 R_4), \\ Z_7^\pm = \frac{1}{N^{1/2}} (R_1 R_4 \pm R_2 R_3).$$

The normalized probability may be derived by  $P^+ / (P^+ + P^-)$ . More simple probabilistic formulae were derived independently by Giacovazzo (1975, 1976):

$$P_7(\Phi) = [2\pi I_0(G)]^{-1} \exp(G \cos \Phi), \quad (2.2.5.20)$$

where

$$G = \frac{2C(1 + \varepsilon_5 + \varepsilon_6 + \varepsilon_7)}{1 + Q/(2N)} \quad (2.2.5.21)$$

$$Q = (\varepsilon_1 \varepsilon_2 + \varepsilon_3 \varepsilon_4) \varepsilon_5 + (\varepsilon_1 \varepsilon_3 + \varepsilon_2 \varepsilon_4) \varepsilon_6 + (\varepsilon_1 \varepsilon_4 + \varepsilon_2 \varepsilon_3) \varepsilon_7 \quad (2.2.5.22)$$

and  $\varepsilon_i = (|E_i|^2 - 1)$ .  $Q$  is never allowed to be negative.

According to (2.2.5.20)  $\cos \Phi$  is expected to be positive or negative according to whether  $(\varepsilon_5 + \varepsilon_6 + \varepsilon_7 + 1)$  is positive or negative: the larger is  $C$ , the more reliable is the phase indication. For  $N \geq 150$ , (2.2.5.18) and (2.2.5.20) are practically equivalent in all cases. If  $N$  is small, (2.2.5.20) is in good agreement with (2.2.5.18) for quartets strongly defined as positive or negative, but in poor agreement for enantiomorph-sensitive quartets (see Fig. 2.2.5.3).

In cs. cases the sign probability for  $E_1 E_2 E_3 E_4$  is

$$P^+ = \frac{1}{2} + \frac{1}{2} \tanh(G/2), \quad (2.2.5.23)$$

where  $G$  is defined by (2.2.5.21).

All three cross magnitudes are not always in the set of measured reflections. From marginal distributions the following formulae arise (Giacovazzo, 1977*c*; Heinemann, 1977*b*):

(a) in the ncs. case, if  $R_7$ , or  $R_6$  and  $R_7$ , or  $R_5$  and  $R_6$  and  $R_7$ , are not in the measurements, then (2.2.5.18) is replaced by

$$P(\Phi | R_1, \dots, R_6) \simeq \frac{1}{L} \exp(-2C \cos \Phi) I_0(R_5 Z_5) I_0(R_6 Z_6),$$

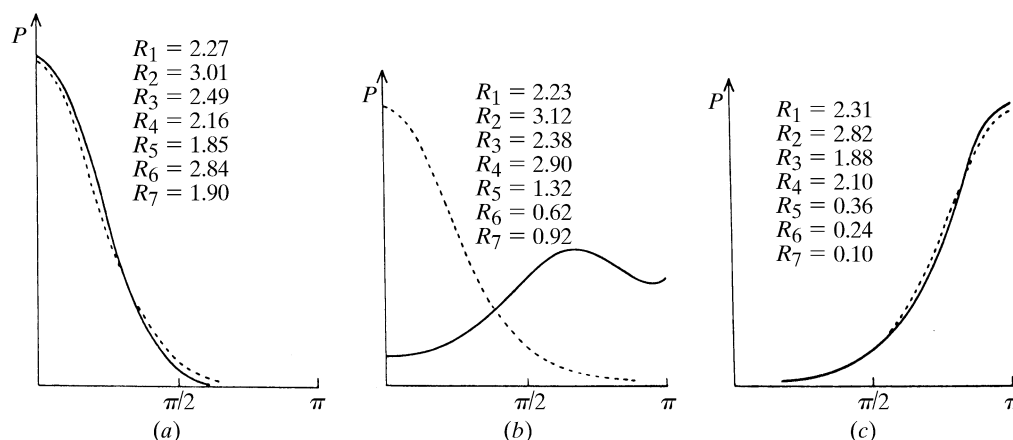


Fig. 2.2.5.3. Distributions (2.2.5.18) (—) and (2.2.5.20) (---) for the indicated  $|E|$  values in three typical cases.

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

Table 2.2.5.1. List of quartets symmetry equivalent to  $\Phi = \Phi_1$  in the class  $mmm$

Quartets	Basis vectors				Cross vectors		
$\Phi_1$	(1, 2, 3)	( $\bar{1}$ , 5, 3)	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, 7, 0)	(0, $\bar{3}$ , 11)	( $\bar{2}$ , 0, 5)
$\Phi_2$	( $\bar{1}$ , 2, 3)	(1, 5, 3)	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, 7, 0)	( $\bar{2}$ , $\bar{3}$ , 11)	(0, 0, 5)
$\Phi_3$	(1, 2, $\bar{3}$ )	( $\bar{1}$ , 5, 3)	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, 7, 0)	(0, $\bar{3}$ , 5)	( $\bar{2}$ , 0, 11)
$\Phi_4$	( $\bar{1}$ , 2, $\bar{3}$ )	(1, 5, 3)	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, 7, 0)	( $\bar{2}$ , $\bar{3}$ , 5)	(0, 0, 11)
$\Phi_5$	( $\bar{1}$ , 2, 3)	( $\bar{1}$ , 5, $\bar{3}$ )	(1, $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	( $\bar{2}$ , 7, 0)	(0, $\bar{3}$ , 11)	(0, 0, 5)
$\Phi_6$	(1, 2, 3)	( $\bar{1}$ , 5, $\bar{3}$ )	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, $\bar{3}$ , 0)	(0, 7, 11)	( $\bar{2}$ , 0, 5)
$\Phi_7$	( $\bar{1}$ , 2, 3)	(1, 5, $\bar{3}$ )	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, $\bar{3}$ , 0)	( $\bar{2}$ , 7, 11)	(0, 0, 5)
$\Phi_8$	( $\bar{1}$ , 2, $\bar{3}$ )	( $\bar{1}$ , 5, 3)	(1, $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	( $\bar{2}$ , 7, 0)	(0, $\bar{3}$ , 5)	(0, 0, 11)
$\Phi_9$	(1, 2, $\bar{3}$ )	( $\bar{1}$ , 5, 3)	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, $\bar{3}$ , 0)	(0, 7, 5)	( $\bar{2}$ , 0, 11)
$\Phi_{10}$	( $\bar{1}$ , 2, $\bar{3}$ )	(1, 5, 3)	( $\bar{1}$ , $\bar{5}$ , 8)	(1, $\bar{2}$ , $\bar{8}$ )	(0, $\bar{3}$ , 0)	( $\bar{2}$ , 7, 5)	(0, 0, 11)
$\Phi_{11}$	( $\bar{1}$ , 2, 3)	( $\bar{1}$ , 5, $\bar{3}$ )	(1, 5, 8)	(1, $\bar{2}$ , $\bar{8}$ )	( $\bar{2}$ , $\bar{3}$ , 0)	(0, 7, 11)	(0, 0, 5)

$$P(\Phi|R_1, \dots, R_5) \simeq \frac{1}{L''} I_0(R_5 Z_5),$$

or

$$P(\Phi|R_1, \dots, R_4) \simeq \frac{1}{L'''} \exp(2C \cos \Phi),$$

respectively.

(b) in the same situations, we have for cs. cases

$$P^\pm \simeq \frac{1}{L'} \exp(\mp C) \cosh(R_5 Z_5^\pm) \cosh(R_6 Z_6^\pm),$$

or

$$P^\pm \simeq \frac{1}{L'''} \cosh(R_5 Z_5^\pm)$$

or

$$P^\pm = \frac{1}{L'''} \exp(\pm C) \simeq 0.5 + 0.5 \tanh(\pm C),$$

respectively.

Equations (2.2.5.20) and (2.2.5.23) are easily modifiable when some cross magnitudes are not in the measurements. If  $R_i$  is not measured then (2.2.5.20) or (2.2.5.23) are still valid provided that in  $G$  it is assumed that  $\varepsilon_i = 0$ . For example, if  $R_7$  and  $R_6$  are not in the data then (2.2.5.21) and (2.2.5.22) become

$$G = \frac{2C(1 + \varepsilon_5)}{1 + Q/(2N)}, \quad Q = (\varepsilon_1 \varepsilon_2 + \varepsilon_3 \varepsilon_4) \varepsilon_5.$$

In space groups with symmetry higher than  $P\bar{1}$  more symmetry-equivalent quartets can exist of the type

$$\psi = \varphi_{\mathbf{h}R_\alpha} + \varphi_{\mathbf{k}R_\beta} + \varphi_{\mathbf{l}R_\gamma} + \varphi_{\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}}R_\delta},$$

where  $R_\alpha, R_\beta, R_\gamma, R_\delta$  are rotation matrices of the space group. The set  $\{\psi\}$  is called the *first representation* of  $\Phi$ . In this case  $\Phi$  primarily depends on more than seven magnitudes. For example, let us consider in  $Pmmm$  the quartet

$$\Phi = \varphi_{123} + \varphi_{1\bar{5}\bar{3}} + \varphi_{1\bar{5}8} + \varphi_{1\bar{2}\bar{8}}.$$

Quartets symmetry equivalent to  $\Phi$  and respective cross terms are given in Table 2.2.5.1.

Experimental tests on the application of the representation concept to quartets have recently been made (Busetta *et al.*, 1980). It was shown that quartets with more than three cross magnitudes are more accurately estimated than other quartets. Also, quartets with a cross reflection which is systematically absent were shown to be of significant importance in direct methods. In this

context it is noted that systematically absent reflections are not usually included in the set of diffraction data. This custom, not exceptionable when only triplet relations are used, can give rise to a loss of information when quartets are used. In fact the usual programs of direct methods discard quartets as soon as one of the cross reflections is not measured, so that systematic absences are dealt with in the same manner as those reflections which are outside the sphere of measurements.

### 2.2.5.6. Quintet phase relationships

A quintet phase

$$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} + \varphi_{\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}}}$$

may be considered as the sum of three suitable triplets or the sum of a triplet and a quartet, *i.e.*

$$\begin{aligned} \Phi = & (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}}) + (\varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} - \varphi_{\mathbf{l}+\mathbf{m}}) \\ & + (\varphi_{\mathbf{h}+\mathbf{k}} + \varphi_{\mathbf{l}+\mathbf{m}} + \varphi_{\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}}}) \end{aligned}$$

or

$$\Phi = (\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}}) + (\varphi_{\mathbf{l}} + \varphi_{\mathbf{m}} + \varphi_{\overline{\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}}} + \varphi_{\mathbf{h}+\mathbf{k}}).$$

It depends primarily on 15 magnitudes: the five *basis magnitudes*

$$R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{l}}, R_{\mathbf{m}}, R_{\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m}},$$

and the ten *cross magnitudes*

$$\begin{aligned} & R_{\mathbf{h}+\mathbf{k}}, R_{\mathbf{h}+\mathbf{l}}, R_{\mathbf{h}+\mathbf{m}}, R_{\mathbf{k}+\mathbf{l}+\mathbf{m}}, R_{\mathbf{k}+\mathbf{l}}, \\ & R_{\mathbf{k}+\mathbf{m}}, R_{\mathbf{h}+\mathbf{l}+\mathbf{m}}, R_{\mathbf{l}+\mathbf{m}}, R_{\mathbf{h}+\mathbf{k}+\mathbf{m}}, R_{\mathbf{h}+\mathbf{k}+\mathbf{l}}. \end{aligned}$$

In the following we will denote

$$R_1 = R_{\mathbf{h}}, R_2 = R_{\mathbf{k}}, \dots, R_{15} = R_{\mathbf{h}+\mathbf{k}+\mathbf{l}}.$$

Conditional distributions of  $\Phi$  in  $P1$  and  $P\bar{1}$  given the 15 magnitudes have been derived by several authors and allow in favourable circumstances in ncs. space groups the quintets having  $\Phi$  near 0 or near  $\pi$  or near  $\pm\pi/2$  to be identified. Among others, we remember:

(a) the semi-empirical expression for  $P_{15}(\Phi)$  suggested by Van der Putten & Schenk (1977):

$$P(\Phi|\dots) \simeq \frac{1}{L} \exp \left[ \left( 6 - \sum_{j=6}^{15} R_j^2 \right) 2C \cos \Phi \right] \prod_{j=6}^{15} I_0(2R_j Y_j),$$

where

$$C = N^{-3/2} R_1 R_2 R_3 R_4 R_5$$

## 2.2. DIRECT METHODS

and  $Y_j$  is an expression related to the  $j$ th of the ten quartets connected with the quintet  $\Phi$ ;

(b) the formula by Fortier & Hauptman (1977), valid in  $\bar{P}\bar{1}$ , which is able to predict the sign of a quintet by means of an expression which involves a summation over 1024 sets of signs;

(c) the expression by Giacovazzo (1977d), according to which

$$P_{15}(\Phi) \simeq [2\pi I_0(G)]^{-1} \exp(G \cos \Phi), \quad (2.2.5.24)$$

where

$$G = \frac{2C}{1 + 6(N)^{1/2}} \left[ \frac{1 + A + B}{1 + D/(2N)} \right] \quad (2.2.5.25)$$

and where

$$A = \sum_{i=6}^{15} \varepsilon_i,$$

$$B = \varepsilon_6\varepsilon_{13} + \varepsilon_6\varepsilon_{15} + \varepsilon_6\varepsilon_{14} + \varepsilon_7\varepsilon_{11} + \varepsilon_7\varepsilon_{15} + \varepsilon_7\varepsilon_{12} \\ + \varepsilon_8\varepsilon_{10} + \varepsilon_8\varepsilon_{14} + \varepsilon_8\varepsilon_{12} + \varepsilon_{10}\varepsilon_{15} + \varepsilon_{10}\varepsilon_9 + \varepsilon_{11}\varepsilon_{14} \\ + \varepsilon_{11}\varepsilon_9 + \varepsilon_{13}\varepsilon_9 + \varepsilon_{13}\varepsilon_{12},$$

$$D = \varepsilon_1\varepsilon_2\varepsilon_6 + \varepsilon_1\varepsilon_3\varepsilon_7 + \varepsilon_1\varepsilon_4\varepsilon_8 + \varepsilon_1\varepsilon_5\varepsilon_9 + \varepsilon_1\varepsilon_{10}\varepsilon_{15} \\ + \varepsilon_1\varepsilon_{11}\varepsilon_{14} + \varepsilon_1\varepsilon_{13}\varepsilon_{12} + \varepsilon_2\varepsilon_3\varepsilon_{10} + \varepsilon_2\varepsilon_4\varepsilon_{11} \\ + \varepsilon_2\varepsilon_5\varepsilon_{12} + \varepsilon_2\varepsilon_7\varepsilon_{15} + \varepsilon_2\varepsilon_8\varepsilon_{14} + \varepsilon_2\varepsilon_{13}\varepsilon_9 + \varepsilon_3\varepsilon_4\varepsilon_{13} \\ + \varepsilon_3\varepsilon_5\varepsilon_{14} + \varepsilon_3\varepsilon_6\varepsilon_{15} + \varepsilon_3\varepsilon_8\varepsilon_{12} + \varepsilon_3\varepsilon_{11}\varepsilon_9 + \varepsilon_4\varepsilon_5\varepsilon_{15} \\ + \varepsilon_4\varepsilon_6\varepsilon_{14} + \varepsilon_4\varepsilon_7\varepsilon_{12} + \varepsilon_4\varepsilon_{10}\varepsilon_9 + \varepsilon_5\varepsilon_6\varepsilon_{13} + \varepsilon_5\varepsilon_7\varepsilon_{11} \\ + \varepsilon_5\varepsilon_8\varepsilon_{10}.$$

For cs. cases (2.2.5.24) reduces to

$$P^+ \simeq 0.5 + 0.5 \tanh(G/2). \quad (2.2.5.26)$$

Positive or negative quintets may be identified according to whether  $G$  is larger or smaller than zero.

If  $R_i$  is not measured then (2.2.5.24) and (2.2.5.25) are still valid provided that in (2.2.5.25)  $\varepsilon_i = 0$ .

If the symmetry is higher than in  $\bar{P}\bar{1}$  then more symmetry-equivalent quintets can exist of the type

$$\psi = \varphi_{\mathbf{h}R_\alpha} + \varphi_{\mathbf{k}R_\beta} + \varphi_{\mathbf{l}R_\gamma} + \varphi_{\mathbf{m}R_\delta} + \varphi_{\overline{(\mathbf{h}+\mathbf{k}+\mathbf{l}+\mathbf{m})}R_\varepsilon},$$

where  $R_\alpha, \dots, R_\varepsilon$  are rotation matrices of the space groups. The set  $\{\psi\}$  is called the first representation of  $\Phi$ . In this case  $\Phi$  primarily depends on more than 15 magnitudes which all have to be taken into account for a careful estimation of  $\Phi$  (Giacovazzo, 1980a).

A wide use of quintet invariants in direct methods procedures is prevented for two reasons: (a) the large correlation of positive quintet cosines with positive triplets; (b) the large computing time necessary for their estimation [quintets are phase relationships of order  $1/(N\sqrt{N})$ , so a large number of quintets have to be estimated in order to pick up a sufficient percentage of reliable ones].

### 2.2.5.7. Determinantal formulae

In a crystal structure with  $N$  identical atoms the joint probability distribution of  $n$  normalized s.f.'s  $E_{\mathbf{h}_1+\mathbf{k}}, E_{\mathbf{h}_2+\mathbf{k}}, \dots, E_{\mathbf{h}_n+\mathbf{k}}$  under the following conditions:

(a) the structure is kept fixed whereas  $\mathbf{k}$  is the primitive random variable;

(b)  $E_{\mathbf{h}_i-\mathbf{h}_j}$ ,  $i, j = 1, \dots, n$ , have values which are known *a priori*; is given (Tsoucaris, 1970) [see also Castellano *et al.* (1973) and Heinermann *et al.* (1979)] by

$$P(E_1, E_2, \dots, E_n) = (2\pi)^{-n/2} D_n^{-1/2} \exp(-\frac{1}{2}Q_n) \quad (2.2.5.27)$$

for cs. structures and

$$P(E_1, E_2, \dots, E_n) = (2\pi)^{-n} D_n^{-1/2} \exp(-Q_n) \quad (2.2.5.28)$$

for ncs. structures. In (2.2.5.27) and (2.2.5.28) we have denoted

$$D_n = \lambda, \quad Q_n = \sum_{p, q=1}^n \Lambda_{pq} E_p E_q^*$$

$$E_j = E_{\mathbf{h}_j+\mathbf{k}}, \quad U_{pq} = U_{\mathbf{h}_p-\mathbf{h}_q}, \quad j, p, q = 1, \dots, n.$$

$\Lambda_{pq}$  is an element of  $\lambda^{-1}$ , and  $\lambda$  is the covariance matrix with elements

$$\langle E_{\mathbf{h}_p+\mathbf{k}} E_{\mathbf{h}_q+\mathbf{k}} \rangle = U_{\mathbf{h}_p-\mathbf{h}_q}$$

$$\lambda = \begin{vmatrix} 1 & U_{12} & \dots & U_{1q} & \dots & U_{1n} \\ U_{21} & 1 & \dots & U_{2q} & \dots & U_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{p1} & U_{p2} & \dots & U_{pq} & \dots & U_{pn} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ U_{n1} & U_{n2} & \dots & U_{nq} & \dots & 1 \end{vmatrix}.$$

$\lambda$  is a K-H determinant: therefore  $D_n \geq 0$ . Let us call

$$\Delta_{n+1} = \frac{1}{N} \begin{vmatrix} 1 & U_{12} & \dots & U_{1n} & E_{\mathbf{h}_1+\mathbf{k}} \\ U_{21} & 1 & \dots & U_{2n} & E_{\mathbf{h}_2+\mathbf{k}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ U_{n1} & U_{n2} & \dots & 1 & E_{\mathbf{h}_n+\mathbf{k}} \\ E_{-\mathbf{h}_1-\mathbf{k}} & E_{-\mathbf{h}_2-\mathbf{k}} & \dots & E_{-\mathbf{h}_n-\mathbf{k}} & N \end{vmatrix};$$

the K-H determinant obtained by adding to  $\lambda$  the last column and line formed by  $E_1, E_2, \dots, E_n$ , and  $E_1^*, E_2^*, \dots, E_n^*$ , respectively. Then (2.2.5.27) and (2.2.5.28) may be written

$$P(E_1, E_2, \dots, E_n) \\ = (2\pi)^{-n/2} D_n^{-1/2} \exp \left[ N \frac{\Delta_{n+1} - D_n}{2D_n} \right] \quad (2.2.5.29)$$

and

$$P(E_1, E_2, \dots, E_n) \\ = (2\pi)^{-n} D_n^{-1/2} \exp \left[ N \frac{\Delta_{n+1} - D_n}{D_n} \right], \quad (2.2.5.30)$$

respectively. Because  $D_n$  is a constant, the maximum values of the conditional joint probabilities (2.2.5.29) and (2.2.5.30) are obtained when  $\Delta_{n+1}$  is a maximum. Thus the *maximum determinant rule* may be stated (Tsoucaris, 1970; Lajz rowicz & Lajz rowicz, 1966): among all sets of phases which are compatible with the inequality

$$\Delta_{n+1}(E_1, E_2, \dots, E_n) \geq 0$$

the most probable one is that which leads to a maximum value of  $\Delta_{n+1}$ .

If only one phase, *i.e.*  $\varphi_q$ , is unknown whereas all other phases and moduli are known then (de Rango *et al.*, 1974; Podjarny *et al.*, 1976) for cs. crystals

$$P^\pm(E_q) \simeq 0.5 + 0.5 \tanh \left\{ \pm |E_q| \sum_{\substack{p=1 \\ p \neq q}}^n \Lambda_{pq} E_p \right\}, \quad (2.2.5.31)$$

and for ncs. crystals

$$P(\varphi_q) = [2\pi I_0(G_q)]^{-1} \exp\{G_q \cos(\varphi_q - \theta_q)\}, \quad (2.2.5.32)$$

where

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

$$G_q \exp(i\theta_q) = 2|E_q| \sum_{p \neq q=1}^n \Lambda_{pq} E_p.$$

Equations (2.2.5.31) and (2.2.5.32) generalize (2.2.5.11) and (2.2.5.7), respectively, and reduce to them for  $n = 3$ . Fourth-order determinantal formulae estimating triplet invariants in cs. and ncs. crystals, and making use of the entire data set, have recently been secured (Karle, 1979, 1980).

Advantages, limitations and applications of determinantal formulae can be found in the literature (Heinermann *et al.*, 1979; de Rango *et al.*, 1975, 1985). Taylor *et al.* (1978) combined K–H determinants with a magic-integer approach. The computing time, however, was larger than that required by standard computing techniques. The use of K–H matrices has been made faster and more effective by de Gelder *et al.* (1990) (see also de Gelder, 1992). They developed a phasing procedure (*CRUNCH*) which uses random phases as starting points for the maximization of the K–H determinants.

### 2.2.5.8. Algebraic relationships for structure seminvariants

According to the representations method (Giacovazzo, 1977a, 1980a,b):

(i) any s.s.  $\Phi$  may be estimated *via* one or more s.i.'s  $\{\psi\}$ , whose values differ from  $\Phi$  by a constant arising because of symmetry;

(ii) two types of s.s.'s exist, first-rank and second-rank s.s.'s, with different algebraic properties;

(iii) conditions characterizing s.s.'s of first rank for any space group may be expressed in terms of seminvariant moduli and seminvariantly associated vectors. For example, for all the space groups with point group 422 [Hauptman–Karle group  $(h+k, l) P(2, 2)$ ] the one-phase s.s.'s of first rank are characterized by

$$(h, k, l) \equiv 0 \pmod{(2, 2, 0)} \text{ or } (2, 0, 2) \text{ or } (0, 2, 2)$$

$$(h \pm k, l) \equiv 0 \pmod{(0, 2)} \text{ or } (2, 0).$$

The more general expressions for the s.s.'s of first rank are

(a)  $\Phi = \varphi_{\mathbf{u}} = \varphi_{\mathbf{h}(\mathbf{I} - \mathbf{R}_\alpha)}$  for one-phase s.s.'s;

(b)  $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} = \varphi_{\mathbf{h}_1 - \mathbf{h}_2 \mathbf{R}_\beta} + \varphi_{\mathbf{h}_2 - \mathbf{h}_1 \mathbf{R}_\alpha}$  for two-phase s.s.'s;

(c)  $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} + \varphi_{\mathbf{u}_3} = \varphi_{\mathbf{h}_1 - \mathbf{h}_2 \mathbf{R}_\beta} + \varphi_{\mathbf{h}_2 - \mathbf{h}_3 \mathbf{R}_\gamma} + \varphi_{\mathbf{h}_3 - \mathbf{h}_1 \mathbf{R}_\alpha}$  for three-phase s.s.'s;

$$(d) \Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2} + \varphi_{\mathbf{u}_3} + \varphi_{\mathbf{u}_4}$$

$$= \varphi_{\mathbf{h}_1 - \mathbf{h}_2 \mathbf{R}_\beta} + \varphi_{\mathbf{h}_2 - \mathbf{h}_3 \mathbf{R}_\gamma} + \varphi_{\mathbf{h}_3 - \mathbf{h}_4 \mathbf{R}_\delta} + \varphi_{\mathbf{h}_4 - \mathbf{h}_1 \mathbf{R}_\alpha}$$

for four-phase s.s.'s; *etc.*

In other words:

(a)  $\varphi_{\mathbf{u}}$  is an s.s. of first rank if at least one  $\mathbf{h}$  and at least one rotation matrix  $\mathbf{R}_\alpha$  exist such that  $\mathbf{u} = \mathbf{h}(\mathbf{I} - \mathbf{R}_\alpha)$ .  $\varphi_{\mathbf{u}}$  may be estimated *via* the special triplet invariants

$$\{\psi\} = \varphi_{\mathbf{u}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\mathbf{R}_\alpha}. \quad (2.2.5.33)$$

The set  $\{\psi\}$  is called the *first representation* of  $\varphi_{\mathbf{u}}$ .

(b)  $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$  is an s.s. of first rank if at least two vectors  $\mathbf{h}_1$  and  $\mathbf{h}_2$  and two rotation matrices  $\mathbf{R}_\alpha$  and  $\mathbf{R}_\beta$  exist such that

$$\begin{cases} \mathbf{u}_1 = \mathbf{h}_1 - \mathbf{h}_2 \mathbf{R}_\beta \\ \mathbf{u}_2 = \mathbf{h}_2 - \mathbf{h}_1 \mathbf{R}_\alpha. \end{cases} \quad (2.2.5.34)$$

$\Phi$  may then be estimated *via* the special quartet invariants

$$\{\psi\} = \varphi_{\mathbf{u}_1 \mathbf{R}_\alpha} + \varphi_{\mathbf{u}_2} - \varphi_{\mathbf{h}_2} + \varphi_{\mathbf{h}_2 \mathbf{R}_\beta \mathbf{R}_\alpha} \quad (2.2.5.35a)$$

and

$$\{\psi\} = \{\varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2 \mathbf{R}_\beta} - \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_1 \mathbf{R}_\alpha \mathbf{R}_\beta}\}. \quad (2.2.5.35b)$$

For example,  $\Phi = \varphi_{123} + \varphi_{7\bar{2}5}$  in  $P2_1$  may be estimated *via*

$$\{\psi\} = \varphi_{123} + \varphi_{7\bar{2}5} - \varphi_{3K\bar{1}} + \varphi_{3K1}$$

and

$$\{\psi\} = \varphi_{123} + \varphi_{7\bar{2}5} - \varphi_{4K4} + \varphi_{4K\bar{4}},$$

where  $K$  is a free index.

The set of special quartets (2.2.5.35a) and (2.2.5.35b) constitutes the *first representations* of  $\Phi$ .

Structure seminvariants of the second rank can be characterized as follows: suppose that, for a given seminvariant  $\Phi$ , it is not possible to find a vectorial index  $\mathbf{h}$  and a rotation matrix  $\mathbf{R}_\alpha$  such that  $\Phi - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\mathbf{R}_\alpha}$  is a structure invariant. Then  $\Phi$  is a structure seminvariant of the second rank and a set of structure invariants  $\psi$  can certainly be formed, of type

$$\{\psi\} = \Phi + \varphi_{\mathbf{h}\mathbf{R}_p} - \varphi_{\mathbf{h}\mathbf{R}_q} + \varphi_{\mathbf{l}\mathbf{R}_r} - \varphi_{\mathbf{l}\mathbf{R}_s},$$

by means of suitable indices  $\mathbf{h}$  and  $\mathbf{l}$  and rotation matrices  $\mathbf{R}_p, \mathbf{R}_q, \mathbf{R}_r$  and  $\mathbf{R}_s$ . As an example, for symmetry class 222,  $\varphi_{240}$  or  $\varphi_{024}$  or  $\varphi_{204}$  are s.s.'s of the first rank while  $\varphi_{246}$  is an s.s. of the second rank.

The procedure may easily be generalized to s.s.'s of any order of the first and of the second rank. So far only the role of one-phase and two-phase s.s.'s of the first rank in direct procedures is well documented (see references quoted in Sections 2.2.5.9 and 2.2.5.10).

### 2.2.5.9. Formulae estimating one-phase structure seminvariants of the first rank

Let  $E_{\mathbf{H}}$  be our one-phase s.s. of the first rank, where

$$\mathbf{H} = \mathbf{h}(\mathbf{I} - \mathbf{R}_n). \quad (2.2.5.36)$$

In general, more than one rotation matrix  $\mathbf{R}_n$  and more than one vector  $\mathbf{h}$  are compatible with (2.2.5.36). The set of special triplets

$$\{\psi\} = \{\varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\mathbf{R}_n}\}$$

is the first representation of  $E_{\mathbf{H}}$ . In cs. space groups the probability that  $E_{\mathbf{H}} > 0$ , given  $|E_{\mathbf{H}}|$  and the set  $\{|E_{\mathbf{h}}|\}$ , may be estimated (Hauptman & Karle, 1953; Naya *et al.*, 1964; Cochran & Woolfson, 1955) by

$$P^+(E_{\mathbf{H}}) \simeq 0.5 + 0.5 \tanh \sum_{\mathbf{h}, n} G_{\mathbf{h}, n} (-1)^{2\mathbf{h} \cdot \mathbf{T}_n}, \quad (2.2.5.37)$$

where

$$G_{\mathbf{h}, n} = |E_{\mathbf{H}}| \varepsilon_{\mathbf{h}} / (2\sqrt{N}), \text{ and } \varepsilon = |E|^2 - 1.$$

In (2.2.5.37), the summation over  $n$  goes within the set of matrices  $\mathbf{R}_n$  for which (2.2.5.35a,b) is compatible, and  $\mathbf{h}$  varies within the set of vectors which satisfy (2.2.5.36) for each  $\mathbf{R}_n$ . Equation (2.2.5.36) is actually a generalized way of writing the so-called  $\sum_1$  relationships (Hauptman & Karle, 1953).

If  $\varphi_{\mathbf{H}}$  is a phase restricted by symmetry to  $\theta_{\mathbf{H}}$  and  $\theta_{\mathbf{H}} + \pi$  in an ncs. space group then (Giacovazzo, 1978)

$$P(\varphi_{\mathbf{H}} = \theta_{\mathbf{H}}) \simeq 0.5 + 0.5 \tanh \left\{ \sum_{\mathbf{h}, n} G_{\mathbf{h}, n} \cos(\theta_{\mathbf{H}} - 2\pi \mathbf{h} \cdot \mathbf{T}_n) \right\}. \quad (2.2.5.38)$$

If  $\varphi_{\mathbf{H}}$  is a general phase then  $\varphi_{\mathbf{H}}$  is distributed according to

$$P(\varphi_{\mathbf{H}}) \simeq \frac{1}{L} \exp\{\alpha \cos(\varphi_{\mathbf{H}} - \theta_{\mathbf{H}})\},$$

where



## 2.2. DIRECT METHODS

$$\tan \theta_{\mathbf{H}} = \frac{\left( \sum_{\mathbf{h}, n} G_{\mathbf{h}, n} \sin 2\pi \mathbf{h} \cdot \mathbf{T}_n \right)}{\left( \sum_{\mathbf{h}, n} G_{\mathbf{h}, n} \cos 2\pi \mathbf{h} \cdot \mathbf{T}_n \right)} \quad (2.2.5.39)$$

with a reliability measured by

$$\alpha = \left\{ \left( \sum_{\mathbf{h}, n} G_{\mathbf{h}, n} \sin 2\pi \mathbf{h} \cdot \mathbf{T}_n \right)^2 + \left( \sum_{\mathbf{h}, n} G_{\mathbf{h}, n} \cos 2\pi \mathbf{h} \cdot \mathbf{T}_n \right)^2 \right\}^{1/2}.$$

The second representation of  $\varphi_{\mathbf{H}}$  is the set of special quintets

$$\{\psi\} = \{\varphi_{\mathbf{H}} - \varphi_{\mathbf{h}} + \varphi_{\mathbf{h}\mathbf{R}_n} + \varphi_{\mathbf{k}\mathbf{R}_j} - \varphi_{\mathbf{k}\mathbf{R}_j}\} \quad (2.2.5.40)$$

provided that  $\mathbf{h}$  and  $\mathbf{R}_n$  vary over the vectors and matrices for which (2.2.5.36) is compatible,  $\mathbf{k}$  over the asymmetric region of the reciprocal space, and  $\mathbf{R}_j$  over the rotation matrices in the space group. Formulae estimating  $\varphi_{\mathbf{H}}$  via the second representation in all the space groups [all the base and cross magnitudes of the quintets (2.2.5.40) now constitute the *a priori* information] have recently been secured (Giacovazzo, 1978; Cascarano & Giacovazzo, 1983; Cascarano, Giacovazzo, Calabrese *et al.*, 1984). Such formulae contain, besides the contribution of order  $N^{-1/2}$  provided by the first representation, a supplementary (not negligible) contribution of order  $N^{-3/2}$  arising from quintets.

Denoting

$$E_1 = E_{\mathbf{H}}, E_2 = E_{\mathbf{h}}, E_3 = E_{\mathbf{k}}, \\ E_{4,j} = E_{\mathbf{h}+\mathbf{k}\mathbf{R}_j}, E_{5,j} = E_{\mathbf{H}+\mathbf{k}\mathbf{R}_j},$$

formulae (2.2.5.37), (2.2.5.38), (2.2.5.39) still hold provided that  $\sum_{\mathbf{h}, n} G_{\mathbf{h}, n}$  is replaced by

$$\sum_{\mathbf{h}, n} G_{\mathbf{h}, n} + \sum'_{\mathbf{h}, \mathbf{k}, n} \frac{|E_{\mathbf{H}}|}{2N^{3/2}} \frac{A_{\mathbf{h}, \mathbf{k}, n}}{1 + B_{\mathbf{h}, \mathbf{k}, n}},$$

where

$$A_{\mathbf{h}, \mathbf{k}, n} = \left[ (2|E_2|^2 - 1) \varepsilon_3 \left( \sum_{\substack{R_j=R_j \\ R_j+R_i, R_n=0}} \varepsilon_{4,i} \varepsilon_{5,j} + \sum_{\substack{R_j=R_i, R_n \\ R_i=R_j, R_n}} \varepsilon_{4,i} \varepsilon_{4,j} \right) - \frac{\varepsilon_3}{2} \sum_{j=1}^m \varepsilon_{4,j} - \frac{1}{2} \sum_{\substack{R_j=R_i \\ R_j+R_i, R_n=0}} \varepsilon_{4,i} \varepsilon_{5,j} \right] / N, \\ B_{\mathbf{h}, \mathbf{k}, n} = \left[ \varepsilon_1 \varepsilon_3 \sum_{j=1}^m \varepsilon_{5,j} + \varepsilon_1 \sum_{\substack{R_j=R_i, R_n \\ R_i=R_j, R_n}} \varepsilon_{4,i} \varepsilon_{4,j} + \varepsilon_2 \varepsilon_3 \sum_{j=1}^m \varepsilon_{4,j} + \varepsilon_2 \sum_{\substack{R_j=R_i \\ R_j+R_i, R_n=0}} \varepsilon_{4,i} \varepsilon_{5,j} + \frac{1}{4} \varepsilon_1 H_4(E_2) \right] / (2N).$$

$m$  is the number of symmetry operators and  $H_4(E) = E^4 - 6E^2 + 3$  is the Hermite polynomial of order four.

$B_{\mathbf{h}, \mathbf{k}, n}$  is assumed to be zero if it is computed negative. The prime to the summation warns the reader that precautions have to be taken in order to avoid duplication in the contributions.

### 2.2.5.10. Formulae estimating two-phase structure seminvariants of the first rank

Two-phase s.s.'s of the first rank were first evaluated in some cs. space groups by the method of coincidence by Grant *et al.* (1957); the idea was extended to ncs. space groups by Debaerdemaeker & Woolfson (1972), and in a more general way by Giacovazzo (1977*e,f*).

The technique was based on the combination of the two triplets

$$\varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2} \simeq \varphi_{\mathbf{h}_1+\mathbf{h}_2} \\ \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2\mathbf{R}} \simeq \varphi_{\mathbf{h}_1+\mathbf{h}_2\mathbf{R}},$$

which, subtracted from one another, give

$$\varphi_{\mathbf{h}_1+\mathbf{h}_2\mathbf{R}} - \varphi_{\mathbf{h}_1+\mathbf{h}_2} \simeq \varphi_{\mathbf{h}_2\mathbf{R}} - \varphi_{\mathbf{h}_2} \simeq -2\pi \mathbf{h} \cdot \mathbf{T}.$$

If all four  $|E|$ 's are sufficiently large, an estimate of the two-phase seminvariant  $\varphi_{\mathbf{h}_1+\mathbf{h}_2\mathbf{R}} - \varphi_{\mathbf{h}_1+\mathbf{h}_2}$  is available.

Probability distributions valid in  $P2_1$  according to the neighbourhood principle have been given by Hauptman & Green (1978). Finally, the theory of representations was combined by Giacovazzo (1979*a*) with the joint probability distribution method in order to estimate two-phase s.s.'s in all the space groups.

According to representation theory, the problem is that of evaluating  $\Phi = \varphi_{\mathbf{u}_1} + \varphi_{\mathbf{u}_2}$  via the special quartets (2.2.5.35*a*) and (2.2.5.35*b*). Thus, contributions of order  $N^{-1}$  will appear in the probabilistic formulae, which will be functions of the basis and of the cross magnitudes of the quartets (2.2.5.35). Since more pairs of matrices  $\mathbf{R}_\alpha$  and  $\mathbf{R}_\beta$  can be compatible with (2.2.5.34), and for each pair ( $\mathbf{R}_\alpha, \mathbf{R}_\beta$ ) more pairs of vectors  $\mathbf{h}_1$  and  $\mathbf{h}_2$  may satisfy (2.2.5.34), several quartets can in general be exploited for estimating  $\Phi$ . The simplest case occurs in  $P\bar{1}$  where the two quartets (2.2.5.35) suggest the calculation of the six-variate distribution function ( $\mathbf{u}_1 = \mathbf{h}_1 + \mathbf{h}_2, \mathbf{u}_2 = \mathbf{h}_1 - \mathbf{h}_2$ )

$$P(E_{\mathbf{h}_1}, E_{\mathbf{h}_2}, E_{\mathbf{h}_1+\mathbf{h}_2}, E_{\mathbf{h}_1-\mathbf{h}_2}, E_{2\mathbf{h}_1}, E_{2\mathbf{h}_2})$$

which leads to the probability formula

$$P^+ \simeq 0.5 + 0.5 \tanh \left( \frac{|E_{\mathbf{h}_1+\mathbf{h}_2} E_{\mathbf{h}_1-\mathbf{h}_2}|}{2N} \cdot \frac{A}{1+B} \right),$$

where  $P^+$  is the probability that the product  $E_{\mathbf{h}_1+\mathbf{h}_2} E_{\mathbf{h}_1-\mathbf{h}_2}$  is positive, and

$$A = \varepsilon_{\mathbf{h}_1} + \varepsilon_{\mathbf{h}_2} + 2\varepsilon_{\mathbf{h}_1} \varepsilon_{\mathbf{h}_2} + \varepsilon_{\mathbf{h}_1} \varepsilon_{2\mathbf{h}_1} + \varepsilon_{\mathbf{h}_2} \varepsilon_{2\mathbf{h}_2} \\ B = (\varepsilon_{\mathbf{h}_1} \varepsilon_{\mathbf{h}_2} \varepsilon_{\mathbf{u}_1} + \varepsilon_{\mathbf{h}_1} \varepsilon_{\mathbf{h}_2} \varepsilon_{\mathbf{u}_2} + \varepsilon_{\mathbf{u}_1} \varepsilon_{\mathbf{u}_2} \varepsilon_{2\mathbf{h}_1} + \varepsilon_{\mathbf{u}_1} \varepsilon_{\mathbf{u}_2} \varepsilon_{2\mathbf{h}_2}) / (2N).$$

It may be seen that in favourable cases  $P^+ < 0.5$ .

For the sake of brevity, the probabilistic formulae for the general case are not given and the reader is referred to the original papers.

### 2.2.6. Direct methods in real and reciprocal space: Sayre's equation

The statistical treatment suggested by Wilson for scaling observed intensities corresponds, in direct space, to the origin peak of the Patterson function, so it is not surprising that a general correspondence exists between probabilistic formulation in reciprocal space and algebraic properties in direct space.

For a structure containing atoms which are fully resolved from one another, the operation of raising  $\rho(\mathbf{r})$  to the  $n$ th power retains