Journal of
Applied
Crystallography

ISSN 0021-8898

Received 30 March 2005
Accepted 30 June 2005

© 2005 International Union of Crystallography
Printed in Great Britain – all rights reserved

Space group determination: improvements in *EXPO2004*

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

[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]Departimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [d]Dipartamento 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

In the ideal process for crystal structure solution, the determination of the space group usually follows the indexation step. When powder data are used, peak overlap, background definition and possible impurity peaks may hinder the identification of the correct extinction group. The recent automatic procedure implemented in *EXPO2004* has been further developed; the modifications are based on additional statistical criteria, with new algorithms for a better definition of the background and for the removal of impurity peaks. The new version also offers a more powerful graphical interface. Failures occur rarely, as shown by examples.

1. Introduction

Crystal structure solution from powder data is not a straightforward process, owing to two main problems: the uncertainty of the background level and, more fundamentally, the overlapping of the diffraction peaks. Such problems affect any procedure for estimating the integrated intensities, often resulting in a failure to recognize the true extinction group. This is the reason why careful visual inspection of the observed diffraction pattern is still the most used approach for determining the space group.

Markvardsen *et al.* (2001) proposed a Bayesian probabilistic approach for estimating the relative probabilities of the different extinction symbols; the intensities were extracted by the Pawley (1981) method. These tests involved a small number of structures and obtained good results, but also showed how difficult the recognition of the correct extinction group can be.

We have recently developed (Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.*, 2004) a new algorithm based on the statistical analysis of the normalized z intensities. Once the unit cell has been determined, the full pattern is decomposed according to the Le Bail method (Le Bail *et al.*, 1988) in the space group having the largest Laue symmetry compatible with the assumed crystal system and no extinction condition. The extracted intensities are then normalized according to the classical Wilson method, and the z statistics are used to define the probability of the extinction symbol. A proper weight is assigned to each z value to take overlapping into account. The algorithm has been proved to work well for a large set of test structures, covering all the crystal systems; however, it was not able, for a few structures, to unambiguously recognize the true extinction group.

To improve efficiency of this alogorithm, we have modified the above procedure as follows: (*a*) a new algorithm has been written to find, and eventually remove, impurity peaks (see §2); (*b*) the automatic procedure for estimating the background has been improved (see §3); (*c*) a more significant expression for the variance σ_z^2 has been derived (see §4); (*d*) an additional criterion has been used to define the extinction group (see §5); (*e*) the graphical interface has been improved, in order to facilitate the direct visual inspection of the systematically absent reflections corresponding to each extinction symbol selected by the user (see §6).

For comparison, the new procedure has been tested with the same set of test structures used by Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.* (2004) (see Table 1), using a modified version of the program *EXPO2004* (Altomare, Caliandro, Camalli, Cuocci, Giacovazzo *et al.*, 2004).

2. Impurity peak removal

Impurity peaks, when not eliminated from the diffraction pattern, can affect the intensity extraction process, with detrimental effects on the efficiency of the procedures for space group determination and for crystal structure solution. A new algorithm has been developed and introduced in *EXPO2004* to identify and delete impurity peaks eventually present in the powder diffraction pattern.

Let us suppose that a peak search procedure has been used (*e.g.* Altomare *et al.*, 2000), that the diffraction pattern has been indexed and that a full-pattern decomposition has been

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 (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 (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).

accomplished. The simplest condition for recognizing an impurity peak is the following: the 2θ distance of the peak from the closest reflection is larger than the FWHM. Impurity peaks are also recognized as follows:

(1) The pairs of contiguous reflections **h** and **k** for which

$$2\theta_{\mathbf{k}} - 2\theta_{\mathbf{h}} \geq [\mathrm{FWHM}(\mathbf{h}) + \mathrm{FWHM}(\mathbf{k})]/2$$

are considered.

(2) The intensity $y(i)$ measured at each ith count step between $2\theta_{\mathbf{h}}$ and $2\theta_{\mathbf{k}}$ is checked; the presence of an impurity peak is assessed if the next conditions are verified for at least one i:

$$2\theta_i - 2\theta_{\mathbf{h}} > \mathrm{FWHM}(\mathbf{h})/2, \quad (1a)$$

$$2\theta_{\mathbf{k}} - 2\theta_i > \mathrm{FWHM}(\mathbf{k})/2, \quad (1b)$$

$$y(i) > y(\mathbf{h}) \text{ and } y(\mathbf{k}).$$

The peaks considered as impurities are deleted by using a filter technique as follows:

(a) A Gaussian profile $y^{\mathrm{c}}(i)$ centred at $2\theta_{\max}$ is calculated; $2\theta_{\max}$ is the 2θ position corresponding to the maximum intensity (say $y_{\max}$) of the points satisfying the conditions stated at step 2. We have

$$y^{\mathrm{c}}(i) = \begin{cases} y_{\max} G(\Delta 2\theta_i) & i \in \chi_{\mathrm{peak}} \\ 0.0 & \text{otherwise,} \end{cases}$$

where $(\Delta 2\theta_i) = 2\theta_{\max} - 2\theta_i$, $G(\Delta 2\theta_i)$ is the Gaussian profile function and χ_{peak} is the set of 2θ points belonging to the assumed peak extension ($2\theta_1 = 2\theta_{\max} - 2\mathrm{FWHM}$, $2\theta_2 = 2\theta_{\max} + 2\mathrm{FWHM}$).

(b) The observed profile count $y(i)$ is replaced by

$$y^n(i) = y(i) - y^{\mathrm{c}}(i).$$

In Fig. 1 we show how this algorithm works for CFCL (see Table 1) into the 2θ interval 26.5–29°. The intensity extraction process (the first step of the procedure for the determination of the extinction group), performed in the space group *Pmmm*, is unable to fit the experimental profile in the range 26.5–29° by means of the two reflections lying at 27.96 and 28.46° (see Fig. 1*a*). The intensities and the FWHM values of the two reflections are y_1 = 736, and y_2 = 855, 0.399 and 0.400, respectively. The 2θ range that satisfies conditions (1*a*) and (1*b*) is (28.16°, 28.26°) and, actually, there are three experimental points in this range with intensities (1392, 1452 and 1454, respectively) higher than both y_1 and y_2.

Our algorithm corrects the experimental profile as in Fig. 1(*b*), providing nearly vanishing intensities to the two reflections. Actually, these two reflections are extinct for the right extinction group *Fdd*2; accordingly the application of our algorithm makes the determination of the correct space group easier.

3. Background improvement

The background estimate plays a central role in the determination of the integrated intensities extracted from a powder diffraction pattern, particularly for small intensities, which are the most sensitive to background modification. The recently developed space group determination procedure (Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.*, 2004) uses statistical analysis of the normalized structure factors extracted from the experimental profile. This technique strongly depends on the peak intensities and requires accurate background evaluation. The automatic background determination procedure in *EXPO* (Altomare *et al.*, 1995, 1999) is based on the assumption that the observed diffraction profile counts outside the reflection peak areas are background counts. The procedure is efficient, but it may fail for two main reasons:

(1) Impurity peaks are present in the pattern. In this case, since no reflection corresponds to the impurity peak, any automatic procedure will assume that the impurity peak counts are background points. Consequently, the background curve fitting those points will be overestimated (sometimes severely).

(2) At very large 2θ values, where peak overlap is strong, it is difficult to discriminate background from Bragg signal.

Owing to the above two considerations, the strategy of *EXPO* allowed the user to easily modify, *via* a graphical interface, the background points fixed by the automatic procedure, by adding or deleting background points. This option is maintained in *EXPO2004*, but we implemented in it a new automatic procedure, which should avoid the two above-mentioned sources of error. It is described below.

The *EXPO* strategy of dividing the observed pattern into intervals (Altomare *et al.*, 1995) is maintained. In each interval a peak search procedure is used for locating the intensity maxima of the experimental profile after having multiplied the count values by -1. The 2θ positions corresponding to such maxima provide, for each interval, an initial list of background points. Since they may occasionally be located in peak areas, some of them do not correspond to noise but to structure contribution. In order to discard them the following steps are executed:

(*a*) Let $y(i)$ be the observed intensity corresponding to the ith background point. The polynomial curve best fitting the background points is calculated by a least-squares procedure, which associates to each point a weight $w(i) = 0.5$ if

$$y(i) \geq 1.1\, y(i-1) \text{ or } y(i) \geq 1.1\, y(i+1);$$

otherwise $w(i)$ is 1.0.

(*b*) Let $\mathrm{bg}(i)$ be the polynomial value at the ith point. We consider the ratio

$$\mathrm{rbg}(i) = [\,y(i) - \mathrm{bg}(i)]/\mathrm{bg}(i),$$

its mean value

$$\mathrm{rbg}_{\mathrm{med}} = \frac{\sum_{i=1}^{N} \mathrm{rbg}(i)}{N}$$

and its standard deviation

$$\mathrm{rbg}_{\mathrm{stdv}} = \left\{ \frac{\sum_{i=1}^{N} [\mathrm{rbg}(i) - \mathrm{rbg}_{\mathrm{med}}]^2}{N-1} \right\}^{1/2}.$$

The summation is over the N points satisfying the condition $y(i) > \mathrm{bg}(i)$ (the points in the starting list with larger intensity values). If $\mathrm{rbg}_{\mathrm{stdv}} < 0.01$, the process passes to step (*d*); if $\mathrm{rbg}(i) > \mathrm{sogh} \times \mathrm{rbg}_{\mathrm{stdv}}$, the ith point is removed from the set of background points [sogh is a suitable threshold whose value is clarified in step (*c*)]. The aim of step (*b*) is to remove background points that have large deviations from the best fitted background polynomial (*e.g.* points corresponding to impurity contributions and points located in the peak areas). If no point is eliminated, the procedure jumps to step (*d*).

(*c*) The selected background points are used for a new polynomial calculation. The process comes back to (*a*) and is repeated up to ten times. The sogh value depends on the cycle number nc:

$$\mathrm{sogh} = 0.5 + 0.05(\mathrm{nc} - 1).$$

(*d*) If the 2θ distance between two adjacent background points is larger than 0.5° then some background points are added (at intervals of 0.5°). Their intensity values are obtained by interpolation. Step (*d*) aims at avoiding large 2θ intervals without background points, which may ill-condition the least-squares background polynomial calculation.

(*e*) The background points are used for a new polynomial calculation.

(*f*) The percentage pg of the pattern points whose current background values $\mathrm{bg}(i)$ are larger than the observed values $y(i)$ is calculated with respect to the full number of observations in the profile. If pg is not in the interval (0.05, 0.08), the background value in each point of the pattern is increased (if $\mathrm{pg} < 0.05$) or decreased (if $\mathrm{pg} > 0.08$) by constant factors (0.98 for decreasing and 1.003 for increasing). Step (*f*) is repeated until the condition on pg is satisfied. This procedure is particularly effective when the background is underestimated (*e.g.* when the pattern is noisy).

The new procedure provides a more reasonable background evaluation and, combined with the other new features described in this paper, contributes to improve the identification of the extinction group (see §6). An example is given for OTHYM (see Table 1) (2θ range: 5.0–69.97°; step: 0.0194; number of data points: 3350). *EXPO2004* divides the experimental pattern into two partially overlapping 2θ intervals. In the first range (5.00°, 35.50°) there are 1573 data points; in the second (35.30°, 69.97°) there are 1890 data points. The initial

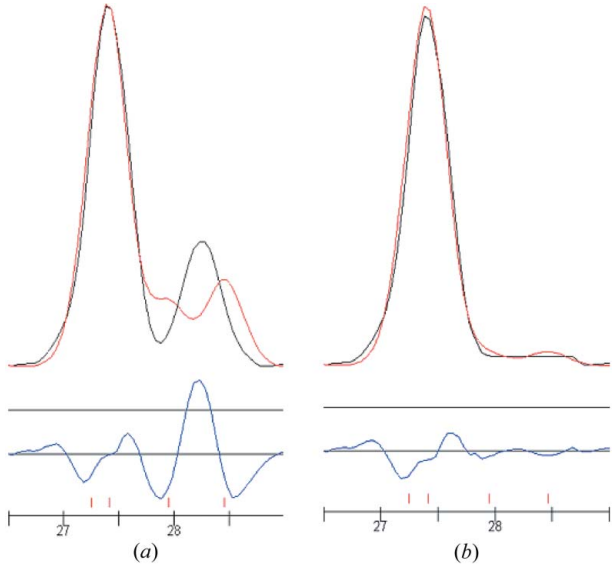


Figure 1
CFCL: experimental diffraction pattern after the intensity extraction procedure. The black line corresponds to the experimental profile, the red line to the calculated profile and the blue line to the difference profile. (*a*) The 2θ interval 26.5–29°; (*b*) the same interval after the application of the impurity removal algorithm.

background point list provided by the peak search consists of 73 points in the first interval and 98 in the second. Steps from (*a*) to (*c*) discard 39 large intensity points in the first and 78 in the second interval. Step (*d*) adds 18 interpolating points in the first and 31 in the second interval.

In Fig. 2 the background evaluations up to 41° 2θ by the old (Fig. 2*a*) and the new (Fig. 2*b*) procedures are given. It may be useful to note the improvement in the new polynomial in the range 26–35°.

4. A more significant σ_z^2

Let S be the measured intensity of a cluster of n perfectly overlapping reflections with intensity $z_1, z_2, \ldots, z_n$. The variance associated with the generic intensity component z may be expressed as (Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.*, 2004)

$$\sigma^2(z) = \int_0^S z^2 P(z)\,\mathrm{d}z - \left[\int_0^S z P(z)\,\mathrm{d}z\right]^2, \qquad (2)$$

where $P(z)$ is the Wilson z distribution and $S = z_1 + z_2 + \cdots + z_n$. In the absence of any information about the space group, $P(z)$ was assumed to be acentric and

$$\sigma^2(z) = 1 - S^2\exp(-S) - (1+S)^2\exp(-2S).$$

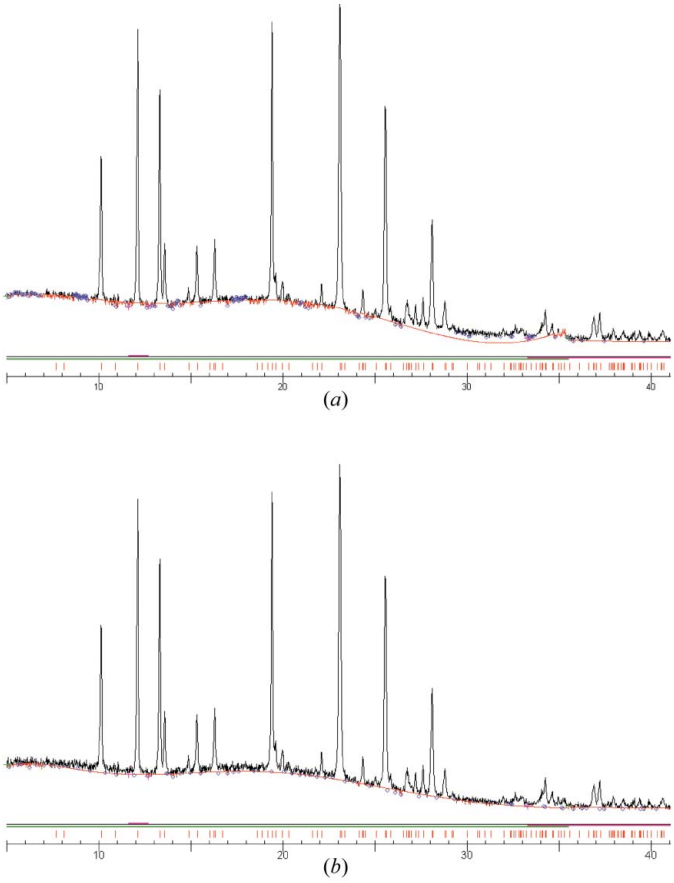


Figure 2
OTHYM: background calculated by the old (*a*) and the new (*b*) procedure.

For a generic cluster for which the overlapping is not complete, equation (2) may be generalized as follows:

$$\sigma^2(z) = (1 - t^q)[1 - S^2\exp(-S) - (1+S)^2\exp(-2S)] \qquad (3)$$

where

$$t = \Delta 2\theta/(0.7\mathrm{FWHM});$$

$\Delta 2\theta$ is the 2θ position distance of the reflection from the closest reflection in the cluster.

The weight

$$w = \exp[-2\sigma^2(z)]$$

was then associated with each reflection, and

$$z_w = wz.$$

The average

$$\langle z_w\rangle = \left(\sum w_j z_j\right)/\left(\sum w_j\right) \qquad (4)$$

was used to establish the probability of each extinction group.

The above mathematical treatment ignores the multiplicity m_i, $i = 1, \ldots, n$ of the reflections belonging to the cluster. In our new model we assume $S = m_1 z_1 + m_2 z_2 + \cdots + m_n z_n$ and we replace equation (2) by

$$\sigma^2(z_i) = \int_0^{S_i'} z_i^2 P(z_i)\,\mathrm{d}z_i - \left[\int_0^{S_i'} z_i P(z_i)\,\mathrm{d}z_i\right]^2,$$

where $S_i' = S/m_i$. As a consequence, equation (3) is replaced by

$$\sigma^2(z_i) = (1 - t^q)[1 - S_i'^2\exp(-S_i') - (1+S_i')^2\exp(-2S_i')].$$

A numerical example will show the difference between the estimates obtained by the old and the new procedure. Let us consider, in the orthorhombic crystal system, two completely overlapping reflections with indices (132) and (070), respectively. Let $z_1 = 1.0$, $m_1 = 8$ and $z_2 = 0.1$, $m_2 = 2$. Then

$$S_1 = 1.025,\ \sigma^2(z_1) = 0.09,\ w_1 = 0.84,$$
$$S_2 = 4.1, \sigma^2(z_2) = 0.71, w_2 = 0.24.$$

While the old treatment associates the same weight with both the completely overlapping reflections, the new one assigns a smaller weight to the reflection with small multiplicity, thus increasing the probability of recognizing a symmetry element with translational symmetry [note that reflections systematically absent because of the symmetry operator **C** = (**R**, **T**) must satisfy the condition **hR** = **h** with non integer **hT** value, and this condition automatically reduces their multiplicity].

5. More selective criteria for identifying the correct extinction group

If we are examining single-crystal data under the hypothesis of the space group SG, and we find a few non-zero intensities corresponding to reflections that should be systematically absent, we should reject (by excluding Renninger effects) the assumed SG, even in the case in which $\langle z\rangle$ is close to zero (the average is over the systematically absent reflections). There-

Table 2
For each test structure, the correct extinction symbol (cexts) and the order number (orn) at which it is ranked by the automatic procedure for the identification of the extinction symbol.

The order number orn2 refers to the case in which the pseudotranslation correction is applied. The columns headed Previous and New refer to the standard *EXPO2004* procedure and to that described in this paper, respectively.

MONOCLINIC		Previous		New	
Code name	cexts	orn	orn2	orn	orn2
AND1	$P12_1/n1$	1		1	
AND1N	$P12_1/n1$	1		1	
AND2	$P12_1/n1$	1		1	
AND2N	$P12_1/n1$	1		1	
BENZ	$P12_1/a1$	1		1	
CIME	$P12_1/n1$	1		1	
FAMO	$P12_1/c1$	1		1	
IBUPS	$P12_1/c1$	1		1	
MES	$P12_1/c1$	1		2	
OTHYM	$P12_1/n1$	2		1	
YONO	$P12_11$	2		1	
YURI	$P12_1/c1$	1		1	
BACO	$C1-1$	1	10	1	10
BAMO	$P12_11$	2	1	2	1
CF3BR	$P12_1/a1$	1		1	
DFQP	$I1a1$	1	5	1	5
LAMO	$P12_1/a1$	1		1	
LASI	$P12_1/c1$	1		1	
LASO	$C1c1$	2	4	1	3
NBPO	$C1c1$	1	5	1	2
NIZR	$P12_1/n1$	3	1	1	1
NORB2	$P12_1/c1$	3		1	
SBPO	$P12_1/n1$	5	1	1	1
UTM1	$C1-1$	2	5	1	3

ORTHORHOMBIC		Previous		New	
Code name	cexts	orn	orn2	orn	orn2
ALDS	$P2_12_12_1$	1		1	
ALDX	$P2_12_12_1$	1		1	
DADA	$P2_12_12_1$	1		3	
GAPO	$Pbca$	1		1	
LASOKA	$Pn-a$	1		1	
SCAMPHOR1	$P2_12_12_1$	7		2	
AGPZ	$Pbca$	34	16	29	2
ALPHA	$P2_12_12_1$	1	1	1	1
CFCL	$Fdd-$	1	1	1	1
CFI	$C-c(ab)$	1	31	1	16
CUPZ	$Pbca$	16	1	10	1
FORMYLUREA	$Pn-a$	1	1	1	1
METYL	$I---$	2	82	2	44
PBS	$Pbca$	1	11	1	2
RUCO	$Iba-$	1	1	3	15
SAPO	$P--n$	1		1	
SULPH	$Pbc-$	1	1	1	1

TETRAGONAL		Previous		New	
Code name	cexts	orn	orn2	orn	orn2
VNI	$P4_12_1-$	18		3	3
KUOS	$Pncc$	1	4	1	9
SGT	$Ia-d$	1	9	1	9

Table 2 (continued)

TRIGONAL/HEXAGONAL		Previous		New	
Code name	cexts	orn	orn2	orn	orn2
EMT	$P--c$	1		1	
MCM	$P---$	6		2	
F1A	$R--$	5	8	7	7
LEV	$R--$	1	6	1	5
VFI	$P6_3--$	2	2	2	2

CUBIC		Previous		New	
Code name	cexts	orn	orn2	orn	orn2
ADAM	$F---$	1		1	7
CeO_2	$F---$	1		3	16
LaB_6	$P---$	4	4	4	4
NAC	$I---$	1	14	1	18
Y_2O_3	$Ia--$	1	15	1	12

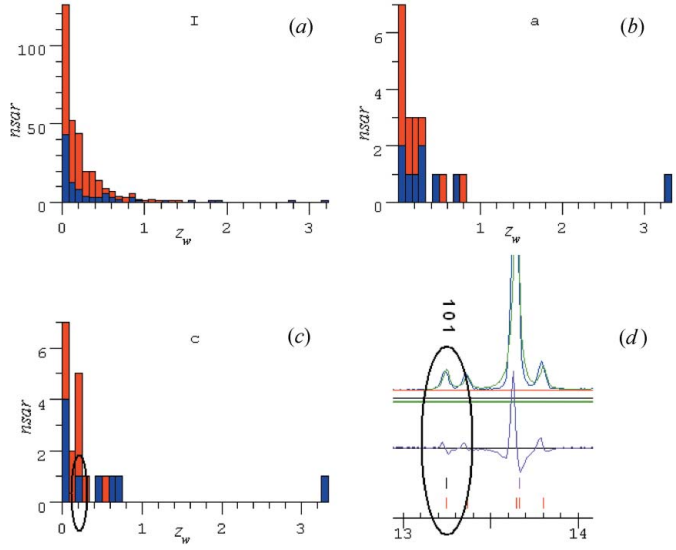


Figure 3
Histogram plots of METYL, corresponding to the symmetry operators (*a*) *I*, (*b*) *a* and (*c*) *c*. (*d*) Part of the experimental pattern.

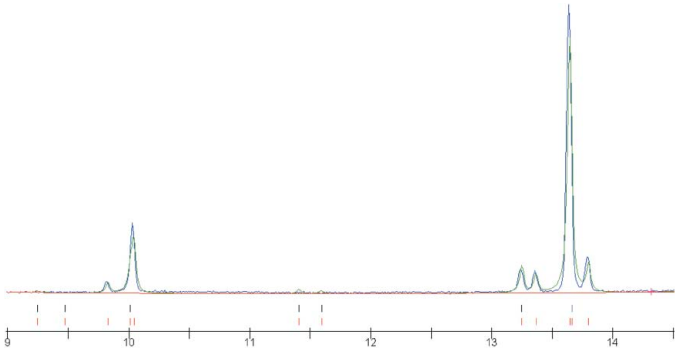


Figure 4
Details of the diffraction pattern for METYL in the range 9–14.5°: vertical bars at the bottom show reflection positions for the $P2/m2/m2/m$ space group, while the upper black bars show extinct reflections for the $I-(ac)-$ extinction group (black bars correspond to reflections that are single or that belong to a whole extinct cluster).

fore the statistical analysis adopted by Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.* (2004) for selecting the most probable extinction group from powder data, based only on the z statistics, may fail in some cases. The risk is higher for powder data because reflections overlap in clusters and each z value may be wrongly evaluated by the decomposition algorithm. In order to improve the efficiency of the score based on the z statistics, we have introduced in the new automatic procedure the following additional criteria:

(*a*) if a cell is supposed to be centred on the basis of the z statistics, but four or more single (*i.e.* not overlapping with any other) reflections, expected to be systematically absent, are found with $d \geq 2.25$ Å and $z_w \geq 0.45$, then the hypothesis is rejected;

(*b*) if a glide plane is supposed to be present on the basis of the z statistics, but two or more single reflections, expected to be systematically absent, are found with $d \geq 2.25$ Å and $z_w \geq 0.45$, then the hypothesis is rejected;

(*c*) the same procedure is adopted for a supposed screw axis when two single reflections are found with $d \geq 2$ Å and $z_w \geq 0.35$.

If, for a given extinction group, at least one of the previous conditions is verified, the probability value of the extinction group provided by the z statistics [say $P_{\rm old}$(ext. group)] is replaced by

$$P_{\rm new}(\text{ext. group}) = 0.01 P_{\rm old}(\text{ext. group}).$$

To gain experimental information, clusters constituted by reflections with reduced multiplicity when this is fixed by the same symmetry operator [*e.g.* in the orthorhombic system when both the reflections are of type ($hk0$) with $h + k$ = odd] are also tested by the above criteria.

6. Experimental results by the automatic procedure

We have applied the new procedure to the test structures quoted in Table 1. The results are summarized in Table 2, where for each test structure we give the codename, the correct extinction symbol (cexts), and the order number (orn) as ranked by the previous procedure and by the new one. In the case in which the program finds a non-negligible pseudotranslational symmetry, the probabilities of the extinction symbols are modified as explained by Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.* (2004); the new values of orn (say orn2) are reported for both the procedures. The code names of the test structures for which a pseudotranslational symmetry is found are shown in bold print; the reader who wants to consider the pseudotranslational vectors and their powers is referred to Table 2 of Altomare, Caliandro, Camalli, Cuocci, da Silva *et al.* (2004). We note the following:

(*a*) In the monoclinic system the efficiency of the new procedure is quite considerable, particularly for the structures for which a pseudotranslational symmetry is found (the most difficult cases for the previous procedure): orn = 2 only for MES. For BAMO orn = 2 but orn2 = 1.

(*b*) For the orthorhombic system the mean value of orn is 4.2 for the old procedure and 3.5 for the new one. In particular, orn (and/or orn2) markedly improves for SCAMPHOR1, AGPZ and CUPZ.

(*c*) Both the old and the new procedure provide orn = 1 for two of the three tetragonal test structures (KUOS and SGT). For the third structure (VNI) orn jumps from 18 to 3. This is a big improvement if we recall that the c axis of VNI is 73.65 Å and that the correct extinction symbol cannot be found without the crystal chemical information employed by McCusker *et al.* (1996).

(*d*) For the five test structures belonging to the hexagonal system, the average value of orn reduces from 3 to 2.6, while for the five structures belonging to the cubic system, using the new procedure increases the average value of orn from 1.6 to 2.

7. Improvement of the graphical interface

The new automatic procedure is able to recognize the correct extinction symbol for 42 of the 54 test structures. For these structures, orn (and/or orn2) is 1. One can ask if the user, by manual inspection of the diffraction diagram, can discard the extinction symbols that have been wrongly assigned a probability higher than the true one. To improve the user control we have modified the standard graphical interface of *EXPO2004*. A new interface has been added, which provides a window showing the extinction symbols ranked according to their probability; when the user clicks on one of them, the list of compatible space groups is supplied. In addition, the new graphical interface allows the user to gain the following information:

(*a*) A list of the reflections that, according to the selected extinction symbol, are expected to be systematically absent. For each of these, the following specifications are given: the type of reflection (single or overlapping with other reflections), the corresponding z_w value [as provided by equation (4)] and the symmetry operator(s) responsible for the extinction [*e.g.* the glide plane operator responsible, in space group $P2_1/c$, for the extinction of the reflection (103).]

(*b*) A histogram for each symmetry operator, giving the number of systematically absent reflections (nsar) *versus* the z_w values. The histogram provides an immediate evaluation of the distribution of the intensities associated with potentially systematically absent reflections. As an example, we consider the test structure METYL; the highest scoring extinction group is $I-(ac)-$ ($P = 0.214$); the second highest is $I---$ ($P = 0.213$), which is the correct one. For $I-(ac)-$ we show in Figs. 3(*a*), 3(*b*) and 3(*c*) three histograms corresponding to the three symmetry operators I, a and c, respectively. Fig. 3(*c*) shows, in the ellipse, a single reflection (blue colour) that is expected to be systematically absent. In Fig. 3(*d*), the corresponding peak is shown; the extinction symbol $I-(ac)-$ is thus clearly eliminated from the list of candidates.

(*c*) The reflections that are expected to be absent for the selected extinction symbol are marked by a vertical bar [see

Fig. 4, where the diffraction pattern of METYL in the range 9–14.5° is shown according to the extinction group $I-(ac)-$]. The bars are situated below the observed pattern, just above the set of bars defining the reflection positions, and have different colours depending on whether the reflection is single (or belongs to a whole extinct cluster) or overlapping. By default, a black vertical bar marks single systematically absent reflections, while the rest of the absent reflections (those that are overlapping) are marked with a purple bar.

The new graphical interface allows easy correction of most of the errors made by the automatic procedure. As an example, we consider the case of METYL. The top-ranked extinction symbol is $I-(ac)-$; $I---$, the true group, is ranked second. The experimental diagram shows that, for example at $2\theta = 13.26°$, an evident peak (see Fig. 4) marked (by the black vertical bar) as being absent is in fact present, and therefore the user can safely discard $I-(ac)-$ in favour of $I---$.

Working with this visual help, for the DADA, METYL, MCM, CeO_2 and LaB_6 cases, the extinction groups with higher probability than the true one can be discarded with just a quick inspection of the pattern.

8. Conclusions

The automatic procedure for space-group determination implemented in *EXPO2004* has been optimized. The new procedure employs new statistical criteria, reduces the errors in the background definition, eliminates impurity peaks and offers to the user a more powerful graphical interface. As a result, the extinction group is often unambiguously determined. When this does not happen, the graphical support helps the user to discard unreliable diffraction symbols even if estimated with large probability.

References

Altomare, A., Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G. & Rizzi, R. (1999). *J. Appl. Cryst.* **32**, 339–340.

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, R., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2004). *J. Appl. Cryst.* **37**, 1025–1028.

Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *J. Appl. Cryst.* **37**, 957–966.

Altomare, A., Giacovazzo, C., Gagliardi, 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). *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. B. (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**, 7668–7676.

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.
SDPDRR (1998). *Structure Determination by Powder Diffractometry Round Robin*, http://www.cristal.Org/SDPDRR/index.html.
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.
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.