Acta Cryst. (1999). A**55**, 314–321

The probability distribution of structure factors with non-integral indices. II. The $P\bar{1}$ case

Carmelo Giacovazzo,[a]* Dritan Siliqi,[a,b] Benedetta Carrozzini,[a] Antonietta Guagliardi[a] and Anna Grazia Moliterni[a]

[a]*Dipartimento Geomineralogico and IRMEC, c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and* [b]*Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania. E-mail: c.giacovazzo@area.ba.cnr.it*

(*Received 11 March 1998; accepted 23 July 1998*)

Abstract

The probability distribution of the structure factors with non-integral indices is derived in $P\bar{1}$. For integral values of at least one of the indices, the intensity distribution coincides with that provided by Wilson's statistics, but may strongly differ when the indices are (or are close to) half-integers. The deviations are stronger when the integral part of the indices is small, and increase with the size of the structure. In favourable circumstances, moduli and phases of the reflections may be accurately estimated.

1. Symbols and notation

N: number of atoms in the unit cell
f_j: scattering factor of the jth atom (thermal factor included)
h: three-dimensional index with integral components (h, k, l)
p: three-dimensional index with rational components (p_1, p_2, p_3)
$p_s = p_1 + p_2 + p_3$
F: structure factor
E: normalized structure factor
φ: phase of the structure factor

$$\Sigma_1 = \sum_{j=1}^{N} f_j$$

$$\Sigma_2 = \sum_{j=1}^{N} f_j^2$$

Z_j: atomic number of the jth atom.

2. Introduction

In paper I of this series (Giacovazzo & Siliqi, 1998; hereafter paper I), the statistical properties of the structure factors with rational indices were investigated. Three assumptions were made:

(*a*) The structure factors refer to a single unit cell, so that rational indices can be considered.

(*b*) The atomic coordinates were assumed to be the primitive random variables, uniformly distributed in the interval (0, 1), while the reflection indices were kept fixed. When **p** has non-integral components, this condition is not equivalent to the assumption that $2\pi\mathbf{h}\cdot\mathbf{r}_j$ is uniformly distributed over the trigonometric circle.

(*c*) No symmetry elements were considered. Thus the conclusive formulae were strictly valid for $P1$.

The main results of paper I may be summarized as follows:

(i) The distribution $P(|F_\mathbf{p}|)$ shows remarkable differences from Wilson's distribution $P(|F_\mathbf{h}|)$; the differences increase when **p** approaches vectors with half-integral indices, and decrease when one of the indices approaches an integral value and/or when the integral parts of the index components increase.

(ii) In the absence of any prior information about the structure, phase predictions may be made with good reliability when $|p|$ is not large and its components are close to half-integers.

(iii) The distribution $P(|E_\mathbf{p}|)$ and $P(\varphi_\mathbf{p})$ are not universal but depend on the structure complexity.

(iv) Conditional probabilities $P(\varphi_\mathbf{p}||F_\mathbf{p}|)$ and $P(|F_\mathbf{p}||\varphi_\mathbf{p})$ are derived, which are able to exploit prior information on the specific structure.

This paper is devoted to the derivation of the statistical properties of the structure factors with rational indices in the space group $P\bar{1}$ under the assumptions (*a*) and (*b*). The relative distributions will be called centric, by analogy with Wilson's statistics (Wilson, 1942). As in $P1$, the distributions will show unconventional features which are expected to be useful for the solution of the phase problem (see the *Introduction* of paper I for useful references on this subject).

3. About the statistical problem

Let us suppose that the variables x_j, y_j, z_j, $j = 1, \ldots, N/2$, are independently and uniformly distributed in the interval (0, 1). In order to keep all the atoms in the unit cell [*i.e.* x_j, y_j, z_j, $j = 1, \ldots, N$, distributed in the interval $(0, 1)$], $N/2$ symmetry-equivalent atoms are generated by applying the inversion centre at $(1/2, 1/2, 1/2)$. Then,

$$F_{\mathbf{p}} = \sum_{j=1}^{N/2} f_j \big(\exp[2\pi i(p_1 x_j + p_2 y_j + p_3 z_j)] + \exp\{2\pi i[p_1(1-x_j) + p_2(1-y_j) + p_3(1-z_j)]\}\big) = A_{\mathbf{p}} + iB_{\mathbf{p}},$$

where

$$A_{\mathbf{p}} = \sum_{j=1}^{N/2} f_j[\cos(2\pi \mathbf{p}\cdot\mathbf{r}_j) + \cos 2\pi(p_s - \mathbf{p}\cdot\mathbf{r}_j)] = 2\cos\pi p_s \left[\sum_{j=1}^{N/2} f_j \cos(\pi p_s - 2\pi \mathbf{p}\cdot\mathbf{r}_j)\right] = \cos(\pi p_s) A_{0\mathbf{p}}, \quad (1)$$

$$B_{\mathbf{p}} = \sum_{j=1}^{N/2} f_j[\sin(2\pi \mathbf{p}\cdot\mathbf{r}_j) + \sin 2\pi(p_s - \mathbf{p}\cdot\mathbf{r}_j)] = 2\sin\pi p_s \left[\sum_{j=1}^{N/2} f_j \cos(\pi p_s - 2\pi \mathbf{p}\cdot\mathbf{r}_j)\right] = \sin(\pi p_s) A_{0\mathbf{p}} \quad (2)$$

and

$$A_{0\mathbf{p}} = 2\sum_{j=1}^{N/2} f_j \cos(\pi p_s - 2\pi \mathbf{p}\cdot\mathbf{r}_j). \quad (3)$$

The phase of $F_{\mathbf{p}}$ is then given by

$$\varphi_{\mathbf{p}} = \tan^{-1}\left\{[\sin(\pi p_s)A_{0\mathbf{p}}]/[\cos(\pi p_s)A_{0\mathbf{p}}]\right\}. \quad (4)$$

$\varphi_{\mathbf{p}}$ coincides with πp_s if $A_{0\mathbf{p}}$ is positive, and with $\pi(1+p_s)$ if $A_{0\mathbf{p}}$ is negative. Accordingly,

$$F_{\mathbf{p}} = A_{0\mathbf{p}} \exp(i\pi p_s)$$

and, *vice versa*,

$$A_{0\mathbf{p}} = F_{\mathbf{p}} \exp(-i\pi p_s) = |F_{\mathbf{p}}| \exp[i(\varphi_{\mathbf{p}} - \pi p_s)] = |F_{\mathbf{p}}| \cos(\varphi_{\mathbf{p}} - \pi p_s). \quad (5)$$

All the phases are therefore symmetry restricted to two values: special cases occur when $p_s = h$ [then $\varphi_{\mathbf{p}} = (0, \pi)$], or when $p_s = (2h+1)/2$ (then $\varphi_{\mathbf{p}} = \pm\pi/2$).

In accordance with (1) and (2), $B_{\mathbf{p}}$ and $A_{\mathbf{p}}$ are algebraically related *via* the relationship

$$B_{\mathbf{p}} = A_{\mathbf{p}} \tan(\pi p_s).$$

Since p_1, p_2, p_3 are fixed parameters in our statistical approach, the joint probability distribution $P(A_{\mathbf{p}}, B_{\mathbf{p}})$ coincides with

$$P(A_{\mathbf{p}})\delta[B_{\mathbf{p}} - A_{\mathbf{p}} \tan(\pi p_s)],$$

where δ is the Dirac delta function. Thus, $P(A_{\mathbf{p}}, B_{\mathbf{p}})$ is completely known if $P(A_{\mathbf{p}})$ is known. In turn, $A_{\mathbf{p}}$ is algebraically related to $A_{0\mathbf{p}}$ *via* relationship (1): again $P(A_{\mathbf{p}})$ is known if $P(A_{0\mathbf{p}})$ is known. In conclusion, we need only calculate the distribution $P(A_{0\mathbf{p}})$; the other distributions will ensue from this one.

4. The distribution $P(A_{0\mathbf{p}})$

The characteristic function $C(u)$ of the distribution $P(A_{0\mathbf{p}})$, *e.g.*

$$C(u) = \langle \exp(iuA_{0\mathbf{p}}) \rangle,$$

may be written in terms of the cumulants of the distribution. If only terms up to the second order are considered, we have

$$C(u) \simeq \exp(iK_1 u - K_2 u^2/2).$$

Then,

$$P(A_{0\mathbf{p}}) \simeq (2\pi)^{-1} \int_{-\infty}^{+\infty} C(u) \exp(-iuA_{0\mathbf{p}})\, \mathrm{d}u = (2\pi)^{-1} \int_{-\infty}^{+\infty} \exp[-iu(A_{0\mathbf{p}} - K_1) - K_2 u^2/2]\, \mathrm{d}u.$$

The integral may be calculated by applying the standard formula (Gradshteyn & Ryzhik, 1965)

$$\int_{-\infty}^{+\infty} \exp(itu - qu^2/2) = (2\pi/q)^{1/2} \exp(-t^2/2q).$$

We obtain

$$P(A_{0\mathbf{p}}) \simeq (2\pi K_2)^{-1/2} \exp\{-[(A_{0\mathbf{p}} - K_1)^2]/(2K_2)\}. \quad (6)$$

In accordance with Appendix *A*, the following expressions of the cumulants arise,

$$K_1 = m_1 = \langle A_{0\mathbf{p}} \rangle = \Sigma_1 c_{p_1/2} c_{p_2/2} c_{p_3/2}, \quad (7a)$$

$$m_2 = \langle A_{0\mathbf{p}}^2 \rangle = \Sigma_2(1 + c_{p_1} c_{p_2} c_{p_3} - 2c_{p_1/2}^2 c_{p_2/2}^2 c_{p_3/2}^2) + \Sigma_1^2 c_{p_1/2}^2 c_{p_2/2}^2 c_{p_3/2}^2, \quad (7b)$$

$$K_2 = m_2 - m_1^2 = \Sigma_2(1 + c_{p_1} c_{p_2} c_{p_3} - 2c_{p_1/2}^2 c_{p_2/2}^2 c_{p_3/2}^2), \quad (7c)$$

where

$$c_{p_i} = \sin(2\pi p_i)/(2\pi p_i),$$

$$c_{p_i/2} = \sin(\pi p_i)/(\pi p_i).$$

The distribution (6) is the first result of this paper: $A_{0\mathbf{p}}$ is normally distributed about K_1 with variance given by K_2. It may be noticed that $c_{p_i/2} = 0$ for integral values of p_i, and that $c_{p_i/2} = \pm 1/(\pi p_i)$ when p_i is a half-integer. Accordingly, $A_{0\mathbf{p}}$ is always estimated to be zero when at least one of the indices is an integer. The most favourable situation occurs when p_1, p_2, p_3 are all half-integers; then, $c_{p_i} = 0$ for $i = 1, 2, 3$, and

$$\langle A_{0\mathbf{p}}\rangle = K_1 = \pm\Sigma_1(\pi^3 p_1 p_2 p_3)^{-1},$$

$$K_2 = \Sigma_2[1 - 2(\pi^3 p_1 p_2 p_3)^{-2}].$$

The best estimates will be obtained for small values of $|p_1|$, $|p_2|$, $|p_3|$; for these, $|\langle A_{0\mathbf{p}}\rangle|$ is large and K_2 (the variance of the estimate) is relatively small. Usually K_2 is much larger than K_1, and when $|p_i| > 1$ for $i = 1, 2, 3$, it is not too far away from Σ_2. Therefore the estimate of $A_{0\mathbf{p}}$ provided by (5) has a standard deviation smaller than, but close to, Wilson's standard deviation.

5. The distributions $P(A_\mathbf{p})$, $P(B_\mathbf{p})$, $P(F_\mathbf{p})$

Combining (1) with (6) gives

$$P(A_\mathbf{p}) \simeq (2\pi K_2)^{-1/2}(\cos\pi p_s)^{-1} \times \exp[-(A_\mathbf{p} - K_1\cos\pi p_s)^2/(2K_2\cos^2\pi p_s)]. \quad (8)$$

Combining (2) with (6) gives

$$P(B_\mathbf{p}) \simeq (2\pi K_2)^{-1/2}(\sin\pi p_s)^{-1} \times \exp[-(A_\mathbf{p} - K_1\sin\pi p_s)^2/(2K_2\sin^2\pi p_s)]. \quad (9)$$

Since

$$|F_\mathbf{p}|^2 = A_\mathbf{p}^2 + B_\mathbf{p}^2 = A_{0\mathbf{p}}^2,$$

we can calculate the distribution $P(|F_\mathbf{p}|)$ by combining (6) with the relation $|F_\mathbf{p}| = |A_{0\mathbf{p}}|$. We have

$$\begin{aligned} P(|F_\mathbf{p}|) &= P(|A_{0\mathbf{p}}|) + P(-|A_{0\mathbf{p}}|) \\ &= (2\pi K_2)^{-1/2}\exp[(-A_{0\mathbf{p}}^2 - K_1^2)/(2K_2)] \\ &\quad \times [\exp(|A_{0\mathbf{p}}|K_1/K_2) + \exp(-|A_{0\mathbf{p}}|K_1/K_2)], \end{aligned}$$

from which

$$P(|F_\mathbf{p}|) = \exp[-K_1^2/(2K_2)][2/(\pi K_2)]^{1/2} \times \exp[-|F_\mathbf{p}|^2/(2K_2)]\cosh(|F_\mathbf{p}|K_1/K_2). \quad (10)$$

The distribution (10) may be considered the product of an exponential term (*i.e.* the normal or, also, Wilson's component) and of a modulation function (*i.e.* the hyperbolic cosine term). This last component is responsible for the unconventional features of the distribution.

Let us now suppose that the ith index is an integer; then $c_{p_i} = c_{p_i/2} = 0$, $K_1 = 0$, $K_2 = \Sigma_2$. In this case, (10) reduces to Wilson's centric distribution,

$$P(|F|) = [2/(\pi\Sigma_2)]^{1/2}\exp[-|F|^2/(2\Sigma_2)].$$

Vice versa, structure factors for which none of the three indices has an integral value are not distributed according to Wilson's statistics. It may be better to study (10) in terms of normalized structure factors, provided we know the features of the function $\langle|F_\mathbf{p}|^2\rangle$.

6. About the expected value of $|F_\mathbf{p}|^2$

If we introduce into (6) the change of variable

$$A_{0\mathbf{p}n} = (A_{0\mathbf{p}} - K_1)/K_2^{1/2},$$

we have

$$P(A_{0\mathbf{p}n}) = (2\pi)^{-1/2}\exp(-A_{0\mathbf{p}n}^2/2).$$

Denoting $|F_{\mathbf{p}n}| = |A_{0\mathbf{p}n}|$ gives

$$\begin{aligned} P(|F_{\mathbf{p}n}|) &= P(|A_{0\mathbf{p}n}|) + P(-|A_{0\mathbf{p}n}|) \\ &= (2/\pi)^{1/2}\exp(-A_{0\mathbf{p}n}^2/2), \end{aligned}$$

which coincides with the Wilson distribution. The above 'normalization' process leads to a distribution that does not depend on the structural complexity and on the data resolution, but it is of little use in practice; indeed, it requires prior knowledge of the sign of $A_{0\mathbf{p}}$. We prefer, therefore, to investigate the properties of the function $\langle|F_\mathbf{p}|^2\rangle$ and normalize the structure factor in the usual way.

The expected value of $|F_\mathbf{p}|^2$ is given by

$$\langle|F_\mathbf{p}|^2\rangle = \int_0^\infty |F_\mathbf{p}|^2 P(|F_\mathbf{p}|)\,\mathrm{d}|F_\mathbf{p}|, \quad (11)$$

which may be estimated *via* the formula (Gradshteyn & Ryzhik, 1965)

$$\begin{aligned} &\int_0^\infty x^2\exp(-\beta x^2)\cosh(\gamma x)\,\mathrm{d}x \\ &\quad = [\pi^{1/2}(2\beta + \gamma^2)/8\beta^{5/2}]\exp(\gamma^2/4\beta). \end{aligned}$$

We obtain

$$\langle|F_\mathbf{p}|^2\rangle = K_2 + K_1^2 \equiv m_2.$$

According to (7*b*), $\langle|F_\mathbf{p}|^2\rangle$ is a rather complicated oscillating function. To describe its basic properties, we analyse the ratio $\langle|F_\mathbf{p}|^2\rangle/\Sigma_2$ for an equal-atom structure. In this case, $\Sigma_1^2 = N\Sigma_2$ and

$$\langle|F_\mathbf{p}|^2\rangle/\Sigma_2 = 1 + c_{p_1}c_{p_2}c_{p_3} + (N-2)c_{p_1/2}^2c_{p_2/2}^2c_{p_3/2}^2. \quad (12)$$

It is easily verified that:

(*a*) For $p_1 = p_2 = p_3 = 0$, then $c_{p_i} = c_{p_i/2} = 1$ for $i = 1$, 2, 3, and, consequently,

$$\langle|F_\mathbf{p}|^2\rangle = N\Sigma_2 = N^2Z^2,$$

in agreement with the well known relation $F_{000} = NZ$. This result suggests that our statistical approach should also hold when all three indices assume vanishing or almost vanishing values.

(*b*) If p_i is an integer then $c_{p_i} = c_{p_i/2} = 0$, and, in accordance with §5, $\langle|F_\mathbf{p}|^2\rangle = \Sigma_2$, as for Wilson's statistics.

(*c*) If p_i is a half-integer then $c_{p_i} = 0$ and $c_{p_i/2} = \pm 1/(\pi p_i)$. Accordingly, $\langle|F_\mathbf{p}|^2\rangle/\Sigma_2$ will attain a local maximum when all three indices have half-integral

Table 1. *NEWQB: structure-factor moduli and phases calculated from published atomic parameters for a selected set of indices*

In the last two columns the values of πp_s and of $P(\varphi_{\mathbf{p}} = \pi p_s)$ are quoted.

p_2	p_1	p_3	$\|F\|$	φ (°)	πp_s (°)	$P(\varphi_{\mathbf{p}} = \pi p_s)$
0.0	0.0	0.0	791.87	360	360	1.0
0.0	0.0	0.5	494.77	90	90	1.0
0.0	0.5	0.0	504.81	90	90	1.0
0.5	0.0	0.0	495.54	90	90	1.0
0.5	0.0	0.5	297.70	180	180	1.0
0.5	0.5	0.5	341.38	270	270	1.0
0.5	0.5	1.0	2.45	360	360	0.5
0.5	0.5	1.5	162.43	270	90	0.01
0.5	1.0	0.0	116.53	270	270	0.5
0.5	1.0	0.5	157.18	360	360	0.5
0.5	1.0	1.0	68.51	90	90	0.5
0.5	1.0	1.5	20.18	360	180	0.5
0.5	1.5	0.0	49.91	180	360	0.09
0.5	1.5	0.5	11.55	270	90	0.42
0.5	1.5	1.0	75.44	180	180	0.5
0.5	1.5	1.5	108.09	270	270	0.75
1.0	0.0	0.0	0.43	360	180	0.5
1.0	0.0	0.5	71.76	90	270	0.5
1.0	0.0	1.0	120.29	180	360	0.5
1.0	0.0	1.5	124.87	270	90	0.5
1.0	0.5	0.0	72.25	270	270	0.5
1.0	0.5	0.5	64.79	360	360	0.5
1.5	0.0	0.0	149.71	90	270	0.0
1.5	0.0	0.5	166.75	180	360	0.0
1.5	0.0	1.0	78.74	270	90	0.5
1.5	0.0	1.5	44.05	360	180	0.67
1.5	0.5	0.0	140.37	180	360	0.0
1.5	0.5	0.5	134.45	270	90	0.02
1.5	0.5	1.0	21.56	90	270	0.56
1.5	1.0	0.0	117.70	270	90	0.5
1.5	1.5	0.5	32.32	90	270	0.58
1.5	1.5	1.5	28.98	270	90	0.47

values. The practical consequence is that the moduli of the structure factors with all three indices close to half-integers are expected to be larger (on average) than the moduli of the reflections with one or more integral indices. This feature is confirmed by Table 1, where the calculated $|F|$ values of some selected reflections of NEWQB (Grigg *et al.*, 1978), a $P\bar{1}$ structure with chemical formula $C_{24}H_{20}N_2O_5$ (two identical molecules in the asymmetric unit), are shown. The atomic coordinates of this structure are given in the file test of structures distributed by G. Sheldrick, University of Göttingen.

To avoid a rather complicated three-dimensional representation of (12), we draw it in the one-dimensional case, for which

$$\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2 = 1 + c_p + (N-2)c_{p/2}^2. \qquad (13)$$

In Fig. 1, we plot (13) against p for RAND250 and RAND500, two random one-dimensional equal-atom structures (all the atoms are assumed to be carbon, with the same isotropic temperature factor $B_T = 5$) with $N =$ 250 and 500, respectively. We observe:

(*a*) $\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2$ is an oscillating function with maxima at half-integral values of p (but for $p = 1/2$) and minima at integral values;

(*b*) the amplitudes of the oscillations are quite large for small values of p, decay with p, but are still non-negligible up to $p = 9.5$ (low convergence to unity);

(*c*) $\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2$ regularly decreases in the interval $(0, 1)$: thus $p = 0.5$ is the only half-integer index which does not correspond to a relative maximum;

(*d*) the absolute maximum is attained at $p = 0$ (where $\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2 = N$);

(*e*) the amplitudes of the oscillations increase with N.

It may be useful to compare (12) with the corresponding expression derived in paper I for the acentric case [equation (I.20)]. For N sufficiently large,

$$[\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2]_c \simeq [\langle |F_{\mathbf{p}}|^2 \rangle]/\Sigma_2]_a + c_{p_1} c_{p_2} c_{p_3},$$

where c and a stand for 'centric' and 'acentric', respectively. Both the ratios $\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2$ attain their maxima at half-integral values, where $c_{p_i} = 0$; therefore, the oscillation sizes in $P1$ and $P\bar{1}$ are almost identical for structures of equivalent complexity.

The reader can now easily extend the above results to the three-dimensional case, with one warning: as an effect of the three-dimensionality, the intensity oscillations die down rapidly with increasing values of $|p_1|, |p_2|, |p_3|$.

7. The normalized centric distribution $P(|E_{\mathbf{p}}|)$

As for the acentric case, we define

$$E_{\mathbf{p}} = F_{\mathbf{p}} / \langle |F_{\mathbf{p}}|^2 \rangle^{1/2}. \qquad (14)$$

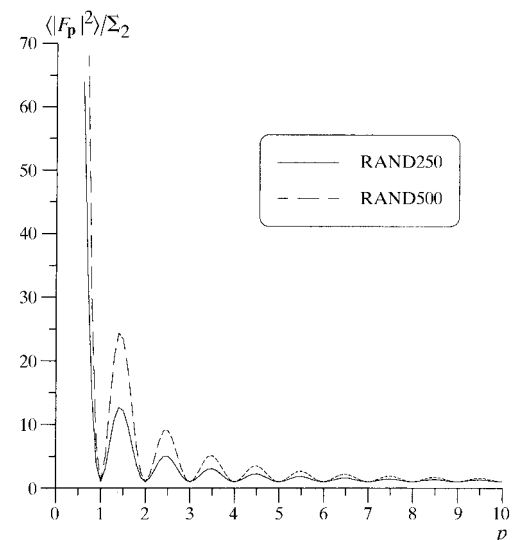

Fig. 1. The ratio $\langle |F_{\mathbf{p}}|^2 \rangle / \Sigma_2$ is plotted against p for the two random structures RAND250 and RAND500.

According to this definition, $E_{000} = 1$, whereas, using Wilson's statistics, $E_{000} = N^{1/2}$.

Combining (10) with (14) gives

$$P(|E_{\mathbf{p}}|) = [2m_2/(\pi K_2)]^{1/2} \exp[-K_1^2/(2K_2)] \times \exp[-|E_{\mathbf{p}}|^2 m_2/(2K_2)] \cosh(|E_{\mathbf{p}}| m_2^{1/2} K_1/K_2). \quad (15)$$

The study of the distribution (15) may be properly made in the one-dimensional case; the extension to three dimensions is trivial. Accordingly, we introduce into (15) the cumulants relative to the one dimensional case,

$$K_1 = m_1 = \Sigma_1 c_{p/2},$$

$$m_2 = \Sigma_2(1 + c_p - 2c_{p/2}^2) + \Sigma_1^2 c_{p/2}^2,$$

$$K_2 = \Sigma_2(1 + c_p - 2c_{p/2}^2).$$

The distribution (15), as obtained for RAND250, is plotted in Fig. 2 for selected values of p. We note:

(a) If p is very close to some integral number and/or is large, then $P(|E_{\mathbf{p}}|)$ will be closer to the centric Wilson distribution (analogous results were found in $P1$). The order of the curves (in terms of deviation from Wilson's distribution) is 1.9, 1.1, 1.8, 1.2, 1.7, 1.3, 1.6, 1.4, 1.5.

(b) As soon as p approaches some half-integer, the mode of the distribution moves towards $|E| = 1$, and the distribution tends to be more symmetric about the mode.

In order to show how the distribution depends on the integral part of p, we draw in Fig. 3, for RAND250, the curves corresponding to selected values of p in the interval (2.0, 2.9). We see that the main features noted in Fig. 2 also hold for Fig. 3, but the deviations from Wilson's distribution are smaller in Fig. 3 than in Fig. 2.

While the $P(|E_{\mathbf{h}}|)$ curves are universal (they do not vary for structures of different complexity), the $P(|E_{\mathbf{p}}|)$ curves are not structure invariants. In Fig. 4 we show the curves calculated for selected values of p in the range (2.0, 3.0) for RAND500. Comparing them with those drawn in Fig. 3 suggests that the deviations of the $P(|E|)$ curves from Wilson's distribution increase with the structural complexity.

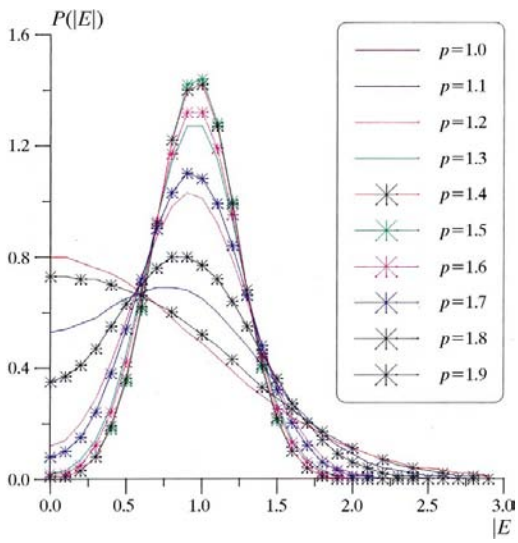

Fig. 2. RAND250: the $P(|E|)$ distribution is plotted for selected values of p between 1 and 2.

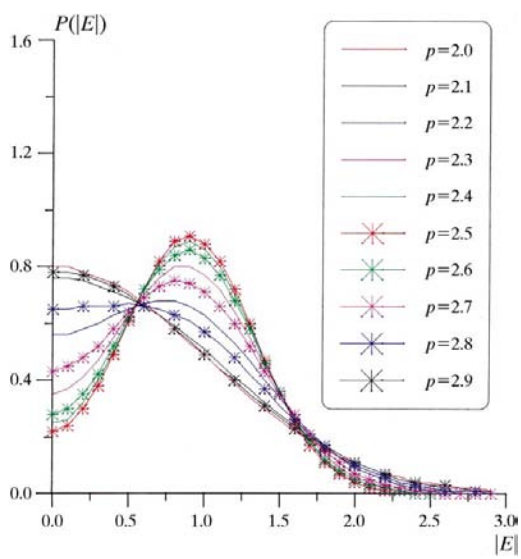

Fig. 3. RAND250: the $P(|E|)$ distribution is plotted for selected values of p between 2 and 3.

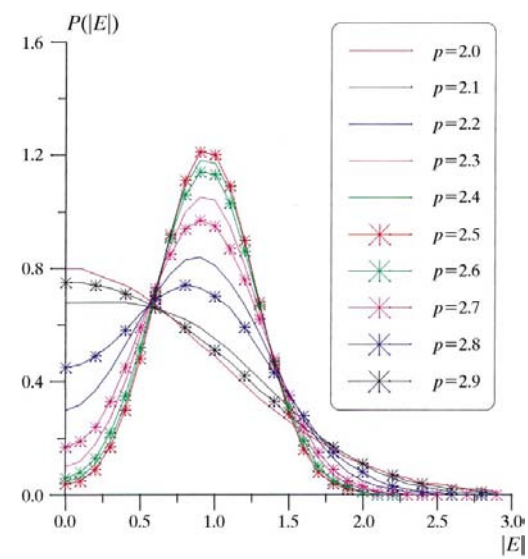

Fig. 4. RAND500: the $P(|E|)$ distribution for selected values of p between 2 and 3.

8. The distribution $P(\varphi_{\mathbf{p}}\|F_{\mathbf{p}}|)$

Let $s_{0\mathbf{p}}$ be the sign of $A_{0\mathbf{p}}$. In accordance with (6), the probability that $A_{0\mathbf{p}}$ is positive or negative is given by

$$P(s_{0\mathbf{p}} = +) = L^{-1}(2\pi K_2)^{-1/2} \exp[-(|A_{0\mathbf{p}}| - K_1)^2/(2K_2)]$$

and

$$P(s_{0\mathbf{p}} = -) = L^{-1}(2\pi K_2)^{-1/2} \exp[-(|A_{0\mathbf{p}}| + K_1)^2/(2K_2)]$$

where L is a suitable scaling factor. Then

$$P(s_{0\mathbf{p}} = -)/P(s_{0\mathbf{p}} = +) = \exp(-2|A_{0\mathbf{p}}|K_1/K_2)$$

and

$$P(s_{0\mathbf{p}} = +) = P(s_{0\mathbf{p}} = +)/[P(s_{0\mathbf{p}} = +) + P(s_{0\mathbf{p}} = -)]$$
$$= 0.5 + 0.5\tanh(|A_{0\mathbf{p}}|K_1/K_2). \quad (16)$$

Since $|A_{0\mathbf{p}}|$ is kept fixed in these calculations, (16) is really the conditional probability of the sign $s_{0\mathbf{p}}$ given $|A_{0\mathbf{p}}|$ [*i.e.* $P(s_0 = \pm 1||A_{0\mathbf{p}}|)$; for shortness we will not emphasize this property in the symbol].

Since $|A_{0\mathbf{p}}| = |F_{\mathbf{p}}|$, we can write (16) in the form

$$P(s_{0\mathbf{p}} = \pm) = 0.5 \pm 0.5\tanh(|F_{\mathbf{p}}|K_1/K_2). \quad (17)$$

Thus the expected sign of $A_{0\mathbf{p}}$ is plus or minus according to whether K_1 is positive or negative; the reliability of the sign prediction increases with increasing values of $|F_{\mathbf{p}}|$.

We now rewrite (4) in the form

$$\varphi_{\mathbf{p}} = \tan^{-1}\{[\sin(\pi p_s)s_{0\mathbf{p}}]/[\cos(\pi p_s)s_{0\mathbf{p}}]\}, \quad (18)$$

which emphasizes the fact that $\varphi_{\mathbf{p}}$ is defined without ambiguity when $s_{0\mathbf{p}}$ is known. Since the sign of $A_{0\mathbf{p}}$ may be probabilistically estimated *via* (17), the following probabilistic formula arises,

$$P(\varphi_{\mathbf{p}} = \pi p_s) = 0.5 + 0.5\tanh(|F_{\mathbf{p}}|K_1/K_2). \quad (19)$$

Values of P smaller than 0.5 indicate that the most probable phase of $\varphi_{\mathbf{p}}$ is $\pi(1 + p_s)$.

In the last two columns of Table 1, we give, for NEWQB, the values of πp_s and $P(\varphi_{\mathbf{p}} = \pi p_s)$. The agreement between predictions and true (*i.e.* calculated from published atomic parameters) phases is excellent; in accordance with theory, no phase prediction is possible for reflections for which one of the indices is an integer different from zero.

9. About the computability of $P(|E_{\mathbf{p}}|)$ and $P(\varphi_{\mathbf{p}}\|F_{\mathbf{p}}|)$ for $0 \le p_i < 1$, $i = 1, 2, 3$

Let us consider the reflections with indices satisfying the relation $0 \le p_i < 1$ for $i = 1, 2, 3$.

As soon as the components (p_1, p_2, p_3) approach $(0, 0, 0)$, very large arguments of the exponential and of the hyperbolic cosine occur in (15); thus, even modern computers can perform the calculations necessary to estimate $P(|E_{\mathbf{p}}|)$ only when $(p_1^2 + p_2^2 + p_3^2)$ is larger than a given threshold. Such behaviour is not unexpected; indeed, $P(|E_{\mathbf{p}}|)$ must coincide, at $p = (0, 0, 0)$, with the Dirac delta function

$$P(|E_{\mathbf{p}}|) = \delta(|E_{\mathbf{p}}| - 1),$$

centred at $E_{000} = 1$ (indeed the value of E_{000} is defined without uncertainty). Luckily, $P(|E_{\mathbf{p}}|)$ is easily predictable where it is not computable. In Fig. 5 we show, for RAND250, the $P(|E|)$ curves calculated for selected indices in the range $(0, 1)$. As soon as p approaches zero, $P(|E_{\mathbf{p}}|)$ will approach a delta function at $|E_{\mathbf{p}}| = 1$. *Vice*

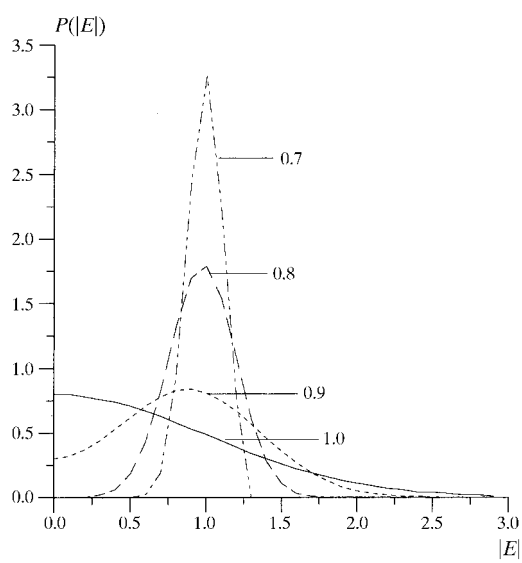

Fig. 5. RAND250: the $P(|E|)$ curves for $p = 1$, 0.9, 0.8, 0.7 are shown.

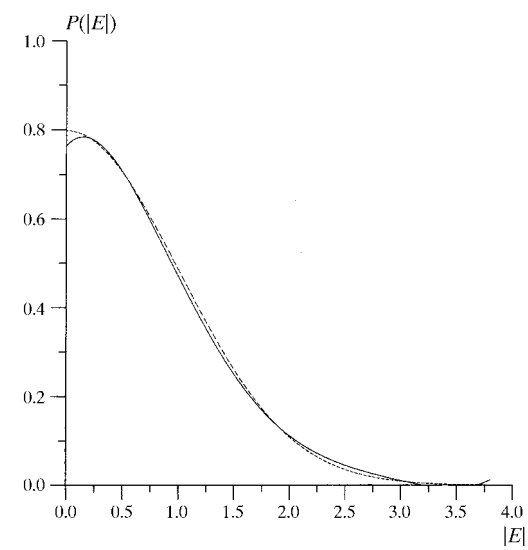

Fig. 6. NEWQB. Experimental distribution $P(|E|)$ for half-integral index reflections compared with Wilson's centric distribution.

versa, it may be assumed that $|E_{\mathbf{p}}|$ is quite close to unity for indices very close to zero.

$P(\varphi_{\mathbf{p}})$ is always computable, even for indices very close to zero. In this domain, very high probability values will be attained as soon as (p_1, p_2, p_3) approaches $(0, 0, 0)$ ($\varphi_{000} = 2\pi$ without uncertainty). Since $s_{0\mathbf{p}}$ is always positive in this domain, $\varphi_{\mathbf{p}}$ is always very close to πp_s.

10. Centric and acentric distributions

It may be observed that, for integral values of p_1, p_2, p_3, the acentric and centric curves have quite different forms (the classical Wilson distributions), while both the centric [*i.e.* equation (15)] and the acentric [*i.e.* equation (I.11)] distributions approach a normal distribution centred on $|E| = 1$ as soon as p_1, p_2, p_3 approach small modulus half-integers. A question arises: can a centric distribution be discriminated from an acentric one for structures of the same complexity *via* the statistical analysis of homologous types of reflections? The problem involves a different type of statistics (that for which the indices are the primitive random variables while the atomic positions are kept fixed) and may be solved on an experimental basis. We have calculated, from the published atomic parameters of NEWQB, the structure-factor moduli corresponding to reflections with half-integral indices (they constitute the subset for which centric and acentric curves are more similar to each other). The corresponding experimental distribution (full line) is shown in Fig. 6; it agrees well with Wilson's centric distribution (broken line), thus suggesting a positive answer to our question. The reason for such behaviour is suggested by referring to Fig. 7, where, for two random three-dimensional structures with equal structural complexity (*i.e.* equal unit cell and equal number of atoms), (I.11) and (15) are shown for selected reflections with half-integral indices. As soon as $|p_1| + |p_2| + |p_3|$ increases, the distributions rapidly converge to the acentric and to Wilson's centric distributions, respectively.

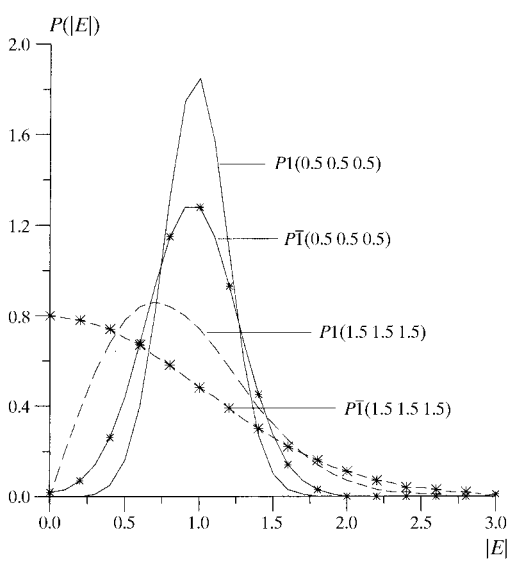

Fig. 7. The $P(|E|)$ distributions (I.11) and (15) for selected half-integral indices. The curves refer to three-dimensional random structures of equal complexity.

11. Conclusions

The distribution of the structure factors with rational indices has been studied in $P\bar{1}$. The main results may be summarized as follows:

(*a*) The distribution $P(|F_{\mathbf{p}}|)$ may be quite different from Wilson's distribution $P(|F_{\mathbf{h}}|)$. The differences increase when $\mathbf{p}$ approaches vectors with half-integral indices, and decrease when the integral part of the index components increase. When one of the components is an integer, $P(|F_{\mathbf{p}}|)$ reduces to Wilson's centric distribution.

(*b*) The phase $\varphi_{\mathbf{p}}$ may be predicted with good reliability when $|\mathbf{p}|$ is small and all three indices are close to a half-integer.

(*c*) $P(|E_{\mathbf{p}}|)$ and $P(\varphi_{\mathbf{p}})$ are not universal, as in Wilson's statistics, but depend on the structural complexity.

The integration of the results obtained in this paper with those derived in paper I shows that Wilson's statistics are part of a larger family, the statistics of the structure factors with rational indices.

APPENDIX *A*
Calculation of the cumulants in $P\bar{1}$

Let us suppose that the variables $x_j, y_j, z_j, j = 1, \ldots, N/2$, are independently and uniformly distributed in the interval (0, 1). Then, according to equation (3),

$$m_1 = \langle A_{0\mathbf{p}} \rangle = \left(2\sum_{j=1}^{N/2} f_j\right)\langle \cos(\pi p_s - 2\pi \mathbf{p}\cdot\mathbf{r}_j)\rangle$$
$$= \Sigma_1[\cos(\pi p_s)c_{\mathbf{p}} + \sin(\pi p_s)s_{\mathbf{p}}], \qquad (20)$$

where

$$c_{\mathbf{p}} = \langle \cos(2\pi\mathbf{p}\cdot\mathbf{r}_j)\rangle$$
$$= c_{p_1}c_{p_2}c_{p_3} - c_{p_1}s_{p_2}s_{p_3} - s_{p_1}s_{p_2}c_{p_3} - s_{p_1}c_{p_2}s_{p_3},$$
$$s_{\mathbf{p}} = \langle \sin(2\pi\mathbf{p}\cdot\mathbf{r}_j)\rangle$$
$$= s_{p_1}c_{p_2}c_{p_3} - s_{p_1}s_{p_2}s_{p_3} + c_{p_1}s_{p_2}c_{p_3} + c_{p_1}c_{p_2}s_{p_3},$$
$$s_{p_i} = [1 - \cos(2\pi p_i)]/(2\pi p_i).$$

It may be shown (for brevity the demonstration is not given) that the following relation holds:

$$\cos(2\pi p_s)c_{2\mathbf{p}} + \sin(2\pi p_s)s_{2\mathbf{p}} = c_{p_1}c_{p_2}c_{p_3}. \qquad (21)$$

Accordingly, (20) reduces to

$$m_1 = \Sigma_1 c_{p_1/2} c_{p_2/2} c_{p_3/2}.$$

The moment m_2 may be calculated as follows,

$$\begin{aligned} m_2 &= 4\left\langle \left[\sum_{j=1}^{N/2} f_j \cos(\pi p_s - 2\pi \mathbf{p}\cdot\mathbf{r}_j) \right]^2 \right\rangle \\ &= 4 \sum_{j=1}^{N/2} f_j^2 [1 + \langle \cos(2\pi p_s - 4\pi \mathbf{p}\cdot\mathbf{r}_j)\rangle]/2 \\ &\quad + 4\left(\sum_{j_1 \neq j_2 = 1}^{N/2} f_{j_1} f_{j_2} \right) [\cos(\pi p_s) c_{\mathbf{p}} + \sin(\pi p_s) s_{\mathbf{p}}]^2. \end{aligned}$$

Since

$$\sum_{j_1 \neq j_2 = 1}^{N/2} f_{j_1} f_{j_2} = \Sigma_1^2/4 - \Sigma_2/2,$$

we have

$$\begin{aligned} m_2 &= \Sigma_2 [1 + \cos(2\pi p_s) c_{2\mathbf{p}} + \sin(2\pi p_s) s_{2\mathbf{p}}] \\ &\quad + 4[\Sigma_1^2/4 - \Sigma_2/2][\cos(\pi p_s) c_{\mathbf{p}} + \sin(\pi p_s) s_{\mathbf{p}}]^2. \end{aligned}$$

Applying (21) gives

$$\begin{aligned} m_2 &= \Sigma_2 (1 + c_{p_1} c_{p_2} c_{p_3} - 2 c_{p_1/2}^2 c_{p_2/2}^2 c_{p_3/2}^2) \\ &\quad + \Sigma_1^2 c_{p_1/2}^2 c_{p_2/2}^2 c_{p_3/2}^2. \end{aligned}$$

References

Giacovazzo, C. & Siliqi, D. (1998). *Acta Cryst.* A**54**, 957–970.

Gradshteyn, I. S. & Ryzhik, I. M. (1965). *Table of Integrals, Series and Products.* New York: Academic Press.

Grigg, R., Kemp, J., Sheldrick, G. M. & Trotter, J. (1978). *J. Chem. Soc. Chem. Commun.* pp. 109–111.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.