Acta Cryst. (1961). **14**, 579

Reliability of Unit Cells Derived from Powder Diffraction Patterns

By P. M. de Wolff

Laboratorium voor Technische Fysica, Technische Hogeschool, Delft, The Netherlands

(*Received* 18 *July* 1960)

It is shown that the reliability of a unit cell, obtained by indexing a powder pattern, depends on the ratio of the actual discrepancies to the average discrepancies Δ which would result from an arbitrary unit cell of the same Bravais type and of about the same size. Expressions for Δ, for all non-cubic Bravais types of lattice, are derived from statistical considerations based on the number of calculated lines below a given Bragg angle.

1. Introduction

The question often arises whether a particular indexing of a given powder pattern is correct. In the present paper we shall attack this question from the other side, namely by trying to judge whether mere chance could be responsible for the apparent success. The ideal answer would be to calculate the probability of that hypothesis. As we shall see, such a calculation, if possible, would be rather complicated. However, a good deal of information on the significance of the proposed unit cell can be obtained by comparing the actual discrepancies with the average discrepancies which could be expected if it were wrong.

To this end, we shall derive expressions for some quantities related to a sequence of calculated line positions.

2. Calculated number of lines

This number, N, is a function of the maximum Bragg angle, which we shall characterize by the corresponding radius r in reciprocal space. Apart from the obvious term in r^3, N may also contain terms of lower degree in r. Only the quadratic additional term will be considered here, since the linear and constant terms are comparable in magnitude with the fluctuations of N around the smooth functions presently to be derived. These fluctuations are surprisingly small provided the reciprocal axes a^*, b^* and c^* have the same order of magnitude. The final results are stated in Table 1, expressed as coefficients in

$$N=r^2(C_0r+C_1a^*+C_2b^*+C_3c^*)/V^*, \qquad (1)$$

where V^* is the volume of the reciprocal unit cell, or, with $Q=r^2$ (cf. Section 3):

$$N=Q(C_0\sqrt{Q}+C_1a^*+C_2b^*+C_3c^*)/V^*. \qquad (1a)$$

Triclinic. N is equal to half the number of reciprocal lattice points within a sphere, hence

$$N=\tfrac{2}{3}\pi r^3/V^*.$$

Monoclinic primitive. N is half the above value, plus half the number of $h0l$ points:

$$N=\tfrac{1}{3}\pi r^3/V^*+\tfrac{1}{2}.\tfrac{1}{2}\pi r^2/a^*c^*\sin\beta$$

from which the final formula follows with $C_2=\frac{1}{4}\pi$.

Orthorhombic and centered lattices. A similar reasoning is followed. Centering of a given lattice causes a reduction of the point density in reciprocal space. The same reduction occurs in the mirror plane(s) of the reciprocal lattice. Hence the appropriate reduction factor applies to all coefficients of N in (1).

Tetragonal. The number of points of a square net with unit period, lying within a circle with radius $\sqrt{p}$, is πp. This leads to the supposition that the number n of integers $<p$ which can be written as a sum of two squares, is $\frac{1}{8}\pi p$. However, many pairs (h, k), with $h\leqq k$, give identical values of h^2+k^2. If (h^2+k^2)-values are supposed to be distributed at random over all integers, the density dn/dp of sums of two squares would be

$$1-\exp(-\tfrac{1}{8}\pi)=0{\cdot}325$$

instead of $\frac{1}{8}\pi=0{\cdot}393$. Moreover, a term in $\sqrt{p}$ can be expected in n. It turns out that the equation

Table 1. *Coefficients in* (1) *for various Bravais types of lattice*

Bravais type	Unique axis	C_0	C_1	C_2	C_3
Triclinic	—	2·095	0	0	0
Monoclinic P	b	1·047	0	0·786	0
Orthorhombic P	—	0·524	0·393	0·393	0·393
Tetragonal P	c	0·214	0·786	0	0·160
Hexagonal	c	0·150	0·681	0	0·113
Rhombohedral (hex. axes)	c	0·050	0·227	0	0·038
C-centered, I-centered	multiply each C by $\frac{1}{2}$				
F-centered	multiply each C by $\frac{1}{4}$				

$$n=0{\cdot}32p+\sqrt{p} \qquad (2)$$

holds with astonishing accuracy up to $p=400$ (Table 2).

Table 2. *Number n or m of integers $<p$ which can be written as h^2+k^2 or as h^2+k^2+hk, respectively*

	n		m	
p	Actual	Eq. (2)	Actual	Eq. (3)
20	11	10·9	9	9·7
40	19	19·1	17	16·7
60	27	27·0	22	23·3
80	34	34·5	30	29·7
100	42	42·0	35	36·0
200	78	78·2	66	66·2
300	113	113·3	92	95·3
400	144	148·0	121	124·0
500	176	182·4	148	152·4
700	239	250·4	199	208·4
900	298	318·0	250	264·0

This good fit is, however, completely fortuitous, because we disregarded the absence of integers of the form $4i-1$ in the series of values of h^2+k^2. Hence the number of coincidences has been underestimated; a precise computation yields $dn/dp=0{\cdot}298$ instead of 0·32 as the asymptotical value. This is corroborated by the behaviour of n for large values of p.

For the calculation of N, we shall use (2) since h^2+k^2 will seldom exceed 400. The tetragonal lattice will be compared with an orthorhombic one with almost the same unit cell. Both reciprocal lattices will be considered layerwise ($l=0, 1, \ldots$). In the orthorhombic lattice, each layer is responsible for

$$n'=\tfrac{1}{4}\pi p+\sqrt{p}$$

diffraction lines, where $\sqrt{p}$ is the radius of the limiting circle in that layer, divided by a^*. For the tetragonal lattice, the number of lines for a given layer is given by (2), p being the same. The term $\sqrt{p}$, in the orthorhombic case, leads to the coefficient C_1+C_2 for $a(\approx b)$ in (1). Hence in the tetragonal case, this coefficient is the same as C_1+C_2 in the orthorhombic expression, whereas C_0 and C_3 are smaller by a factor $0{\cdot}32/\frac{1}{4}\pi=0{\cdot}407$.

Hexagonal: Again a 'semi-empirical' formula obtains for the number m of integers $<p$ which can be written as h^2+k^2+hk:

$$m=0{\cdot}26p+\sqrt{p} \qquad (3)$$

provided p is not too large (Table 2). The asymptotical density is not $0{\cdot}26=1-\exp(-\pi/6\sqrt{3})$ but smaller (0·238) owing to the fact that these integers never have the form $3i-1$. Again, this is confirmed for large p. The final N-formula follows from comparison with a pseudotetragonal lattice with the same a^* and c^* and a reciprocal cell volume $2/\sqrt{3}$ times the actual V^*. Hence all coefficients are smaller than in the orthorhombic case by $\frac{1}{2}\sqrt{3}$, and C_0 and C_3 by a further factor $0{\cdot}26/\frac{1}{4}\pi=0{\cdot}331$.

The data in Table 2 were derived from the *International Tables for X-ray Crystallography* (1959).

3. Distribution of intervals between calculated lines

The term 'distribution' is used here in the sense of a frequency distribution. It refers, not to the collection of all intervals of a sequence of calculated lines for a given unit cell, but to the intervals situated at, or close to, a given Bragg angle. The collection would thus be very small if only one unit cell were considered. However, by varying the cell constants somewhat, a sufficiently large collection can be defined. The individual fluctuations of intervals for any given unit cell, corresponding to the fluctuations of the actual N with respect to its average given by (1), will be sufficiently random in this collection to permit statistical considerations. These will be used to predict the interval length to be expected at a given angle for a given unit cell, without actually calculating all possible line positions near this angle.

As usual, we shall express line positions by Q, which may be the recently recommended $Q=10^4/d^2$ (*Tables of Q*, 1959) or any other quantity proportional to $\sin^2\theta$. The reciprocal quantities used before must be expressed in such units that $Q=r^2$, $Q_{100}=a^{*2}$ etc.

Table 3. *Cumulative distribution of* 214 *intervals* (*cf. Section* 3)

	Number of intervals $>x-\frac{1}{2}$	
x	Actual	$214\exp\{-(x-\frac{1}{2})/18{\cdot}8\}$
5	180	170
10	141	129
15	107	99
20	77	76
25	59	58
30	42	44
35	29	34
40	22	26
45	19	20
50	13	15
60	9	9
75	2	1
100	1	1

The main parameter of the interval distribution is, of course, the average interval between successive Q-values, in the sense indicated above. It follows that this interval, which we shall call 2Δ, is

$$2\Delta=dQ/dN=2r/(dN/dr)$$
$$=V^*/(\tfrac{3}{2}C_0\sqrt{Q}+C_1a^*+C_2b^*+C_3c^*)\,. \qquad (4)$$

If the intervals were distributed exponentially as free path lengths of gas molecules, this would be the only parameter. There is indeed little reason to believe that a different distribution obtains. By way of example, we choose a two-dimensional anorthic net. For this case $2\Delta=(a^*b^*\sin\gamma)/\frac{1}{2}\pi$ is independent of Q, so that the collection of 214 successive intervals for

this special lattice ($Q=23h^2+38k^2+4hk$) can be regarded as representative of a collection of the kind defined above. The cumulative distribution is shown in Table 3. In the last column, the theoretical frequency $\exp -(x/2\Delta)$ for intervals larger than x is listed, where for this case $2\Delta=(23\times 38-4)^{\frac{1}{2}}/\frac{1}{2}\pi=18{\cdot}8$. There appears to be a slight deviation towards a narrower distribution, indicating that the Q's are more 'regularly' spaced than points chosen at random (cf. next section).

4. Average discrepancy

We are now in a position to examine the discrepancy to be expected between a given observed line and the calculated line which is closest to it. Since we want to ascertain what the chances are for a particular unit cell to be only apparently the right one, we must for the moment assume that it is wrong. The position of the observed line can then be supposed to be unrelated to the sequence of calculated lines and the average discrepancy is unrelated to the accuracy either of the line position or of the cell constants, but only to the interval distribution of calculated lines at the position of the observed line.

Under these conditions, and assuming that distribution to be exponential, the average discrepancy $\bar{\varepsilon}$ is half the average interval and thus equal to Δ:

$$\bar{\varepsilon}=\Delta=\tfrac{1}{2}V^*/(\tfrac{3}{2}C_0/Q+C_1a^*+C_2b^*+C_3c^*) \qquad (5)$$

because the discrepancy is evenly distributed between zero and half the length of the interval in which the observed line is situated, whereas the conditional average length of that interval is 4Δ.†

Moreover, the distribution of discrepancies for a certain observed line is easily demonstrated to be exponential as well, so discrepancies larger than x will occur with the frequency $\exp[-x/\Delta]$.

A non-exponential distribution of intervals can be expected in particular for unit cells of high symmetry, where successive calculated Q's may be somewhat more regularly spaced. This will not influence the average interval 2Δ, but it will make $\bar{\varepsilon}<\Delta$; e.g. for the extreme case of equidistant Q's, $\bar{\varepsilon}=\frac{1}{2}\Delta$. Therefore, by supposing an exponential distribution, the reliability of the assumed unit cell can only be overestimated. In the following sections we shall identify $\bar{\varepsilon}$ with Δ.

† Statistically, the observed line cannot be distinguished from the surrounding calculated lines. Hence each of the two parts in which it divides the interval in question is itself equivalent to an arbitrary interval, averaging 2Δ. Indeed, the whole interval is not arbitrary because large intervals are more likely to contain observed lines than small ones.

This 'Weglängenparadox' was solved originally for a sequence of points distributed at random on a line. The sequence of Q-values is not random; there is a very pronounced negative correlation between neighbouring interval lengths. However, the above conclusions are independent from such a correlation.

5. Example; discussion

By way of illustration, a recently published indexing of the pattern of an aluminium orthoarsenate (Sharan, 1959) will be analyzed. The quadratic form proposed by this author is

$$10^4\sin^2\theta=Q=50h^2+74k^2+88l^2.$$

Applying the expression for N for an orthorhombic lattice, we find

$$N=Q(0{\cdot}00092\sqrt{Q}+0{\cdot}0173);$$

e.g. $N=117$ for $Q=2000$ and $N=638$ for $Q=6850$, which is the largest observed value. (The actual number of values of $Q<2000$ is 115; the difference with the N just mentioned is typical of the accuracy of equation (1).

The number of observed lines below 2000 and 6850, respectively, is 12 and 28, which is 10% and 4½% of the corresponding N. These low percentages might still be compatible with a good reliability if the actual discrepancies were low. The latter are listed in Table 4, together with some values of the average expected discrepancy Δ given in this case by

$$\Delta=1/(0{\cdot}00138\sqrt{Q}+0{\cdot}0173)\,.$$

A comparison between the two shows that for most observed lines the agreement is not at all better than could be expected for an arbitrary orthorhombic unit cell of similar size.‡

Table 4. *Observed values of Q and discrepancies for an indexed pattern (Sharan, 1959)*

Q_o	Q_o-Q_c	Δ	Q_o	Q_o-Q_c	Δ
197	−3	13·5	2490	−6	
287	−1		2811	9(−7)‡	
494	−2		3057	−9	5·4
718	2		3572	−12	
790	−2		3765	+1	
882	−6		4527	+3	
999	7	8·2	4869	−7	
1107	11		5540	−6	
1477	11(±5)‡		5841	9(−1)‡	
1565	10		6037	11(−5)‡	
1641	7		6309	11	4·0
1821	3		6597	3	
2023	−1	6·3	6857	−9	
2263	−11(−5)‡				

It should perhaps be stressed that this fact does not prove that the proposed unit cell is wrong. However, Sharan's statement that 'the validity of the lattice constants is apparent from the good agreement between $\sin^2\theta_o$ and $\sin^2\theta_c$' proves that the implications

‡ For the higher angles, the agreement is indeed so much worse than the expectation, that we were led to examine the indices in detail. It turned out that 5 large discrepancies can be substantially reduced, as exemplified by the line $Q=1477$ for which no less than 3 triplets (014, 241, 332) yield smaller discrepancies than the one given by Sharan.

of the term 'good agreement' have not been realized by this author nor, for that matter, by the authors of quite a few other papers containing similar indexings and similar statements.

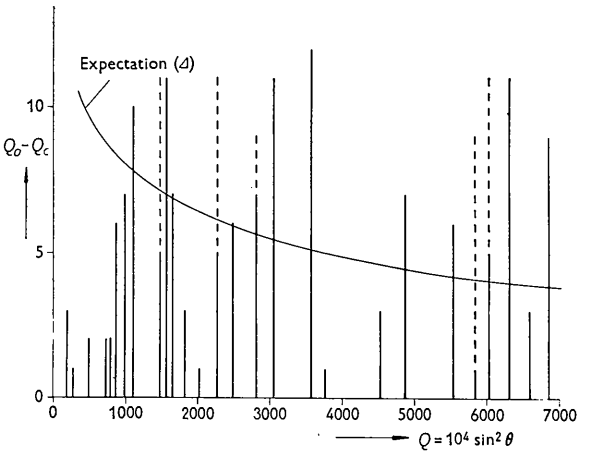

Fig. 1. Actual discrepancies for an indexed pattern and expected discrepancy Δ for a false unit cell (cf. Table 4).

The underlying misconception is probably the idea that the agreement is good when the discrepancies do not exceed the experimental errors. It is true that such an agreement is sufficient when these errors are small compared with Δ for a large number of observed lines. If, for instance, the error limit is equal to $0{\cdot}1\Delta$, (Δ being defined for a given Bravais type and a certain size of unit cell since it depends chiefly on V*), an arbitrary unit cell of this type and size has a chance of $1-\exp(-0{\cdot}1) \approx 0{\cdot}1$ of yielding a discrepancy smaller than the expected error limit, for one observed line. The chance of obtaining the same quality of fit for, say, 20 lines is, therefore, 10^{-20}. Actually, each line has a different ratio of error limit: Δ, because the two tend to vary in opposite directions with θ. Anyhow, it will be clear that the presence of some 20 consecutive lines each having a small value for that ratio practically excludes the existence of a false unit cell of the assumed size.

In Sharan's case, we found in the same way that the combined probability for all discrepancies not to exceed the error limit is about 0·01. (The first 3 lines were excluded because there are 3 parameters; $Q=790$ was excluded as a probable second order of the first line. The error limit was set as closely as the actual discrepancies permit, namely at 12 beyond, and at 8 below, $Q=1000$.) Now there are hundreds of choices possible for the indices of the three parameter-determining lines, even with the restriction of a limit for the cell volume.† Hence the existence of a false unit cell yielding an agreement at least similar to Sharan's result is fairly certain. If, furthermore, the experimental errors are really as large as the discrepancies allowed by Sharan, neither the indexing method he used nor any other method will be able to distinguish between false and true unit cells.

As a conclusion, we may say that it is not sufficient to obtain a good agreement with respect to errors. A reliable indexing should have a substantial number of discrepancies much smaller than the corresponding values of Δ. If this should leave some weak lines unexplained, the chance of a second phase responsible for these lines is usually far greater than the chance of existence of a false unit cell fulfilling the criterion just mentioned.

The author wishes to thank Miss M. van Roon and Miss E. Brouns for their assistance in calculating the data of Table 3.

References

International Tables for X-ray Crystallography. (1959), Vol. II, pp. 110, 113. Birmingham: The Kynoch Press.
SHARAN, B. (1959). *Acta Cryst.* **12**, 948.
Tables of Q, Announcement (1959). *Acta Cryst.* **12**, 421.

† A precise evaluation of the probability of the 'false unit cell hypothesis', as alluded to in the introduction, would involve a separate calculation of Δ for each choice. Many choices, in the present case, would give unit cells smaller than Sharan's; for these, the combined chance of success would be much smaller than the above figure 0·01. However, even in a range of cell volumes within $\pm 20\%$ from that of Sharan, the number of choices is already of the order of 100.