Z. Kristallogr. Suppl. **26** (2007) 191-196
© by Oldenbourg Wissenschaftsverlag, München

Powder pattern indexing and the dichotomy algorithm

D. Louër[1,*], A. Boultif[2]

[1]Rennes, France
[2]Département de Cristallographie, Institut de Physique, Université Mentouri, 25000 Constantine, Algeria
[*]Contact author; e-mail: Daniel.Louer@free.fr

Keywords: powder diffraction, pattern indexing, dichotomy method, computer program, DICVOL

Abstract. A short review of the development of the successive dichotomy method for indexing powder diffraction patterns is presented. The use of logical filters in the tree-type approach is discussed, particularly for searching smallest unit cells according to the crystallographic convention. An additional option, based on an extended search, has been implemented for unstable cases. The new version of the program, DICVOL06, offers the two options to the users, i.e. DICVOL04 with its optimized search procedure and an extended search. A few examples are used for comparing CPU times with the two options.

Introduction

Indexing of powder diffraction patterns is a major application of powder diffraction [1]. With the advent of the computer technology, the first efficient approaches were introduced in the late sixties and seventies (see the reviews on computer programs for pattern indexing reported by Shirley [2,3]). Among the modern methods, the successive dichotomy algorithm [4] is based on an optimal exhaustive strategy. It has been applied in many *ab initio* structure determinations from powder data. The program developed in parallel to the speed of the computers. Its modern version [5] has been considerably improved, with facilities adapted to the experimental features of the measured powder data. A short review of the successive developments of the dichotomy method and a new option based on an extended search, incorporated in DICVOL06, are reported in the present study.

The indexing problem and the dichotomy algorithm

Pattern indexing is based on the quadratic equation obtained by squaring the diffraction vectors $\mathbf{d}_{hkl}$*, i.e. $Q_i(hkl) = f(a,b,c,\alpha,\beta,\gamma,h,k,l)$ $(= d_i^{*2})$ where a, b, c, α, β, γ are the lattice parameters and $h_ik_il_i$ the Miller indices. Indexing a powder diffraction pattern consists in obtaining a, b, c, α, β, γ, h_i, k_i, l_i, which are unknown, from the known set of measured Q_i. The problem cannot be solved by ordinary algebra, since whatever how many equations are con-

sidered, there are always more unknowns than equations. The dichotomy method (method of division in halves) is based on a scan of the parameter space by dividing cell edges and angles into equal intervals (i.e. 0.4 Å and 5° for linear and angular parameters, respectively), so that the entire space is divided into volume domains. Each volume domain is then transformed into a calculated pattern expressed in terms of ranges $[Q_-(hkl),Q_+(hkl)]$. A domain can contain a solution if for the observed lines Q_i:

$$Q_-(hkl) - \Delta Q_i < Q_i < Q_+(hkl) + \Delta Q_i, \quad (1)$$

where ΔQ_i is the absolute error on observed Q_i. If so, each domain edge is bisected to generate volume sub-domains analysed as before. The dichotomy process is applied at six levels, with volume domain edges progressively reduced. Each crystal lattice is explored separately, starting from the cubic side of the symmetry sequence. The strategy of the dichotomy algorithm consists of searching smallest cell volume(s) according to the universal crystallographic convention, i.e. the smallest repeat unit that has the full symmetry of the crystal. The procedure is a tree-type approach, which is optimized with criteria for discarding or retaining mathematical solutions. The figure of merit (FoM) M_N [6] is used as a filter to discard solutions with too low FoMs. M_N includes two ideas, the average discrepancy between observed and calculated Q_i and the size of the unit cell (the smaller they are, the greater is M_N). Final checks on validity of solutions are based on a review of all available data; new refined parameters and FoM are then obtained. In the WinPLOTR/FULLPROF suite [7], in which DICVOL04 is implemented, an additional useful check is performed with a pattern matching facility (Le Bail algorithm).

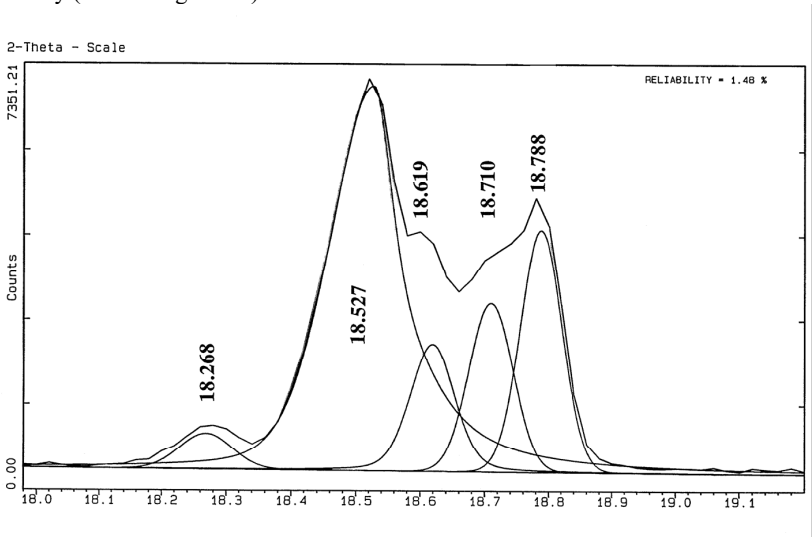

Figure 1. Pattern decomposition of a cluster of diffraction lines, using pseudo-Voigt functions, with the interactive fitting program PROFILE (λ = 1.5406 Å). Peak positions are expressed in °(2θ).

Data requirements

Powder pattern indexing requires high quality data to be confident in the physical solution suggested by a computer program. They must be collected with a correctly adjusted powder diffractometer. The effect of systematic and random errors has been discussed elsewhere [8]. The refinement of an extra degree of freedom, named 'zero-point', available in DICVOL04 [5], allows to a certain extent an overall correction of usual aberrations (e.g. zero-point shift, sample surface displacement, transparency and umbrella effects, capillary displacement). In practice, peak positions are extracted with automatic procedures, see for instance the software WinPLOTR [7]. Interactive pattern decomposition is useful too, particularly for the detection of shoulders in line profiles. An example of decomposition into elementary components of a cluster of lines, carried out with the profile fitting PC-program PROFILE from Socabim, is shown in figure 1. The peak positions of the diffraction lines are extracted with the program and used subsequently for indexing.

Successive programs until DICVOL06

With the introduction of the successive dichotomy method in 1972 [4] two computer programs for the indexing of high symmetry patterns were also reported. With the advances in computer technology, the method and the programs have been progressively extended to lower symmetry cases. The successive programs appeared as follows.

- P1, down to orthorhombic lattice symmetry [4]. In the related program P2, density and molecular weight information were introduced, speeding the calculations.
- DICVOL, down to monoclinic symmetry [9]. A search strategy based on a partition of the volume space into 400 $Å^3$ shells of volumes was introduced in order to find first solutions with smallest cell volumes.
- DICVOL91, down to triclinic symmetry [10]. A detailed mathematical analysis of dichotomy domain limits for monoclinic and triclinic systems was reported. For computing time considerations, unlike the higher symmetry cases for which the dichotomy method has been applied in the parameter space (direct space), the triclinic case has been investigated in the Q-space with constraints on the indices of the first two lines.
- DICVOL04 [5] was a significant improvement of the DICVOL91 version. In addition to an optimization of the triclinic case, the program includes new facilities, such as a tolerance for spurious lines, a 'zero-point' error analysis, a refinement of the 'zero-point' shift together with the cell parameters, a cell analysis in terms of reduced cell and a review of all available powder data for helping space group derivation. The tolerance for spurious lines concerns both impurity lines and lines for which the input error on peak position is too small. This tolerance is applied at all levels of the dichotomy process. However, at the final stage the lattice parameters are least-squares refined from the peak positions with a greater error range, which means that lines rejected at the last dichotomy level may, by chance, be indexed in the output file.
- DICVOL06, reported here, offers to the user two options in which logical filters are applied differently, i.e. DICVOL04 and an extended search in shells of volumes where a solution is found. Compared to the initial program (P1), the reduction of CPU times with the more recent programs can be seen in Table 1.

Table 1. CPU times (sec) for indexing the patterns of two orthorhombic compounds using the first program P1 (CII 10070 computer) and DICVOL91 or DICVOL06 (Dell Latitude D800 laptop).

	P1 [4]	DICVOL91/06
BiF_3	260	0.01
β-Cd(OH)Br	300	0.01

Additionally, in the program LOSH [11] a dichotomy search was also used for locating the reciprocal angles α* and β* of a triclinic cell. This subroutine was subsequently combined to the Visser's zone indexing algorithm [12] in the combination-strategy program LZON [11].

Logical filters

The dichotomy algorithm is a tree-type approach devised for searching the mathematical solutions to the indexing of a powder pattern, within volume space limits. Solutions are then discarded/retained, namely filtered, according to two types of logical filters.
(i) *User-decision-dependent filters*, which are the ranges scanned by the parameters and volume space, the absolute errors on peak position measurements, the minimum value of the figure of merit M_N for retaining a solution.
(ii) *Strategy dependent filters*, which drive the tree-type approach. As described above, the crystallographic definition of the unit cell requires searching 'mathematical' solutions with smallest cell volumes, which is an important feature in the dichotomy algorithm. Indeed, as soon as a solution is found, only new solutions with smaller volumes should be searched. This strategy is used in DICVOL04, with however a tolerance to take into account experimental errors and dichotomy interval limits. This is an optimized strategy with normal quality data. It has been found powerful from the indexing of a great number of powder diffraction data collected with usual care. For some data sets, where precision is low or spurious lines are present or, simply, unstable cases (e.g. dominant zones, large axis case, solutions with close cell volumes, etc.), an extended search can be helpful. This is the purpose of an option implemented in the new version of DICVOL, i.e. DICVOL06. As soon as a solution is found in a 400 $Å^3$ shell of volume, unlike the DICVOL04 option, the exhaustive search is extended to the whole domain. Then, all solutions within 400 $Å^3$ are displayed, according to the filters defined in (i). Of course, the number of mathematical solutions is, generally, greater. To make easier the selection of the physical solution, they are ranked according to their figure of merit M_N in a second output file with the extension 'ord'. However, the suggested solutions must be critically analysed.
DICVOL06 is available via the CCP14 website (http://www.ccp14.ac.uk).

Tests with the new extended search option

The two options, i.e. DICVOL04 and extended search, have been tested with powder data used elsewhere [5,13]. Calculations were carried out with a Dell Latitude D800 laptop. Applications to powder data of pharmaceutical compounds collected with the capillary tech-

nique (Cu $K\alpha_1$) are reported in Table 2. The increase of CPU time is small for most compounds. However, it can be significant as soon as impurity lines are present (see aspirin).

Table 2. CPU times (sec) for indexing the powder data of pharmaceutical compounds.

	M_{20}	DICVOL04	Extended search
Barbital sodium (Orthorhombic)	69	< 1	< 1
Phenacetin (Monoclinic)	56	2	3
Theophylline (Orthorhombic)	97	< 1	< 1
Sulfaguanidine (Monoclinic)	49	2	3
Aspirin (Monoclinic) [3 spurious lines]	25	20	48
Urea (Tetragonal)	235	< 1	< 1
Piracetam Form I (Monoclinic)	158	< 1	2
Piracetam Form II (Triclinic)	69	2	3
Piracetam Form III (Monoclinic)	26	2	3

The patterns of the 71 substances in section 20 of the NBS Monograph 25 [14] have also been studied. The average CPU times (sec) for the two 'DICVOL04/Extended-search' options were 0.01/0.08 for 9 cubic patterns, 0.03/0.12 for 11 tetragonal patterns, 0.03/0.30 for 17 hexagonal patterns, 0.06/0.11 for 16 orthorhombic patterns and 2.50/14.30 for 18 monoclinic patterns. Only the data of monoclinic Eu_2O_3 were indexed with a shorter CPU time using the extended search, i.e. 51 sec against 152 sec with DICVOL04. This is due to the relaxation in the 'extended search' of a constraint used in DICVOL04 and to the presence of a dominant zone in the powder data ($a/b = 3.92$). Indeed, in DICVOL04, to speed the calculation the search is first carried out with linear parameter limits equal to twice the first observed d-spacing, which is generally appropriate for monoclinic cases. If there is no solution, as for Eu_2O_3, the search is then extended to the input maximum values.

All 29 triclinic powder patterns in the NBS monograph 25 have also been studied. From the DICVOL04 to the extended search options, CPU times increase by variable percentages. However, with the extended search, 23 patterns are indexed in less than 360 sec, including 15 patterns in less that 60 sec.

In case of lattice metric singularity, where two (or more) cells have an identical set of d-spacings, the indexing solution is not unique. As discussed elsewhere [5], the strategy used in DICVOL, based on lattices scanned successively is appropriate. Using data reported recently (Table 3 in ref. 15) for the quaternary lattice metric singularity described for tetragonal mawsonite [15], the two options of DICVOL06 give in automatic mode three of the cells with smallest volumes in each symmetry, i.e. cubic ($V = 1238.8$ Å^3), tetragonal ($V = 309.7$ Å^3) and orthorhombic ($V = 154.8$ Å^3). The fourth cell (orthorhombic, $V = 825.9$ Å^3) is only displayed among about 15 mathematical solutions by acting on the facilities available in DICVOL06 (extended search).

Concluding remarks

The implementation in the program DICVOL of an option with an extended search of solutions can be helpful in presence of difficult indexing problems. However, this option should not be regarded as a panacea, since for low quality data, the number of solutions increases and their discrimination on the basis of FOMs may not be easy. With DICVOL06, a good practice remains the use of the option based on DICVOL04 with good quality data. For pattern indexing, the collection of precise data remains a basic recommendation. Practices in presence of impurity lines and short or long axis cases have been discussed elsewhere [5]. Additional aspects related to the dichotomy algorithm, e.g. the number of lines required for indexing and two-phases patterns, have also been discussed [13]. Finally, the indexing of the pattern of turkey egg-white lysozyme with a large hexagonal cell volume, 374 057 $Å^3$ [16], is a representative example of the performance of the dichotomy method.

References

1. Langford, J.I. & Louër, D., 1996, *Rep. Prog. Phys.*, **59**, 131-234.
2. Shirley, R., 1978, in *Computing in Crystallography*, edited by H. Schenk, R. Oltyof-Hazekamp, H. van Koningsveld & G.C. Bassi (Delft: Delft University Press), pp. 221-234.
3. Shirley, R., 1980, in *Accuracy in Powder Diffraction*, NBS Spec. Pub. No. 567, edited by S. Block & C.R. Hubbard, (Gaithersburg, MA: US Department of Commerce), pp. 361-382.
4. Louër, D. & Louër, M., 1972, *J. Appl. Crystallogr.*, **5**, 271-275.
5. Boultif, A. & Louër, D., 2004, *J. Appl. Crystallogr.*, **37**, 724-731.
6. Wolff, P.M. de, 1968, *J. Appl. Crystallogr.*, **1**, 108-112.
7. Roisnel, T. & Rodriguez-Carvajal, J. (2001). *Mat. Sci. Forum*, **378-381**, 118-123.
8. Louër, D., 1992, in *Accuracy in Powder Diffraction II*, NIST Spec. Pub. No. 846, edited by E. Prince & J.K. Stalick, (Gaithersburg, MA: US Department of Commerce), pp. 92-104.
9. Louër, D. & Vargas, R., 1982, *J. Appl. Crystallogr.*, **15**, 542-545.
10. Boultif, A. & Louër, D., 1991, *J. Appl. Crystallogr.*, **24**, 987-993.
11. Shirley, R. & Louër, D., 1978, *Acta Crystallogr.*, **A34**, S382 (abstract).
12. Visser, J.W., 1969, *J. Appl. Crystallogr.*, **2**, 89-95.
13. Louër, D. & Boultif, A., 2006, *Z. Kristallogr. Suppl.*, **23**, 225-230.
14. Morris M.C., McMurdie, H.F., Evans, E.H., Paretzkin, B., Parker, H.S., Pyrros, N.P. & Hubbard, C.R., 1984, in *Monograph No. 25*, Sect. 20, National Bureau of Standards (Washington: U. S. Department of Commerce).
15. Mighell, A.D., 2006, *J. Res. Natl. Inst. Stand. Technol.*, **111**, 393-399.
16. Margiolaki, I., Wright, J.P., Fitch, A.N., Fox, G.C. & Von Dreele, R.B., 2005, *Acta Crystallogr.*, **D61**, 423-432.