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

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

where $\langle |F^o_{\mathbf{h}}|^2\rangle$ is the expected value of $|F^o_{\mathbf{h}}|^2$ 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)

$${}_{\bar{1}}P(|E|)\ \mathrm{d}|E| = \sqrt{\frac{2}{\pi}}\exp\left(-\frac{E^2}{2}\right)\ \mathrm{d}|E| \qquad (2.2.4.4)$$

and

$${}_{1}P(|E|)\ \mathrm{d}|E| = 2|E|\exp(-|E|^2)\ \mathrm{d}|E|, \qquad (2.2.4.5)$$

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

$${}_{\bar{1}}N(|E|) = \sqrt{\frac{2}{\pi}}\int_0^{|E|}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t = \mathrm{erf}\left(\frac{|E|}{\sqrt{2}}\right),$$

$${}_{1}N(|E|) = \int_0^{|E|} 2t\exp(-t^2)\ \mathrm{d}t = 1-\exp(-|E|^2).$$

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.

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

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 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:

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

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

$$P1: \quad |U_{h,k,l}|^2 \le 1$$
$$P\bar{1}: \quad U^2_{h,k,l} \le 0.5 + 0.5U_{2h,2k,2l}$$
$$P2_1: \quad |U_{h,k,l}|^2 \le 0.5 + 0.5(-1)^k U_{2h,0,2l}.$$

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:

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \le \frac{1}{m}\left\{\sum_{s=1}^{m} a_s(-\mathbf{h})U_{\mathbf{h}(\mathbf{I}-\mathbf{R}_s)} + \sum_{s=1}^{m} a_s(-\mathbf{h}')U_{\mathbf{h}'(\mathbf{I}-\mathbf{R}_s)} \pm 2\mathscr{R}e\left[\sum_{s=1}^{m} a_s(-\mathbf{h}')U_{\mathbf{h}-\mathbf{h}'\mathbf{R}_s}\right]\right\} \qquad (2.2.5.2)$$

where $\mathscr{R}e$ stands for 'real part of'. Equation (2.2.5.2) applied to $P1$ gives

$$|U_{\mathbf{h}} \pm U_{\mathbf{h}'}|^2 \le 2 \pm 2|U_{\mathbf{h}-\mathbf{h}'}|\cos\varphi_{\mathbf{h}-\mathbf{h}'}.$$

references