

2.2. DIRECT METHODS

$$\begin{vmatrix} 7 & 0 \\ 8 & 1 \end{vmatrix} \xrightarrow{\text{reduced mod } (2,0)} \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

$$E_{\mathbf{h}} = \frac{F_{\mathbf{h}}}{(\varepsilon_{\mathbf{h}} \sum_N)^{1/2}}. \tag{2.2.4.2}$$

(i) If an s.s. or an s.i. has a general value  $\varphi$  for a given structure, it will have a value  $-\varphi$  for the enantiomorph structure. If  $\varphi = 0, \pi$  the s.s. has the same value for both enantiomorphs. Once the origin has been assigned, in ncs. space groups the sign of a given s.s.  $\varphi \neq 0, \pi$  can be assigned to fix the enantiomorph. In practice it is often advisable to use an s.s. or an s.i. whose value is as near as possible to  $\pm\pi/2$ .

2.2.4. Normalized structure factors

2.2.4.1. Definition of normalized structure factor

The normalized structure factors  $E$  (see also Chapter 2.1) are calculated according to (Hauptman & Karle, 1953)

$$|E_{\mathbf{h}}|^2 = |F_{\mathbf{h}}|^2 / \langle |F_{\mathbf{h}}|^2 \rangle, \tag{2.2.4.1}$$

where  $|F_{\mathbf{h}}|^2$  is the squared observed structure-factor magnitude on the absolute scale and  $\langle |F_{\mathbf{h}}|^2 \rangle$  is the expected value of  $|F_{\mathbf{h}}|^2$ .

$\langle |F_{\mathbf{h}}|^2 \rangle$  depends on the available *a priori* information. Often, but not always, this may be considered as a combination of several typical situations. We mention:

(a) No structural information. The atomic positions are considered random variables. Then

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \sum_{j=1}^N f_j^2 = \varepsilon_{\mathbf{h}} \sum_N$$

so that

$\varepsilon_{\mathbf{h}}$  takes account of the effect of space-group symmetry (see Chapter 2.1).

(b)  $P$  atomic groups having a known configuration but with unknown orientation and position (Main, 1976). Then a certain number of interatomic distances  $r_{j_1j_2}$  are known and

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \left( \sum_N + \sum_{i=1}^P \sum_{j_1 \neq j_2=1}^{M_i} f_{j_1} f_{j_2} \frac{\sin 2\pi q r_{j_1j_2}}{2\pi q r_{j_1j_2}} \right),$$

where  $M_i$  is the number of atoms in the  $i$ th molecular fragment and  $q = |\mathbf{h}|$ .

(c)  $P$  atomic groups with a known configuration, correctly oriented, but with unknown position (Main, 1976). Then a certain group of interatomic vectors  $\mathbf{r}_{j_1j_2}$  is fixed and

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} \left( \sum_N + \sum_{i=1}^P \sum_{j_1 \neq j_2=1}^{M_i} f_{j_1} f_{j_2} \exp 2\pi i \mathbf{h} \cdot \mathbf{r}_{j_1j_2} \right).$$

The above formula has been derived on the assumption that primitive positional random variables are uniformly distributed over the unit cell. Such an assumption may be considered unfavourable (Giacovazzo, 1988) in space groups for which the allowed shifts of origin, consistent with the chosen algebraic form for the symmetry operators  $\mathbf{C}_s$ , are arbitrary displacements along any polar axes. Thanks to the indeterminacy in the choice of origin, the first of the shifts  $\tau_i$  (to be applied to the  $i$ th fragment in order to translate atoms in the correct positions) may be restricted to a region which is smaller than the unit cell (e.g. in  $P2$  we are free to specify

Table 2.2.3.4. (cont.)

H-K group							
$(h, l)I(2, 2)$	$(h+k+l)F(2)$	$(h+k+l)F(4)$	$(l)I(0)$	$(l)I(2)$	$(2k-l)I(4)$	$(l)F(0)$	$I$
$I222$ $I2_12_12_1$	$F432$ $F4_132$	$F222$ $F23$ $F\bar{4}3m$ $F\bar{4}3c$	$I4$ $I4_1$ $I4mm$ $I4cm$ $I4_1md$ $I4_1cd$	$I422$ $I4_122$ $I\bar{4}2m$ $I\bar{4}2d$	$I\bar{4}$ $I\bar{4}m2$ $I\bar{4}c2$	$Fmm2$ $Fdd2$	$I23$ $I2_13$ $I432$ $I4_132$ $I\bar{4}3m$ $I\bar{4}3d$
$(0, 0, 0)$ $(0, 0, \frac{1}{2})$ $(0, \frac{1}{2}, 0)$ $(\frac{1}{2}, 0, 0)$	$(0, 0, 0)$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	$(0, 0, 0)$ $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ $(\frac{3}{4}, \frac{3}{4}, \frac{3}{4})$	$(0, 0, z)$	$(0, 0, 0)$ $(0, 0, \frac{1}{2})$	$(0, 0, 0)$ $(0, 0, \frac{1}{2})$ $(\frac{1}{2}, 0, \frac{3}{4})$ $(\frac{1}{2}, 0, \frac{1}{4})$	$(0, 0, z)$	$(0, 0, 0)$
$(h, l)$	$(h+k+l)$	$(h+k+l)$	$(l)$	$(l)$	$(2k-l)$	$(l)$	$(h, k, l)$
$(2, 2)$	$(2)$	$(4)$	$(0)$	$(2)$	$(4)$	$(0)$	$(1, 1, 1)$
$\varphi_{eee}$	$\varphi_{eee}$	$\varphi_{hkl}$ with $h+k+l \equiv 0$ $(\text{mod } 4)$	$\varphi_{hk0}$	$\varphi_{hke}$	$\varphi_{hkl}$ with $(2k-l) \equiv 0$ $(\text{mod } 4)$	$\varphi_{hk0}$	All
$\ 2\ $	$\ 2\ $	$\ 2\ $ if $h+k+l \equiv 0$ $(\text{mod } 2)$ $\ 4\ $ if $h+k+l \equiv 1$ $(\text{mod } 2)$	$\ \infty\ $	$\ 2\ $	$\ 2\ $ if $h+k+l \equiv 0$ $(\text{mod } 2)$ $\ 4\ $ if $2k-l \equiv 1$ $(\text{mod } 2)$	$\ \infty\ $	All
2	1	1	1	1	1	1	0

## 2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

the origin along the diad axis by restricting  $\tau_1$  to the family of vectors  $\{\tau_1\}$  of type  $[x0z]$ . The practical consequence is that  $\langle |F_{\mathbf{h}}|^2 \rangle$  is significantly modified in polar space groups if  $\mathbf{h}$  satisfies

$$\mathbf{h} \cdot \tau_1 = 0,$$

where  $\tau_1$  belongs to the family of restricted vectors  $\{\tau_1\}$ .

(d) Atomic groups correctly positioned. Then (Main, 1976; Giacovazzo, 1983a)

$$\langle |F_{\mathbf{h}}|^2 \rangle = |F_{p, \mathbf{h}}|^2 + \varepsilon_{\mathbf{h}} \sum_q,$$

where  $F_{p, \mathbf{h}}$  is the structure factor of the partial known structure and  $q$  are the atoms with unknown positions.

(e) A pseudotranslational symmetry is present. Let  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots$  be the pseudotranslation vectors of order  $n_1, n_2, n_3, \dots$ , respectively. Furthermore, let  $p$  be the number of atoms (symmetry equivalents included) whose positions are related by pseudotranslational symmetry and  $q$  the number of atoms (symmetry equivalents included) whose positions are not related by any pseudotranslation. Then (Casarano *et al.*, 1985a,b)

$$\langle |F_{\mathbf{h}}|^2 \rangle = \varepsilon_{\mathbf{h}} (\zeta_{\mathbf{h}} \sum_p + \sum_q),$$

where

$$\zeta_{\mathbf{h}} = \frac{(n_1 n_2 n_3 \dots) \gamma_{\mathbf{h}}}{m}$$

and  $\gamma_{\mathbf{h}}$  is the number of times for which algebraic congruences

$$\mathbf{h} \cdot \mathbf{R}_s \mathbf{u}_i \equiv 0 \pmod{1} \quad \text{for } i = 1, 2, 3, \dots$$

are simultaneously satisfied when  $s$  varies from 1 to  $m$ . If  $\gamma_{\mathbf{h}} = 0$  then  $F_{\mathbf{h}}$  is said to be a *superstructure reflection*, otherwise it is a *substructure reflection*.

Often substructures are not ideal: *e.g.* atoms related by pseudotranslational symmetry are ideally located but of different type (replacive deviations from ideality); or they are equal but not ideally located (displacive deviations); or a combination of the two situations occurs. In these cases a correlation exists between the substructure and the superstructure. It has been shown (Mackay, 1953; Casarano *et al.*, 1988a) that the scattering power of the substructural part may be estimated *via* a statistical analysis of diffraction data for ideal pseudotranslational symmetry or for displacive deviations from it, while it is not estimable in the case of replacive deviations.

### 2.2.4.2. Definition of quasi-normalized structure factor

When probability theory is not used, the *quasi-normalized structure factors*  $\varepsilon_{\mathbf{h}}$  and the *unitary structure factors*  $U_{\mathbf{h}}$  are often used.  $\varepsilon_{\mathbf{h}}$  and  $U_{\mathbf{h}}$  are defined according to

$$|\varepsilon_{\mathbf{h}}|^2 = \varepsilon_{\mathbf{h}} |E_{\mathbf{h}}|^2$$

$$U_{\mathbf{h}} = F_{\mathbf{h}} / \left( \sum_{j=1}^N f_j \right).$$

Since  $\sum_{j=1}^N f_j$  is the largest possible value for  $F_{\mathbf{h}}$ ,  $U_{\mathbf{h}}$  represents the fraction of  $F_{\mathbf{h}}$  with respect to its largest possible value. Therefore

$$0 \leq |U_{\mathbf{h}}| \leq 1.$$

If atoms are equal, then  $U_{\mathbf{h}} = \varepsilon_{\mathbf{h}} / N^{1/2}$ .

### 2.2.4.3. The calculation of normalized structure factors

N.s.f.'s cannot be calculated by applying (2.2.4.1) to observed s.f.'s because: (a) the observed magnitudes  $I_{\mathbf{h}}$  (already corrected for Lp factor, absorption, . . .) are on a relative scale; (b)  $\langle |F_{\mathbf{h}}|^2 \rangle$  cannot

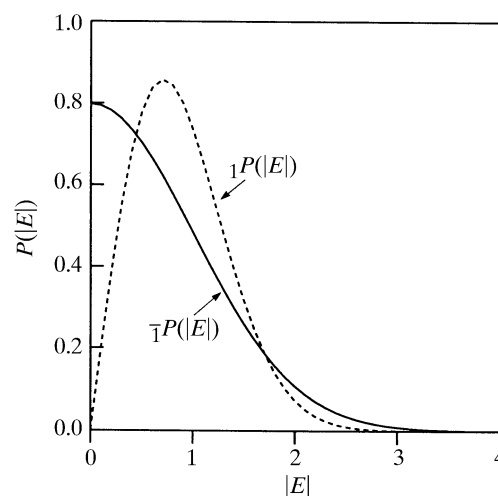


Fig. 2.2.4.1. Probability density functions for cs. and ncs. crystals.

be calculated without having estimated the vibrational motion of the atoms.

This is usually obtained by the well known Wilson plot (Wilson, 1942), according to which observed data are divided into ranges of  $s^2 = \sin^2 \theta / \lambda^2$  and averages of intensity  $\langle I_{\mathbf{h}} \rangle$  are taken in each shell. Reflection multiplicities and other effects of space-group symmetry on intensities must be taken into account when such averages are calculated. The shells are symmetrically overlapped in order to reduce statistical fluctuations and are restricted so that the number of reflections in each shell is reasonably large. For each shell

$$K \langle I \rangle = \langle |F|^2 \rangle = \langle |F^o|^2 \rangle \exp(-2Bs^2) \quad (2.2.4.3)$$

should be obtained, where  $K$  is the scale factor needed to place X-ray intensities on the absolute scale,  $B$  is the overall thermal parameter and  $\langle |F^o|^2 \rangle$  is the expected value of  $|F|^2$  in which it is assumed that all the atoms are at rest.  $\langle |F^o|^2 \rangle$  depends upon the structural information that is available (see Section 2.2.4.1 for some examples).

Equation (2.2.4.3) may be rewritten as

$$\ln \left\{ \frac{\langle I \rangle}{\langle |F^o|^2 \rangle} \right\} = -\ln K - 2Bs^2,$$

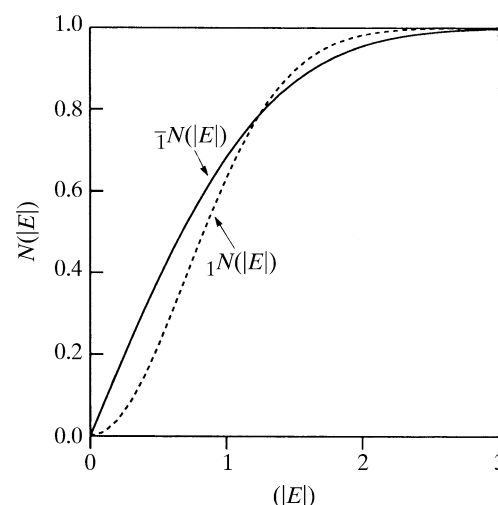


Fig. 2.2.4.2. Cumulative distribution functions for cs. and ncs. crystals.

## 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,  $E_h$  values can be obtained from experimental data by

$$|E_h|^2 = \frac{KI_h}{\langle |F_h^o|^2 \rangle \exp(-2Bs^2)},$$

where  $\langle |F_h^o|^2 \rangle$  is the expected value of  $|F_h^o|^2$  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

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

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

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

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

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