Journal of
Applied Crystallography
ISSN 0021-8898

Received 2 July 2004
Accepted 24 September 2004

© 2004 International Union of Crystallography
Printed in Great Britain – all rights reserved

Space-group determination from powder diffraction data: a probabilistic approach

Angela Altomare,[a] Rocco Caliandro,[a] Mercedes Camalli,[b] Corrado Cuocci,[c] Iván da Silva,[a,d] Carmelo Giacovazzo,[a,c]* Anna Grazia Giuseppina Moliterni[a] and Riccardo Spagna[b]

[a]IC, Bari, Via Amendola 122/o, 70126 Bari, Italy, [b]IC, Sezione di Roma, Via Salaria Km. 29.300 Montelibretti Stazione, Roma, Italy, [c]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [d]Departamento Física Fundamental II, Universidad de La Laguna, Avda. Astrofísico Fco. Sánchez s/n, E-38204, La Laguna, Tenerife, Spain. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

Experimental powder diffraction diagrams, once indexed and decomposed into single diffraction intensities, can be submitted to statistical analysis for the determination of space-group symmetry. A new algorithm is illustrated, which is able to provide, on a quantitative basis, a probability value for each extinction symbol compatible with the previously established lattice symmetry. The algorithm has been implemented in *EXPO2004* [Altomare, Caliandro, Camalli, Cuocci, Giacovazzo, Moliterni & Rizzi (2004). *J. Appl. Cryst.* **37**, 1025–1028] and has been successfully tested using a large set of experimental data.

1. Notation

$F_{\mathbf{h}}$: structure factor with vectorial index $\mathbf{h} = (h, k, l)$.
$E_{\mathbf{h}}$: normalized structure factor.
$z = |E|^2$.
ITCr = International Tables for Crystallography, Vol. A (1992).
$\sum_N = \sum_N f_j^2$, where N is the number of atoms in the unit cell and f is the atomic scattering factor.

2. Introduction

In the structure solution pathway from powder diffraction data, a critical step follows the pattern indexation; the determination of the space group is fundamental for the solution process. This task is usually carried out by a careful investigation of the experimental intensities, aiming at characterizing the extinction group. Because of the collapse of the experimental data onto the Bragg angle one-dimensional space, the experimental pattern is affected by peak overlap and uncertain background definition. As a consequence, judgements about the 'presence' or 'absence' of reflections may provide ambiguous results.

An innovative approach avoiding manual inspection of the diffraction pattern was developed by Markvardsen *et al.* (2001), who elaborated an algorithm for estimating the relative probabilities of the different extinction symbols. The procedure, based on a Bayesian probabilistic approach, uses the intensities (and their correlation matrix) extracted *via* the Pawley (1981) method to check the probability of each extinction group compatible with the crystal system under investigation. Markvardsen *et al.* (2001) showed, for a small number of test structures, that the most probable extinction symbol usually corresponds to the true one, but the discrimination in terms of probabilities between the correct and incorrect groups is not always significant.

A new probabilistic procedure aimed at recognizing the extinction symbol will be described here. This approach uses the statistics of the normalized z intensities [as extracted by the Le Bail method (Le Bail *et al.* 1988)] to calculate the probability of each extinction symbol compatible with the crystal system suggested by the *N-TREOR* (Altomare *et al.*, 2000) indexing procedure. A suitable weight is associated with each z value, depending on the expected reliability of its estimate. The algorithm, which has been implemented in *EXPO2004* (Altomare *et al.*, 2004), provides a probability for each extinction symbol of the considered crystal system. The algorithm has been tested on a large number of test structures and on all the crystal systems, providing good results. Once the extinction symbol has been determined, the choice of the space group (the list of the space groups compatible with the extinction symbol is promptly available by means of the graphic interface in *EXPO2004*) is not problematic, firstly because the number of space groups that are compatible with the extinction symbol is not usually large, and secondly because additional chemical information can be taken into account to help with the selection.

3. The algorithm

The steps of the algorithm are as follows.

Step 1. The unit cell is determined by indexing the experimental powder pattern; the crystal system is therefore assumed to be known.

Step 2. For each crystal system, the relative extinction symbols are stored (see Table 3.2 of *ITCr*). They are 14 for the monoclinic system, unique axis *b*; 111 for the orthorhombic; 31 for the tetragonal; 12 for the trigonal–hexagonal systems (only hexagonal axes considered); 18 for the cubic.

Step 3. The full pattern decomposition is performed, according to the Le Bail algorithm, in the space group having the largest Laue symmetry and no extinction conditions (*e.g.* *P*12/*m*1 in the monoclinic system, *P*2/*m*2/*m*2/*m* for the orthorhombic, *P*4/*mmm* for the tetragonal, *P*6/*mmm* for the trigonal–hexagonal systems, and $Pm\bar{3}m$ for the cubic).

Step 4. The extracted intensities are then normalized, according to the classical Wilson method, using as space groups the Laue groups mentioned in step 3. This is an arbitrary choice, because the statistical Wilson parameter ϵ varies with the space group, and this is unknown at this stage. The choice, however, proved not to be critical for the success of the algorithm.

Step 5. The *z* statistics are used to define the probability of the extinction symbol.

To give a representative example, let us consider the orthorhombic system, where any space group may be represented by the general string

$$M\, r_1/s_1\, r_2/s_2\, r_3/s_3. \tag{1}$$

M denotes the unit-cell type; r_j, j = 1–3, are the symmetry elements along the three axes; s_j, j = 1–3, are the symmetry elements perpendicular to the axes. We assume that the occurrence probabilities for the 2_1 screw axes are

$$p(2_{1[100]}) = 1 - \langle z_{h00}\rangle_{h=2n+1}, \tag{2a}$$

$$p(2_{1[010]}) = 1 - \langle z_{0k0}\rangle_{k=2n+1}, \tag{2b}$$

$$p(2_{1[001]}) = 1 - \langle z_{00l}\rangle_{l=2n+1}. \tag{2c}$$

The probabilities are equal to unity if the *z* averages are close to zero, and vanish if the averages are equal to (or larger than) unity. The same equations define the probabilities for the twofold axes:

$$p(2_{[100]}) = 1 - p(2_{1[100]}),$$

$$p(2_{[010]}) = 1 - p(2_{1[010]}),$$

$$p(2_{[001]}) = 1 - p(2_{1[001]}).$$

Let us now consider s_1 (the same criterion will hold for s_2 and s_3). It may coincide with *b*, *c*, *n* and *d*, for which the following probabilities arise:

$$p(b) = 1 - \langle z_{0kl}\rangle_{k=2n+1},$$

$$p(c) = 1 - \langle z_{0kl}\rangle_{l=2n+1},$$

$$p(n) = 1 - \langle z_{0kl}\rangle_{k+l=2n+1},$$

$$p(d) = 1 - \langle z_{0kl}\rangle_{k+l\neq 4n}.$$

The probability for the mirror plane is assumed to be

$$p(m) = 1 - \max[p(b), p(c), p(n), p(d)].$$

The probabilities for the allowed unit-cell types may be derived as follows. First we calculate

$$p'(A) = 1 - \langle z_{hkl}\rangle_{k+l=2n+1},$$

$$p'(B) = 1 - \langle z_{hkl}\rangle_{h+l=2n+1},$$

$$p'(C) = 1 - \langle z_{hkl}\rangle_{h+k=2n+1},$$

$$p'(I) = 1 - \langle z_{hkl}\rangle_{h+k+l=2n+1},$$

from which we obtain the probabilities

$$p(A) = p'(A)[1 - p'(B)][1 - p'(C)], \tag{3a}$$

$$p(B) = p'(B)[1 - p'(A)][1 - p'(C)], \tag{3b}$$

$$p(C) = p'(C)[1 - p'(A)][1 - p'(B)], \tag{3c}$$

$$p(I) = p'(I). \tag{3d}$$

Equations (3*a*)–(3*c*) express the probabilities that the cell is *A*, *B* or *C* without being *F*. The probabilities for the *F* and the *P* cells are, respectively,

$$p(F) = 1 - \langle z_{hkl}\rangle_{[g]},$$

where [*g*] is the subset of reflections for which not all the indices are odd or even, and

$$p(P) = 1 - \max[p(A), p(B), p(C), p(I), p(F)].$$

Each extinction symbol of the orthorhombic system is compared with the general string given in (1); the probability is calculated as the product of the probabilities of the symmetry elements involved in the extinction symbol. For example, for the symbol *P*−−− we can write the following expression:

$$p(P{-}{-}{-}) = p(P)\, p(2_{[100]})\, p(m \perp \mathbf{a})\, p(2_{[010]})\, p(m \perp \mathbf{b}) \times p(2_{[001]})\, p(m \perp \mathbf{c}).$$

Let us give the probability expression for an extinction symbol with a centred cell, *e.g.* *Bb*−*b*. The string *b*−*b* implies the following symmetry elements:

$$b \perp \mathbf{a},\ b \perp \mathbf{c},\ 2_{1[010]}.$$

The presence of *B* additionally implies

$$c \perp \mathbf{a},\ n \perp \mathbf{b},\ a \perp \mathbf{c},\ 2_{1[100]},\ 2_{1[001]}.$$

Accordingly (see Table 3.2 of *ITCr*), the probability for the extinction symbol *Bb*−*b* has to be calculated as

$$p(Bb{-}b) = p(B)\ p(2_{1[100]})\, p(b, c \perp \mathbf{a})\, p(2_{1[010]})\, p(n \perp \mathbf{b}) \times p(2_{1[001]})\, p(a, b \perp \mathbf{c}).$$

The scheme is easily extended to other crystal systems; we will examine some distinctive examples to clarify possible obscurities.

(*a*) The occurrence probability of the unit-cell type may be fixed as follows.

(i) Monoclinic system: the 14 extinction symbols include *A*, *C*, *I* and *P* cells; since the extinction conditions are mutually exclusive, we have $p(A) = p'(A)$, $p(C) = p'(C)$, $p(I) = p'(I)$, $p(P) = 1 - \max[p(A), p(C), p(I)]$.

(ii) Tetragonal system: $p(P) = 1 - p(I)$.

(iii) Trigonal–hexagonal systems: $p(P) = 1 - p(R)$.

(iv) Cubic system. $p(I) = p'(I)$, $p(F) = p'(F)$, $p(P) = 1 - \max[p(I), p(F)]$.

(*b*) In the trigonal–hexagonal systems, *P*−−− may represent both trigonal (*e.g. P*3, $P\bar{3}$, *P*321 *etc.*) and hexagonal (*P*6, $P\bar{6}$, *P*6*mm*, $P\bar{6}2m$ *etc.*) space groups. Then:

(i) $p(3 \text{ or } 6) = 1 - \max[p(3_1), p(6_1), p(6_3)]$.

(ii) $p(m \perp \mathbf{c}) = 1$ (no glide plane perpendicular to **c** in these systems).

(iii) $p(2_{[100]})$ or $p(2_{[1\bar{1}0]}) = 1$ (no extinction condition for screws in the plane normal to **c**).

(iv) $p(m \perp \mathbf{a}) = 1 - p(c \perp \mathbf{a})$ (only *c* glide planes can occur).

In conclusion, *P*−−− is equivalent to the string *P*(3 or 6)/*m*2/*m*2/*m* and therefore $p(P---) = p(P)\ p(3 \text{ or } 6)\ p(m \perp \mathbf{a})\ p[m \perp (\mathbf{a} - \mathbf{b})]$.

Step 6. For each crystal system, the probabilities of the various extinction symbols are calculated and presented to the user in order of probability. Clicking on an extinction symbol in the list opens the sublist of the compatible space groups, from which the user can choose the preferred one.

4. Taking care of peak overlapping and of the scarcity of experimental information

The above simple scheme, when applied to experimental powder patterns, proved to be not sufficiently robust (details of the results are not shown for brevity). The procedure was then modified to take into account peak overlapping and the eventual scarcity of reflections obeying the extinction conditions.

4.1. About peak overlapping

As stated in §2, the collapse of the three-dimensional lattice onto the 2θ axis hinders the unambiguous estimation of the diffracted intensities; it may frequently occur that for most of the reflections the assigned z value is uncertain, and that uncertainty can critically weaken any criterion based on the standard z statistics. It seems more useful to associate each reflection with a proper weight, depending on the accuracy of the intensity estimate.

Let us suppose that two symmetry-independent reflections perfectly overlap. For non-centrosymmetric space groups, the joint probability distribution $P(z_1, z_2)$ is given by

$$P(z_1, z_2) = \exp(-z_1 - z_2).$$

The probability that the sum of z_1 and z_2 is equal to S is given by

$$\begin{aligned} P(S) &= \int_0^S P(z_1, S - z_1)\,\mathrm{d}z_1 \\ &= \int_0^S \exp(-z_1)\exp(-S + z_1)\,\mathrm{d}z_1 \\ &= \int_0^S \exp(-S)\,\mathrm{d}z_1 = S\exp(-S). \end{aligned} \tag{4}$$

For a cluster of n perfectly overlapping reflections, with normalized diffraction moduli $z_1, z_2, \ldots, z_n$, the probability $P(S)$ is given by

$$\begin{aligned} P(S) &= \int_0^S \cdots \int_0^S \exp[-z_1 - z_2 - \cdots \\ &\quad - (S - z_1 - z_2 - \cdots - z_{n-1})]\,\mathrm{d}z_1 \ldots \mathrm{d}z_{n-1} \\ &= S^{n-1}\exp(-S), \end{aligned} \tag{5}$$

where $S = z_1 + z_2 + \cdots + z_n$. Equation (5) suggests that the overall intensity of a cluster of n overlapping reflections is distributed differently from the single reflection intensity (see Fig. 1). In accordance with the central limit theorem, for increasing values of n the distribution $P(S)$ becomes more similar to a Gaussian, with mean value $\langle S\rangle$ equal to the sum of the mean values ($\langle S\rangle = n\ \langle z\rangle = n$) and variance equal to the sum of the z variances. This trend is clearly visible in Fig. 1.

Let us now calculate, when S is the measured intensity of a cluster of n perfectly overlapping reflections, the variance σ_z^2 associated with the generic intensity component z (note that the Le Bail algorithm, in the absence of additional prior information, would assign the same intensity S/n to all the reflections of the cluster). We have

$$\sigma_z^2 = \int_0^S z^2\exp(-z)\,\mathrm{d}z - \left[\int_0^S z\exp(-z)\,\mathrm{d}z\right]^2.$$

Since

$$\int_0^S z^n\exp(-z)\,\mathrm{d}z = n! - \exp(-S)\sum_{k=0}^{n}\frac{n!}{k!}S^k$$

we have

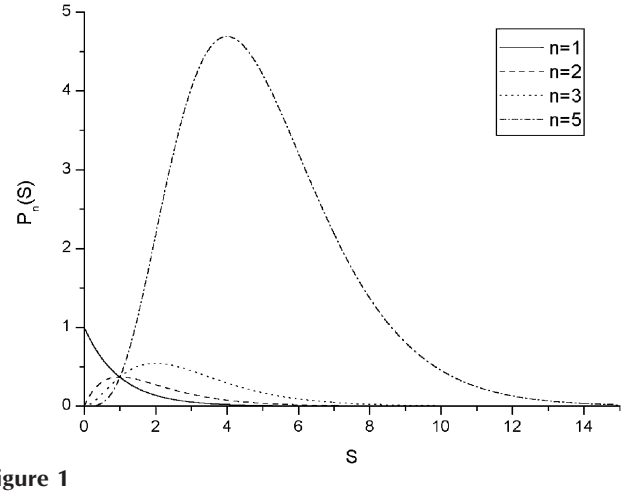


Figure 1
Equation (5) plotted for different values of *n*.

$$\int_0^S z^2 \exp(-z)\,\mathrm{d}z = 2 - \exp(-S)(2 + 2S + S^2),$$

$$\int_0^S z \exp(-z)\,\mathrm{d}z = 1 - \exp(-S)(1 + S),$$

from which we obtain

$$\sigma_z^2 = 1 - S^2 \exp(-S) - (1 + S)^2 \exp(-2S). \qquad (6)$$

It is easily seen that:

(*a*) $\sigma_z^2 = 0$ when $S = 0$. In this case the error on the intensity estimate due to overlapping is vanishing, as should be expected.

(*b*) If S tends to infinity then σ_z^2 tends to unity (the value of the variance for the acentric Wilson distribution). This is the maximal uncertainty caused by the perfect overlapping condition.

(*c*) Equation (6) always provides σ_z^2 values lying in the interval (0,1) (see Fig. 2).

To treat clusters not constituted by perfectly overlapping reflections, we first specify when a reflection is defined as single, namely when no other reflection is closer than

$$\mathrm{tr}\theta = s \times \mathrm{FWHM},$$

where FWHM is the full width at half-maximum calculated at the cluster 2θ angle and s is a parameter fixed by the user (in our tests $s = 0.7$). Accordingly, two reflections for which

$$t = (2\theta_1 - 2\theta_2)/\mathrm{tr}\theta < 1$$

belong to the same cluster.

Let us now relax the condition of perfect overlapping. For a general cluster the following variance may be associated with the single intensity components:

$$\sigma_z^2 = (1 - t^q)[1 - S^2 \exp(-S) - (1 + S)^2 \exp(-2S)]. \qquad (7)$$

Equation (7) agrees well with common sense. Indeed:

(*a*) When t tends to unity σ_z^2 tends to zero, as for single reflections (for these reflections there is no error due to overlapping).

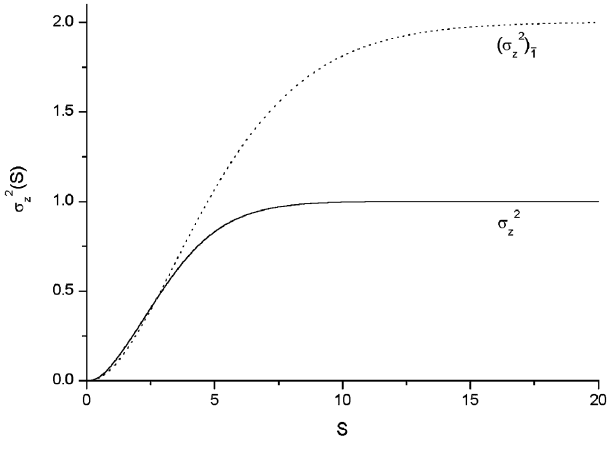


Figure 2
σ_z^2 and $(\sigma_z^2)_{\bar{1}}$ *versus* the overall cluster intensity S.

(*b*) If $t = 0$ we have a case of perfectly overlapping reflections, and (7) coincides with (6). Larger values of t diminish the variance value.

(*c*) q may be a positive, integral or fractional number – it damps or strengthens the effect of the t variable. In our tests we used $q = 2$.

The above results suggest that the presence of a symmetry element with a non-vanishing translation component may be better assessed by the average

$$\langle z_w \rangle = \left(\sum w_j z_j\right)/\left(\sum w_j\right), \qquad (8)$$

than by $\langle z \rangle$, where $w = 1/\sigma_z^2$. According to (8), $w = 1$ is the minimum weight to assign in the case of strong overlapping with a very intense reflection; in all other cases $w > 1$. Since $\sigma_z^2 = 0$ for single reflections and for reflections in a cluster with overall vanishing intensity, the criterion $w = 1/\sigma_z^2$ would cause a singularity. We prefer then to use the following weight:

$$w = \exp(-2\sigma_z^2),$$

giving $w = 1$ for single reflections and for reflections belonging to a cluster with vanishing intensity, and $w < 1$ for overlapping reflections.

Similar calculations may be performed for the centrosymmetric case. For example,

$$P(S) = (2\pi)^{-n/2} \exp(-S/2) \int_0^S \cdots \int_0^S \big[(z_1 z_2 \ldots z_{n-1}) \times (S - z_1 - z_2 - \cdots - z_{n-1})\big]^{-1/2}\,\mathrm{d}z_1 \ldots \mathrm{d}z_{n-1}.$$

The exact value of the right-hand side of the above equation is difficult to calculate for n larger than two. More practical is the use of the central limit theorem, which may be applied by calculating the following expression:

$$(\sigma_z^2)_{\bar{1}} = (2\pi)^{-1/2} \int_0^S z^{3/2} \exp(-z/2)\,\mathrm{d}z - \times (2\pi)^{-1} \left[\int_0^S z^{1/2} \exp(-z/2)\right]^2 \mathrm{d}z.$$

Since

$$\int_0^u z^{\nu-1} \exp(-\mu z)\,\mathrm{d}z = \mu^{-\nu}\gamma(\nu, \mu u),$$

where

$$\gamma(\alpha, x) = \int_0^x \exp(-t) t^{\alpha-1}\,\mathrm{d}t$$

is the incomplete gamma function, we obtain

$$(\sigma_z^2)_{\bar{1}} = \big[4/(\pi)^{1/2}\big]\big[\gamma(5/2, S/2) - (\pi)^{-1/2}\gamma^2(3/2, S/2)\big].$$

The values of $(\sigma_z^2)_{\bar{1}}$ (shown in Fig. 2) are always in the interval $(0, 2)$. It may be noted that, for $S = \infty$, $(\sigma_z^2)_{\bar{1}}$ coincides with the variance of the Wilson distribution for the centrosymmetric case.

Since the space group is unknown at this stage, we decided to simplify the procedure by applying the results obtained for

Table 1
Code names of the test structures, their space groups and the main references.

(1) ADAM	$Fm\bar{3}m$	Amoureux & Bee (1980)
(2) AGPZ	*Pbca*	Masciocchi *et al.* (1994)
(3) ALDS, ALDX	$P2_12_12_1$	SDPDRR (1998)
(4) ALPHA	$P2_12_12_1$	Harris *et al.* (2001)
(5) AND1, AND2, AND1N, AND2N	$P2_1/n$	Chernyshev *et al.* (1999)
(6) BACO	*C*2/*m*	Christensen (1992)
(7) BAMO	$P2_1$	Werner *et al.* (1997)
(8) BENZ	$P2_1/a$	Williams *et al.* (1992)
(9) CeO_2	$Fm\bar{3}m$	Brauer & Gradinger (1954)
(10) CF3BR	$P2_1/a$	Jouanneaux *et al.* (1992)
(11) CFCL	*Fdd*2	Cockcroft & Fitch (1991)
(12) CFI	*Cmca*	Clarke *et al.* (1993)
(13) CIME	$P2_1/n$	Cernik *et al.* (1991)
(14) CUPZ	*Pbca*	Masciocchi *et al.* (1994)
(15) DADA	$P2_12_12_1$	Dadachov & Le Bail (1997)
(16) DFQP	*I*2/*a*	Smrčok *et al.* (2001)
(17) EMT	$P6_3/mmc$	Baerlocher *et al.* (1994)
(18) F1A	$R\bar{3}$	Perez *et al.* (2000)
(19) FAMO	$P2_1/c$	Shankland *et al.* (2002)
(20) FORMYLUREA	$Pn2_1a$	Arumugam *et al.* (1992)
(21) GAPO	*Pbca*	Meden *et al.* (1997)
(22) IBUPS	$P2_1/c$	Shankland *et al.* (1998)
(23) KUOS	*P*4/*ncc*	Fitch & Cole (1991)
(24) LaB_6	$Pm\bar{3}m$	Eliseev *et al.* (1986)
(25) LAMO	$P2_1/a$	Hibble *et al.* (1988)
(26) LASI	$P2_1/c$	Christensen (1994)
(27) LASO	*C*2/*c*	Łasocha & Schenk (1997)
(28) LASOKA	*Pnma*	Łasocha *et al.* (2001)
(29) LEV	$R\bar{3}m$	McCusker (1993)
(30) MCM-22	*P*6/*mmm*	Camblor *et al.* (1998)
(31) MES	$P2_1/c$	Christensen *et al.* (1993)
(32) METYL	*I*222	Weiss *et al.* (1990)
(33) NAC	$I2_13$	Courbion & Ferey (1988)
(34) NBPO	*C*2/*c*	Zah-Letho *et al.* (1992)
(35) NIZR	$P2_1/n$	Jouanneaux, Verbaere, Piffard *et al.* (1991)
(36) NORB2	$P2_1/c$	Benet-Buchholz *et al.* (1998)
(37) OTHYM	$P2_1/n$	Kariuki *et al.* (1997)
(38) PBS	*Pbca*	Christensen *et al.* (1991)
(39) RUCO	*Ibam*	Masciocchi *et al.* (1993)
(40) SAPO-40	*Pmmn*	Estermann *et al.* (1992)
(41) SBPO	$P2_1/n$	Jouanneaux, Verbaere, Guyomard *et al.* (1991)
(42) SCAMPHOR1	$P2_12_12_1$	Brunelli *et al.* (2002)
(43) SGT	$I4_1/amd$	McCusker (1988)
(44) SULPH	*Pbcm*	Cockcroft & Fitch (1990)
(45) UTM1	*C*2/*m*	Plévert *et al.* (1999)
(46) VFI	$P6_3$	McCusker *et al.* (1991)
(47) VNI	$P4_12_12$	McCusker *et al.* (1996)
(48) Y_2O_3	$Ia\bar{3}$	Antic *et al.* (1995)
(49) YONO	$P2_1$	Christensen *et al.* (1992)
(50) YURI	$P2_1/c$	Andreev *et al.* (1997)

the non-centrosymmetric case to all the extinction symbols. This choice proved not to be critical.

4.2. About the scarcity of the experimental information

Quite often the number, n, of reflections obeying the extinction rule of a screw axis is quite small, thus weakening the predictive power of our probabilistic criteria. Assuming that $\langle z_w \rangle$ is sufficiently indicative only if it contains at least four reflections, we replace p(screw) as follows:

if p(screw) < 0.5 then p(screw) = p(screw) exp[$(4 - n)/8$],
if p(screw) > 0.5 then p(screw) = p(screw) exp[$-(4 - n)/8$].

4.3. A more stable statistical criterion

Let us denote by $\langle z_w \rangle_{[g]}$ and $\langle z_w \rangle_{[ng]}$ two weighted z averages, the first with respect to [g] (the set of reflections obeying the extinction rule) and the second with respect to [ng] (the complementary set). For example, for $c \perp \mathbf{a}$, [g] = {$0kl$} with $l = 2n + 1$, and [ng] = {$0kl$} with $l = 2n$.

It may occur that both the conditions $\langle z_w \rangle_{[g]} \ll 1$ and $\langle z_w \rangle_{[ng]} \ll 1$ hold; the first inequality suggests the presence of an extinction rule, the second reduces this expectation. It may also occur that $\langle z_w \rangle_{[g]} \simeq 1$ and $\langle z_w \rangle_{[ng]} \gg 1$. The first condition excludes the presence of an extinction rule, the second suggests more care. We then decided to derive the elementary probabilities of glide planes and centred unit cells by comparing the average $\langle z_w \rangle_{[g]}$ with $\langle z_w \rangle_{[ng]}$:

$$p(\text{glide}) \text{ and } p'(\text{centred cell}) = 1 - \langle z_w \rangle_{[g]} / \langle z_w \rangle_{[ng]}. \quad (9)$$

The criterion (9) proved to be sufficiently robust for practical applications; it does have a statistical basis, which may be described by illustrating a simple example. Let us suppose that the true space group is $P2_12_12_1$; if the space group was known *a priori*, we would prefer to decompose the diffraction pattern in that space group and normalize the intensities according to the Wilson rule:

$$z = |F|^2 \Big/ \Big(\varepsilon \sum_N \Big),$$

where $\varepsilon = 1$ always, except for the reflections (h00), (0k0) and (00l), for which $\varepsilon = 2$.

In the absence of any prior information and in accordance with §3, the pattern decomposition and normalization are both made in the space group $P2/m2/m2/m$, where

$$\epsilon = 1 \text{ for } (hkl); \ \epsilon = 2 \text{ for } (0kl), (h0l), (hk0); \ \epsilon = 4 \text{ for } (h00), (0k0), (00l).$$

As a practical result, the reflections (0kl), (h0l) and (hk0) will have $\langle z \rangle$ values smaller than their true values, so simulating the presence of glide planes. Since the same ε is applied to [g] and to [ng], the ratio $\langle z_w \rangle_{[ng]} / \langle z_w \rangle_{[g]}$ is insensitive to the error introduced by the normalization process and may constitute a more efficient criterion.

4.4. A limit for the elementary probabilities

The occurrence probability for an extinction symbol is the product of several elementary probabilities. If one of them is zero the other probability can no longer influence the final value. We have therefore imposed the following limitation on the elementary probabilities:

if $p < 0.1$ then $p = 0.1$.

Table 2
The correct extinction symbol (cexts), the order number orn at which it is ranked (list 1) by the probabilistic criteria (orn = 1 means that cexts has the highest probability value) and, when orn is not equal to 1, the first ranked extinction symbol (frexts); the order number orn2 as ranked in list 2 is given in parentheses.

MONOCLINIC Codename	cexts	orn	frexts	**u** (perc)	orn2
AND1	$P12_1/n1$	1			
AND1N	$P12_1/n1$	1			
AND2	$P12_1/n1$	1			
AND2N	$P12_1/n1$	1			
BENZ	$P12_1/a1$	1			
CIME	$P12_1/n1$	1			
FAMO	$P12_1/c1$	1			
IBUPS	$P12_1/c1$	1			
MES	$P12_1/c1$	1			
OTHYM	$P12_1/n1$	2	$P12_11$		
YONO	$P12_11$	2	$P12_1/c1$		
YURI	$P12_1/c1$	1			
BACO	$C1-1$	1		**a**/2 + **b**/2 (0.45)	(10)
BAMO	$P12_11$	2	$P12_1/n1$	**a**/2 + **b**/4 + **c**/2 (0.20)	(1)
CF3BR	$P12_1/a1$	1		**a**/3 + **b**/3 (0.17)	
DFQP	$I1a1$	1		**a**/2 + **b**/2 + **c**/2 (0.32)	(5)
LAMO	$P12_1/a1$	1		**a**/3 + **b**/3 (0.19)	
LASI	$P12_1/c1$	1		**a**/4 + **b**/4 (0.18)	
LASO	$C1c1$	2	$P12_11$	**a**/2 + **b**/2 (0.19)	(4)
NBPO	$C1c1$	1		**a**/2 + **b**/2 (0.30)	(5)
NIZR	$P12_1/n1$	3	$A1\ n1$	**b**/2 + **c**/2 (0.47)	(1)
NORB2	$P12_1/c1$	3	$P12_11$	**c**/3 (0.22)	
SBPO	$P12_1/n1$	5	$A1n1$	**b**/2 + **c**/2 (0.45)	(1)
UTM1	$C1-1$	2	$C1c1$	**a**/2 + **b**/2 (0.19)	(5)

ORTHORHOMBIC Codename	cexts	orn	frexts	**u** (perc)	orn2
ALDS	$P2_12_12_1$	1			
ALDX	$P2_12_12_1$	1			
DADA	$P2_12_12_1$	1			
GAPO	$Pbca$	1			
LASOKA	$Pn-a$	1			
SCAMPHOR1	$P2_12_12_1$	7	$P-nb$		
AGPZ	$Pbca$	34	$Ic-a$	**a**/2 + **b**/2 + **c**/2 (0.40)	(16)
ALPHA	$P2_12_12_1$	1		**b**/2 + **c**/2 (0.23)	(1)
CFCL	$Fdd-$	1		**a**/2 + **b**/2 (0.36)	(1)
CFI	$C-c(ab)$	1		**a**/2 + **b**/2 (0.28)	(31)
CUPZ	$Pbca$	16	$Ibca$	**a**/2 + **b**/2 + **c**/2 (0.45)	(1)
FORMYLUREA	$Pn-a$	1		**a**/2 (0.16)	(1)
METYL	$I---$	2	$I-(ac)-$	**a**/2 + **b**/2 + **c**/2 (0.49)	(82)
PBS	$Pbca$	1		**a**/2 + **b**/2 (0.46)	(11)
RUCO	$Iba-$	1		**c**/2 (0.58)	(1)
SAPO	$P--n$	1		**a**/3 (0.19)	
SULPH	$Pbc-$	1		**a**/2 + **b**/2 (0.17)	(1)

TETRAGONAL Codename	cexts	orn	frexts	**u** (perc)	orn2
VNI	$P4_12_1-$	18	$Pn-c$		
KUOS	$Pncc$	1		**c**/2 (0.89)	(4)
SGT	$Ia-d$	1		**a**/2 + **b**/2 + **c**/2 (0.59)	(9)

TRIGONAL/HEXAGONAL Codename	cexts	orn	frexts	**u** (perc)	orn2
EMT	$P--c$	1			
MCM	$P---$	6	$P6_3--$		
F1A	$R--$	5	$P6_2--$	2**a**/3 + **b**/3 + **c**/3 (0.40)	(8)

Table 2 (continued)

TRIGONAL/HEXAGONAL Codename	cexts	orn	frexts	**u** (perc)	orn2
LEV	$R--$	1		2**a**/3 + **b**/3 + **c**/3 (0.81)	(6)
VFI	$P6_3--$	2	$P-c-$	**a**/3 + **c**/3 (0.18)	(2)

CUBIC Codename	cexts	orn	frexts	**u** (perc)	orn2
ADAM	$F---$	1		**a**/2 + **b**/2, **a**/2 + **c**/2 (0.49)	
CeO_2	$F---$	1		**a**/2 + **b**/2, **a**/2 + **c**/2 (0.94)	
LaB_6	$P---$	4	$P4_2--$	**a**/4 + **c**/2 (0.20)	(4)
NAC	$I---$	1		**a**/2 + **b**/2 + **c**/2 (0.99)	(14)
Y_2O_3	$Ia--$	1		**a**/2 + **b**/2 + **c**/2 (0.98)	(15)

5. Applications

We have applied the statistical criteria described in the preceding sections to 50 test structures; their code names are quoted in Table 1. The structures cover all (but the triclinic) crystal systems; for the sake of brevity, for each structure we provide not the main crystallochemical data but only the true space group and references to where the reader can find supplementary information. We decided to verify the robustness of our algorithm by using a large set of test structures for several reasons:

(i) Each pattern shows features that are experiment dependent, *e.g.* a different degree of overlapping, a preferred orientation or a background of variable complexity. In particular, the set of test structures contains traditional X-ray generator data as well as synchrotron and neutron data.

(ii) Each pattern reflects eventual pseudosymmetries (*e.g.* of pseudotranslational nature) present in the crystal structure, thus making the identification of the correct extinction symbol difficult.

(iii) The definition of the correct extinction symbol is more difficult in some crystal systems.

All the above-mentioned features can influence the robustness of any probabilistic approach aimed at determining space-group symmetry; the method may be successful in some cases and inadequate in others. The use of a large set of test structures allowed us to verify the general applicability of our algorithm, and suggests to the reader the nature of the obstacles that they have to overcome in defining the correct space group.

We have applied the following procedure:

(*a*) The experimental diffraction patterns are indexed by the modulus *N-TREOR* of *EXPO2004*. In a few cases, *N-TREOR* is unable to index the pattern; some details about the failures are given in Appendix *A*.

(*b*) The experimental patterns are decomposed into single diffraction intensities by the modulus *EXTRA* (Altomare *et al.*, 1995) of *EXPO2004*; the resulting intensities are normalized according to §2.

(*c*) The probabilistic algorithm illustrated above is applied.

The results of our tests are shown in Table 2, where we provide, for each structure, the correct extinction symbol

(cexts), the order number at which it is ranked by our probabilistic criteria (orn; orn = 1 means that cexts has the highest probability value) and the first ranked extinction symbol (frexts).

Non-crystallographic symmetry can hinder the identification of the correct extinction group (*e.g.* non-crystallographic symmetry can simulate centred cells). To illustrate this statement, for each crystal system we regrouped the test structures into two sets; set *A* contains the structures for which *EXPO2004* did not find (*via* a statistical analysis of the diffraction intensities) any important pseudotranslational symmetry, and set *B* contains the structures with relevant pseudotranslational symmetry effects. For set *B*, perc > 0.16, where perc is the percentage of electrons satisfying the pseudosymmetry [see Altomare *et al.* (1996)]. The set *B* structures are the best candidates to show a centred cell or to simulate it. For this set, Table 2 indicates the most important pseudotranslational vector (in general, more vectors can coexist) and the corresponding perc value.

From Table 2, we observe the following:

(i) The extinction symbol is correctly identified (orn = 1) for 37 structures; orn = 2 for the other seven.

(ii) Identifying the extinction symbol is easier for the structures belonging to set *A*. There are a few exceptions, which will be commented on shortly. For OTHYM and YONO orn = 2, mostly because of the imperfect automatic estimation of the background; the user can easily improve the result after a visual inspection and obtain orn = 1 for both structures. For SCAMPHOR1 orn = 7. Fig. 3 shows that a non-negligible percentage of atoms have their *x* coordinates close to zero or their *z* coordinates close to 0.25. In such conditions the space group $P2_12_12_1$ may simulate the glide planes $n \perp \mathbf{b}$, $a \perp \mathbf{c}$ and $b \perp \mathbf{c}$, which explains why the selected extinction symbol is *P*−*nb* (and why *P*−*n*−, *Pbn*−, *Pbnb*, *Pnn*− and *P*−*nn* follow in order of orn) rather than $P2_12_12_1$.

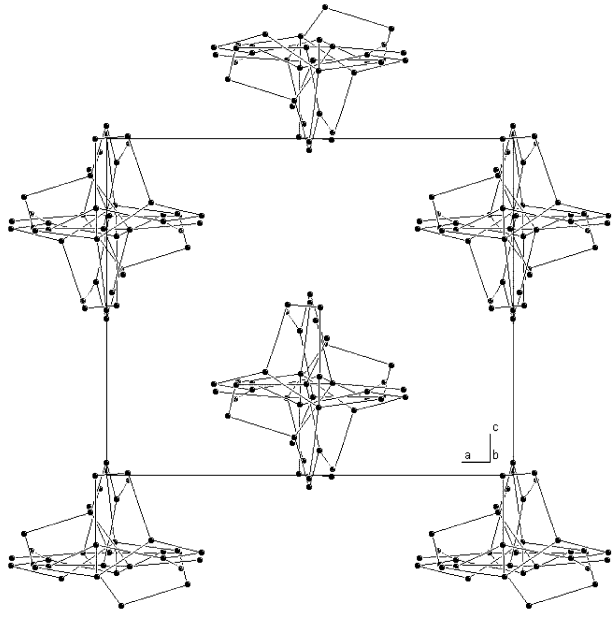


Figure 3
The crystal structure of SCAMPHOR1, viewed along the **b** axis.

The value of orn for VNI is 18; the extinction symbol of this structure (for which $c = 73.65$ Å) cannot be found without the crystallochemical information employed by McCusker *et al.* (1996).

(iii) The recognition of the correct extinction symbol for the test structures belonging to the set *B* is more difficult. The algorithm works well in some cases but fails badly in others (see, for example, AGPZ and CUPZ, where orn is 34 and 16, respectively). The failure is often due to the erroneous identification of the unit-cell type (it may be difficult to distinguish centred from pseudocentred cells), but also to erroneous identification of symmetry elements. We will take two examples to illustrate how the algorithm may be modified to face such problems.

For BAMO, *EXPO2004* found a pseudotranslational vector $\mathbf{u} = \mathbf{a}/2 + \mathbf{b}/4 + \mathbf{c}/2$, with perc = 0.20; the corresponding condition for the substructure reflections is $2h + k + 2l = 4n$. The pseudotranslational symmetry simulates a glide plane, $n \perp \mathbf{b}$, which contributes to the large probability value of the selected symbol $P-2_1/n-$ rather than to $P-2_1-$. In order to simplify the recognition of the correct extinction symbol we have modified the algorithm as follows. Suppose that a pseudotranslational symmetry and the corresponding perc value have been found by *EXPO2004*, and that our probabilistic algorithm ranks the extinction symbols in order of probability (list 1). If a *P* cell is chosen as the most probable one, a new run is automatically launched in which glide planes and screw axes simulated by the pseudotranslational symmetry are identified, and their established elementary probabilities are multiplied by $[1 - \mathrm{perc}(\mathbf{u})]^2$ (actually we do not modify the probabilities of the screw axes). A supplementary list of the most probable extinction symbols (list 2) is then added to the original one, which can help the user to identify the correct extinction symbol. In Table 2 we show in parentheses the orn value (orn2) corresponding to list 2. Remaining with the BAMO example, the probability of $P-2_1-$ and $P-2_1/n-$ are evaluated as follows: in list 1,

$$p(P-2_1/n-) = p(P)\,p(2_{1[010]})\,p(n \perp \mathbf{b}),$$

$$p(P-2_1-) = p(P)\,p(2_{1[010]})\,p(m \perp \mathbf{b}),$$

where $p(m \perp \mathbf{b}) = 1 - p(n \perp \mathbf{b})$.

In list 2

$$p(P-2_1/n-) = p(P)\,p(2_{1[010]})\,p(n \perp \mathbf{b})[1 - perc(\mathbf{u})]^2,$$

$$p(P-2_1-) = p(P)\,p(2_{1[010]})\left\{1 - [p(n \perp \mathbf{b})(1 - perc(\mathbf{u})]^2\right\}.$$

The above modifications lead to orn2 = 1 for BAMO (see Table 3). Generally speaking, list 2 provides the user with a different rank order alternative from that suggested by list 1.

Conversely, let us assume that a centred unit cell has been selected by our algorithm. If we suspect that the cell is actually pseudocentred, we have to diminish the elementary probability of screw axes and glide planes simulated by the centring. For example, for NIZR, *EXPO2004* found the

Table 3
BAMO; the first eight extinction symbols (ext. symb.) and their probability values (p) (in order of orn and orn2, respectively) as ranked in list 1 and list 2.

List1		List2	
Ext. symb.	p	Ext. symb.	p
$P12_1/n1$	0.527	$P12_1/n1$	0.378
$P12_11$	0.185	$P12_11$	0.334
$P1n1$	0.108	$P1n1$	0.077
$I1-1$	0.087	$P12_1/c1$	0.071
$P12_1/a1$	0.071	$P12_1/a1$	0.071
$P12_1/c1$	0.071	$I1-1$	0.068
$I1a1$	0.038	$P1-1$	0.055
$P1-1$	0.037	$I1a1$	0.037
$C1c1$	0.026	$C1c1$	0.026
$A1n1$	0.026	$A1n1$	0.026
$P1c1$	0.015	$P1c1$	0.015
$P1a1$	0.015	$P1a1$	0.015
$A1-1$	0.008	$A1-1$	0.008
$C1-1$	0.008	$C1-1$	0.008

pseudotranslation vector $\mathbf{u} = \mathbf{b}/2 + \mathbf{c}/2$ with perc = 0.47; the corresponding condition for the substructure reflections is $k + l = 2n$, corresponding to an A cell, and the selected extinction group is $A1n1$. We can rerun the program to reduce the probabilities of the centred cell, and of screw axes and glide planes simulated by the A centring (*e.g.* $n \perp \mathbf{a}$, $c \perp \mathbf{b}$, $b \perp \mathbf{c}$, $2_{1[010]}$ and $2_{1[001]}$); the corresponding elementary probabilities (actually, only of the centred cell and glide planes) are multiplied by $[1 - \mathrm{perc}(\mathbf{u})]^2$, and a supplementary list of the most probable extinction symbols (list 2) is again added to the first list. NIZR has orn2 = 1 in the new list.

AGPZ and CUPZ are more complicated examples, both crystallizing in space group *Pbca*, for which an *I*-centred cell has been erroneously selected (orn = 34 and 16, respectively). Both suffer from the same complicated pseudotranslational symmetry; $\mathbf{u} = \mathbf{a}/2$ (with perc = 0.37 and 0.41, respectively) and $\mathbf{u} = (\mathbf{b} + \mathbf{c})/2$ (with perc = 0.38 and 0.37, respectively), generating the additional *I*-centring vector $\mathbf{u} = (\mathbf{a} + \mathbf{b} + \mathbf{c})/2$. The corrections for the centring effects lead to orn2 = 1 and 16 for CUPZ and AGPZ, respectively.

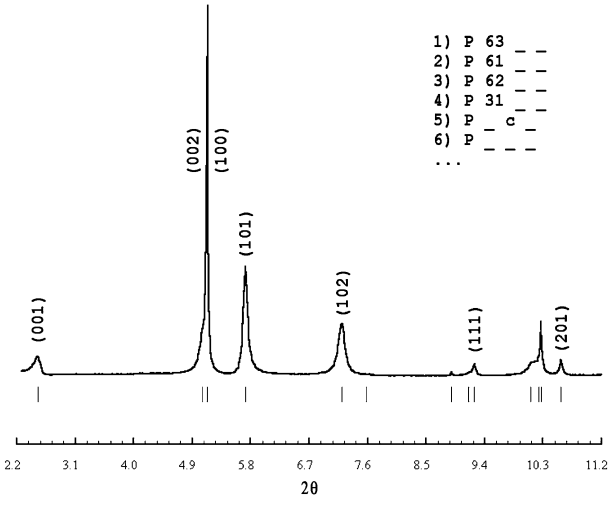


Figure 4
The first 11.2° of the diffraction pattern for MCM, with some indexed peaks. The figure contains a list of the first six diffraction symbols, in order of orn.

Table 4
List of the pseudotranslational vectors (**u**) managed by the algorithm.

For each **u**, the reflection condition and the simulated symmetry elements are given.

u	Reflection condition	Simulated symmetry elements
$\mathbf{a}/2$	$\leftrightarrow h = 2n$	$\rightarrow a \perp \mathbf{b}$; $a \perp \mathbf{c}$; $2_{1[100]}$
$\mathbf{b}/2$	$\leftrightarrow k = 2n$	$\rightarrow b \perp \mathbf{a}$; $b \perp \mathbf{c}$; $2_{1[010]}$
$\mathbf{c}/2$	$\leftrightarrow l = 2n$	$\rightarrow c \perp \mathbf{a}$; $c \perp \mathbf{b}$; $2_{1[001]}$
$\mathbf{a}/2 + \mathbf{b}/2$	$\leftrightarrow h+k = 2n$	$\rightarrow C$; $b \perp \mathbf{a}$; $a \perp \mathbf{b}$; $n \perp \mathbf{c}$; $2_{1[100]}$; $2_{1[010]}$
$\mathbf{a}/2 + \mathbf{b}/2$	$\leftrightarrow h+l = 2n$	$\rightarrow B$; $a \perp \mathbf{c}$; $c \perp \mathbf{a}$; $n \perp \mathbf{b}$; $2_{1[100]}$; $2_{1[001]}$
$\mathbf{b}/2 + \mathbf{c}/2$	$\leftrightarrow k+l = 2n$	$\rightarrow A$; $b \perp \mathbf{c}$; $c \perp \mathbf{b}$; $n \perp \mathbf{a}$; $2_{1[010]}$; $2_{1[001]}$
$\mathbf{a}/2 + \mathbf{b}/2 + \mathbf{c}/2$	$\leftrightarrow h+k+l = 2n$	$\rightarrow I$; $n \perp \mathbf{a}$; $n \perp \mathbf{b}$; $n \perp \mathbf{c}$; $2_{1[100]}$; $2_{1[010]}$; $2_{1[001]}$
$-\mathbf{a}/3 + \mathbf{b}/3 + \mathbf{e}/3$	$\leftrightarrow -h+k+l = 3n$	$\rightarrow R$; $3_{1[001]}$ or $6_{2[001]}$
$\mathbf{a}/2 + \mathbf{b}/4$	$\leftrightarrow 2h+k = 4n$	$\rightarrow a \perp \mathbf{b}$; $2_{1[100]}$
$\mathbf{a}/4 + \mathbf{b}/2$	$\leftrightarrow h+2k = 4n$	$\rightarrow b \perp \mathbf{a}$; $2_{1[010]}$
$\mathbf{a}/2 + \mathbf{c}/4$	$\leftrightarrow 2h+l = 4n$	$\rightarrow a \perp \mathbf{c}$; $2_{1[100]}$; $4_{1[001]}$
$\mathbf{a}/4 + \mathbf{c}/2$	$\leftrightarrow h+2l = 4n$	$\rightarrow c \perp \mathbf{a}$; $2_{1[001]}$
$\mathbf{b}/2 + \mathbf{c}/4$	$\leftrightarrow 2k+l = 4n$	$\rightarrow b \perp \mathbf{c}$; $2_{1[010]}$; $4_{1[001]}$
$\mathbf{b}/4 + \mathbf{c}/2$	$\leftrightarrow k+2l = 4n$	$\rightarrow c \perp \mathbf{b}$; $2_{1[001]}$
$\mathbf{a}/4 + \mathbf{b}/2 + \mathbf{c}/2$	$\leftrightarrow h+2k+2l = 4n$	$\rightarrow n \perp \mathbf{a}$; $2_{1[010]}$; $2_{1[001]}$
$\mathbf{a}/2 + \mathbf{b}/4 + \mathbf{c}/2$	$\leftrightarrow 2h+k+2l = 4n$	$\rightarrow n \perp \mathbf{b}$; $2_{1[100]}$; $2_{1[001]}$
$\mathbf{a}/2 + \mathbf{b}/2 + \mathbf{c}/4$	$\leftrightarrow 2h+2k+l = 4n$	$\rightarrow n \perp \mathbf{c}$; $2_{1[100]}$; $2_{1[010]}$

The program automatically manages a limited list of pseudotranslational vectors; these are listed in Table 4, together with the corresponding reflection conditions and simulated symmetry elements.

It may be worth noting that the two lists offer two different ranking criteria: the first is the pure application of the statistical criteria defined above, while the second aims to reduce the effects of the pseudotranslational symmetry. Usually, if an extinction symbol is highly ranked in one list, it is badly ranked in the second. Examination of the two lists can make it easier for the user to identify the correct extinction symbol.

(iv) Finally, we observe that in several cases (about ten) the discrimination in terms of probabilities between the correctly identified extinction group and the other groups is significant. For example, for ADAM $p = 0.547$ for the correct symbol and $p = 0.07$ for the next one.

Our last observation concerns the usefulness of direct visual inspection of the diffraction pattern. As with any other statistical method, our approach provides criteria based on the averages of useful variables rather than on the study of the profiles of specific peaks. Thus visual inspection of the pattern is a useful complement to our technique. We will describe two examples:

(*a*) MCM, space group *P*6/*mmm*. The extinction symbols suggested by the program are listed in Fig. 4 in order of orn, up to orn = 6, corresponding to the true symbol; in the same figure the first 11.2° of the diffraction diagram are shown (only some peaks are indexed). The first peak is (001), which immediately excludes all the extinction symbols up to orn = 5.

(*b*) CUPZ, space group *Pbca*. The extinction symbols suggested by the program are listed in Fig. 5 in order of orn, up to the correct one (orn = 16). In the same figure the 2θ interval (10, 25°) of the diffraction diagram is shown; the presence of reflection (131) excludes the *I* cell and the presence of reflection (121) excludes the *A* cell, so placing the correct extinction symbol at the first position.

6. Conclusions

We have described a statistically based algorithm that is able to provide a probability value for each diffraction symbol compatible with the established lattice symmetry. The algorithm has been implemented in the program *EXPO2004* and is fully automatic. In most cases the correct diffraction symbol has the largest probability value; in several cases (when the probability value associated with the correct diffraction symbol is much larger than the others) the diffraction symbol is unambiguously defined. For some test structures the correct diffraction symbol is not associated with the largest probability value; these cases have been analysed and it has been shown that the failures frequently result from the effects of pseudosymmetry.

APPENDIX *A*
About the indexing

Most of the test structures were indexed *via* default runs of *EXPO2004*. In some cases, a few changes in the automatic peak selection (addition and/or deletion of peaks) and special directives were necessary in order to detect the correct cell. We give below some details of the tricky cases.

AND1N and AND2N (*neutron data*). The correct cells are in the 'long output' file created by *N-TREOR*, but the values of the M_{20} figure of merit (M_{20} = 7 and M_{20} = 9, respectively) are too low. Synchrotron data of the two structures are easily and successfully indexed by *N-TREOR*, in agreement with Chernyshev *et al.* (1999). We used the unit-cell parameters arising from synchrotron data.

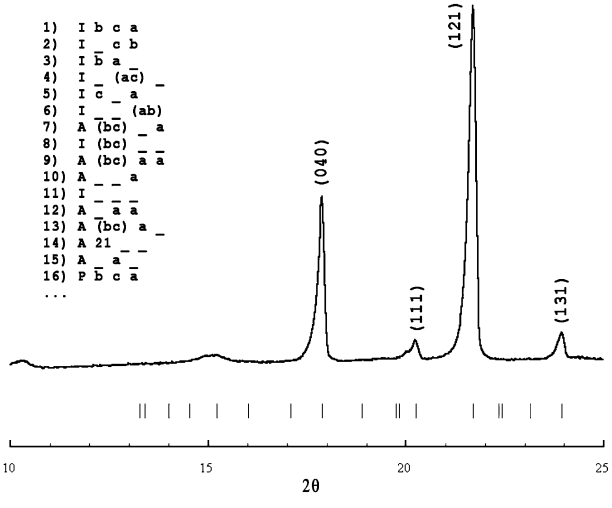


Figure 5
The interval 10–25° of the diffraction pattern for CUPZ, with some indexed peaks. The figure contains the list of the first 16 diffraction symbols, in order of orn.

CFI (*neutron data*). The correct indexing requires very high values of the Miller indices for the lowest-angle reflections. We used the unit cell provided by Clarke *et al.* (1993).

RUCO (*X-ray laboratory data*). The difficulty in indexing is related to the presence of few and very broad experimental peaks. *N-TREOR* finds an orthorhombic cell, in agreement with Masciocchi *et al.* (1993), the *c* axis of which was doubled by these authors '*... in order to avoid stacking (along c) of eclipsed Ru–C–O units of planar* $Ru(CO)_4$ *moieties*'. We used the same criterion for our applications.

VNI (*synchrotron data*). The *c* axis is quite long (*c* = 73.6505 Å). In agreement with McCusker *et al.* (1996), *N-TREOR* is able (by enlarging the default maximum value of the investigated axis cell) to find a cell that is one-half of the correct one (*a* = 9.8823 Å and *c* = 36.8277 Å). The authors doubled the *c* axis in order to index a few small peaks. If these peaks, in place of some low-angle peaks, are used by *N-TREOR* and *ad hoc* directives are employed, our program is able to index the structure (*a* = 9.8837 Å and c = 73.6505 Å) correctly.

References

Altomare, A., Burla, M. C., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G. & Polidori, G. (1995). *J. Appl. Cryst.* **28**, 842–846.
Altomare, A., Caliandro, C., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2004). *J. Appl. Cryst.* **37**, 1025–1028.
Altomare, A., Foadi, J., Giacovazzo, C., Guagliardi, A. & Moliterni, A. G. G. (1996). *J. Appl. Cryst.* **29**, 674–681.
Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). *J. Appl. Cryst.* **33**, 1180–1186.
Amoureux, J. P. & Bee, M. (1980). *J. Phys. C Solid State Phys.* **13**, 3577–3583.
Andreev, Y. G., MacGlashan, G. S. & Bruce, P. G. (1997). *Phys. Rev. B*, **55**, 12011–12017.
Antic, B., Mitric, M. & Rodic, D. (1995). *J. Magn. Magn. Mater.* **145**, 349–356.
Arumugam, S., Glidewell, C. & Harris, K. D. M. (1992). *J. Chem. Soc. Chem. Commun.* pp. 724–726.
Baerlocher, C., McCusker, L. B. & Chiappetta, R. (1994). *Microporous Mater.* **2**, 269–280.
Benet-Buchholz, J., Haumann, T. & Boese, R. (1998). *J. Chem. Soc. Chem. Commun.* pp. 2003–2004.
Brauer, G. & Gradinger, H. (1954). *Z. Anorg. Allg. Chem.* **277**, 89–95.
Brunelli, M., Fitch, A. N. & Mora, A. J. (2002). *J. Solid State Chem.* **163**, 253–258.
Camblor, M. A., Corma, A., Díaz-Cabañas, M.-J. & Baerlocher, C. (1998). *J. Phys. Chem. B*, **102**, 44–51.
Cernik, R. J., Cheetham, A. K., Prout, C. K., Watkin, D. J., Wilkinson, A. P. & Willis, B. T. M. (1991). *J. Appl. Cryst.* **24**, 222–226.
Chernyshev, V. V., Fitch, A. N., Sonneveld, E. J., Kurbakov, A. I., Makarov, V. A. & Tafenko, V. A. (1999). *Acta Cryst.* B**55**, 554–562.
Christensen, A. N. (1992). *Acta Chem. Scand.* **46**, 240–243.
Christensen, A. N. (1994). *Z. Kristallogr.* **209**, 7–13.
Christensen, A. N., Hazell, R. G., Hewat, A. W. & O'Reilly, P. J. (1991). *Acta Chem. Scand.* **45**, 469–473.
Christensen, A. N., Hazell, R. G., Lehmann, M. S. & Nielsen, M. (1993). *Acta Chem. Scand.* **47**, 753–756.

Christensen, A. N., Nielsen, M., O'Reilly, K. P. J. & Wroblewski, T. (1992). *Acta Chem. Scand.* **46**, 224–230.
Clarke, S., Cockcroft, J. K. & Fitch, A. N. (1993). *Z. Kristallogr.* **206**, 87–95.
Cockcroft, J. K. & Fitch, A. N. (1990). *Z. Kristallogr.* **193**, 1–19.
Cockcroft, J. K. & Fitch, A. N. (1991). *Z. Kristallogr.* **197**, 121–130.
Courbion, G. & Ferey, G. (1988). *J. Solid State Chem.* **76**, 426–431.
Dadachov, M. S. & Le Bail, A. (1997). *Eur. J. Solid State Inorg. Chem.* **34**, 381–390.
Eliseev, A. A., Efremmov, V. A., Kuzmicheva, G. M., Konovalova, E. S., Lazorenko, V. I., Paterno, Y. B. & Khlyustova, S. Y. (1986). *Kristallografiya*, **31**, 803–805.
Estermann, M. A., McCusker, L. B. & Baerlocher, C. (1992). *J. Appl. Cryst.* **25**, 539–543.
Fitch, A. N. & Cole, M. (1991). *Mater. Res. Bull.* **26**, 407–414.
Harris, K. D. M., Treymayne, M. & Kariuki, B. M. (2001). *Angew. Chem. Int. Ed.* **40**, 1626–1651.
Hibble, S. J., Cheetham, A. K., Bogle, A. R. L., Wakerley, H. R. & Cox, D. E. (1988). *J. Am. Chem. Soc.* **110**, 3295–3296.
Jouanneaux, A., Fitch, A. N. & Cockcroft, J. K. (1992). *Mol. Phys.* **71**, 45–50.
Jouanneaux, A., Verbaere, A., Guyomard, D., Piffard, Y., Oyetola, S. & Fitch, A. N. (1991). *Eur. J. Solid State Inorg. Chem.* **28**, 755–765.
Jouanneaux, A., Verbaere, A., Piffard, Y., Fitch, A. N. & Kinoshita, M. (1991). *Eur. J. Solid State Inorg. Chem.* **28**, 683–699.
Kariuki, B. M., Serrano-González, H., Johnston, R. L. & Harris, H. D. M. (1997). *Chem. Phys. Lett.* **280**, 189–195.
Łasocha, W., Milart, P., Rafalska-Łasocha, A. & Schenk, H. (2001). *Z. Kristallogr.* **216**, 117–121.
Łasocha, W. & Schenk, H. (1997). *J. Appl. Cryst.* **30**, 909–914.
Le Bail, A., Duroy, H. & Fourquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.
McCusker, L. (1988). *J. Appl. Cryst.* **21**, 305–310.
McCusker, L. B. (1993). *Mater. Sci. Forum*, **133–136**, 423–434.
McCusker, L. B., Baerlocher, Ch., Jahn, E. & Bülow, M. (1991). *Zeolites*, pp. 308–313.
McCusker, L. B., Grosse-Kunstleve, R. W., Baerlocher, C., Yoshikawa, M. & Davis, M. E. (1996). *Microporous Mater.* **6**, 295–309.
Markvardsen, A. J., David, W. I. F., Johnson, J. C. & Shankland, K. (2001). *Acta Cryst.* A**57**, 47–54.
Masciocchi, N., Moret, M., Cairati, P., Ragaini, F. & Sironi, A. (1993). *J. Chem. Soc. Dalton Trans.* pp. 471–475.
Masciocchi, N., Moret, M., Cairati, P., Sironi, A., Ardizzoia, G. A. & La Monica, G. (1994). *J. Am. Chem. Soc.* **116**, 7668.
Meden, A., Grosse-Kunstleve, R. W., Baerlocher, C. & McCusker, L. B. (1997). *Z. Kristallogr.* **212**, 801–807.
Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
Perez, C., Espinola, C. G., Foces-Foces, C., Nunez-Coello, P., Carrasco, H. & Martin, J. D. (2000). *Org. Lett.* **2**, 1185–1188.
Plévert, J., Yamamoto, K., Chiari, G. & Tatsumi, T. (1999). *J. Phys. Chem. B*, **103**, 8647–8649.
Shankland, K., David, W. I. F., Csoka, T. & McBride, L. (1998). *Int. J. Pharm.* **165**, 117–126.
Shankland, K., McBride, L., David, W. I. F., Shankland, N. & Steele, G. (2002). *J. Appl. Cryst.* **35**, 443–454.
Smrčok, L'., Kopelhuber, B., Shankland, K., David, W. I. F., Tunega, D. & Resel, R. (2001). *Z. Kristallogr.* **216**, 63–66.
SDPDRR (1998). *Structure Determination by Powder Diffractometry Round Robin*, http://www.cristal.Org/SDPDRR/index.html.
Weiss, E., Corbelin, S., Cockcroft, J. K. & Fitch, A. N. (1990). *Chem. Ber.* **123**, 1629–1634.
Werner, P.-E., Moustiakimov, M., Marinder, B.-O. & Knight, K. S. (1997). *Z. Kristallogr.* **212**, 665–670.
Williams, J. H., Cockcroft, J. K. & Fitch, A. N. (1992). *Angew. Chem. Int. Ed. Engl.* **31**, 1655–1657.
Zah-Letho, J. J., Jouanneaux, A., Fitch, A. N., Verbaere, A. & Tournoux, M. (1992). *Eur. J. Solid State Inorg. Chem.* **29**, 1309–1320.