2.1. Statistical properties of the weighted reciprocal lattice

By U. Shmueli and A. J. C. Wilson

2.1.1. Introduction

The structure factor of the *hkl* reflection is given by

$$F(hkl) = \sum_{j=1}^{N} f_j \exp[2\pi i(hx_j + ky_j + lz_j)], \tag{2.1.1.1}$$

where f_j is the atomic scattering factor [complex if there is appreciable dispersion; see Chapter 1.2 and *IT* C (1999, Section 4.2.6)], $x_j y_j z_j$ are the fractional coordinates of the *j*th atom and N is the number of atoms in the unit cell. The present chapter is concerned with the statistical properties of the structure factor F and the intensity $I = FF^*$, such as their average values, variances, higher moments and their probability density distributions.

Equation (2.1.1.1) expresses F as a function of two conceptually different sets of variables: *hkl* taking on integral values in reciprocal space and *xyz* in general having non-integral values in direct space, although the special positions tabulated for each space group in *IT* A (1983) may include the integers 0 and 1. In special positions, the non-integers often include rational fractions, but in general positions they are in principle irrational. Although *hkl* and *xyz* appear to be symmetrical variables in (2.1.1.1), these limitations on their values mean that one can consider two different sets of statistical properties. In the first we seek, for example, the average intensity of the *hkl* reflection (indices fixed) as the positional parameters of the N atoms are distributed with equal probability over the continuous range 0–1. In the second, we seek, for example, the average intensity of the observable reflections (or of a subgroup of them having about the same value of $\sin\theta$) with the values of *xyz* held constant at the values they have, or are postulated to have, in a crystal structure. Other examples are obtained by substituting the words 'probability density' for 'average intensity'. For brevity, we may call the statistics resulting from the first process *fixed-index* (continuously variable parameters being understood), and those resulting from the second process *fixed-parameter* (integral indices being understood). Theory based on the first process is (comparatively) easy; theory based on the second hardly exists, although there is a good deal of theory concerning the conditions under which the two processes will lead to the same result (Hauptman & Karle, 1953; Giacovazzo, 1977, 1980). Mathematically, of course, the condition is that the phase angle

$$\vartheta = 2\pi(hx + ky + lz) \tag{2.1.1.2}$$

should be distributed with uniform probability over the range 0–2π, whichever set of variables is regarded as fixed, but it is not clear when this distribution can be expected in practice for fixed-parameter averaging. The usual conclusion is that the uniform distribution will be realized if there are enough atoms, if the atomic coordinates do not approximate to rational fractions, if there are enough reflections and if stereochemical effects are negligible (Shmueli *et al.*, 1984).

Obviously, the second process (fixed parameters, varying integral indices) corresponds to the observable reality, and various approximations to it have been attempted, in preference to assuming its equivalence with the first. For example, a third (approximate) method of averaging has been used (Wilson, 1949, 1981): *xyz* are held fixed and *hkl* are treated as continuous variables.

2.1.2. The average intensity of general reflections

2.1.2.1. *Mathematical background*

The process may be illustrated by evaluating, or attempting to evaluate, the average intensity of reflection by the three processes.

The intensity of reflection is given by multiplying equation (2.1.1.1) by its complex conjugate:

$$I = FF^* \tag{2.1.2.1}$$

$$= \sum_{j,k} f_j f_k^* \exp\{2\pi i[h(x_j - x_k) + \ldots]\} \tag{2.1.2.2}$$

$$= \Sigma + \sum_{j \neq k} f_j f_k^* \exp\{2\pi i[h(x_j - x_k) + \ldots]\}, \tag{2.1.2.3}$$

where

$$\Sigma = \sum_j f_j f_j^* \tag{2.1.2.4}$$

is the sum of the squares of the moduli of the atomic scattering factors. Wilson (1942) argued, without detailed calculation, that the average value of the exponential term would be zero and hence that

$$\langle I \rangle = \Sigma. \tag{2.1.2.5}$$

Averaging equation (2.1.2.3) for *hkl* fixed, *xyz* ranging uniformly over the unit cell – the first process described above – gives this result identically, without complication or approximation. Ordinarily the second process cannot be carried out. We can, however, postulate a special case in which it is possible. We take a homoatomic structure and before averaging we correct the f's for temperature effects and the fall-off with $\sin\theta$, so that ff^* is the same for all the atoms and is independent of *hkl*. If the range of *hkl* over which the expression for I has to be averaged is taken as a parallelepiped in reciprocal space with h ranging from $-H$ to $+H$, k from $-K$ to $+K$, l from $-L$ to $+L$, equation (2.1.2.2) can be factorized into the product of the sums of three geometrical progressions. Algebraic manipulation then easily leads to

$$\langle I \rangle = ff^* \sum_j \sum_k \frac{\sin \pi N_H (x_j - x_k)}{N_H \sin \pi (x_j - x_k)}$$
$$\times \frac{\sin \pi N_K (y_j - y_k)}{N_K \sin \pi (y_j - y_k)} \frac{\sin \pi N_L (z_j - z_k)}{N_L \sin \pi (z_j - z_k)}, \tag{2.1.2.6}$$

where $N_H = 2H + 1$, $N_K = 2K + 1$ and $N_L = 2L + 1$. The terms with $j = k$ give Σ, but the remaining terms are not zero. Because of the periodic nature of the trigonometric terms, the effective coordinate differences are never greater than 0.5 and in a structure of any complexity there will be many much less than 0.5. For $HKL = 000$, in fact, $\langle I \rangle$ becomes the square of the modulus of the sum of the atomic scattering factors,

$$\langle I \rangle = \Phi\Phi^*, \tag{2.1.2.7}$$

where

$$\Phi = \sum_{j=1}^{N} f_j, \tag{2.1.2.8}$$

and not the sum of the squares of their moduli; for larger *HKL*, $\langle I \rangle$ rapidly decreases to Σ and then oscillates about that value. Wilson (1949, especially Section 2.1.1) suggested that the regions of averaging should be chosen so that at least one index of every reflection is ≥ 2 if $\langle I \rangle$ is to be identified with Σ, and this has proven to be a useful rule-of-thumb.

The third process of averaging replaces the sum over integral values of the indices by an integration over continuous values, the appropriate values of the limits in this example being $-(H + 1/2)$ to $+(H + 1/2)$. The effect is to replace the sines in the

references