Computer Programs

J. Appl. Cryst. (1985). **18**, 367–370

TREOR, a semi-exhaustive trial-and-error powder indexing program for all symmetries.

By P.-E. WERNER, L. ERIKSSON and M. WESTDAHL, *Department of Structural Chemistry, Arrhenius Laboratory, University of Stockholm, S-106 91 Stockholm, Sweden*

(*Received* 2 *January* 1985; *accepted* 16 *April* 1985)

Abstract

An indexing program, *TREOR*, mainly based on trial-and-error methods is described. The program contains separate routines for cubic, tetragonal, hexagonal, orthorhombic, monoclinic and triclinic symmetries. Ten years usage has been analysed to improve the original program. For monoclinic indexing a specific short-axis test has been developed. The over-all success rate of the program has been found to be better than 90%, and considerably more for orthorhombic and higher symmetries.

Introduction

A number of programs to facilitate the indexing of powder patterns, based on trial-and-error methods, were written and described twenty years ago (Werner, 1964). The computer technology of that time did not allow a rigorous implementation of the principles. The original programs had to be coded in machine language, but in the 1974 the first version of a trial-and-error indexing program, *TREOR*, was written in Fortran. Some new principles were introduced, such as use of the powerful de Wolff figure-of-merit test (de Wolff, 1968) for assessing the reliability of a unit cell derived solely from powder data.

TREOR searches for solutions in index space by varying the Miller indices, and it has been classified by Shirley (1980) as semi-exhaustive. The term was proposed for programs containing 'judicious deductions to limit the size of the solution field in order to gain speed'. The program was written as a collection of procedures for separate treatment of cubic, tetragonal, hexagonal, orthorhombic and monoclinic symmetry.

In the new version of *TREOR*, described in the present article, triclinic symmetry and more efficient algorithms for monoclinic symmetry have been included.

New features

The original idea to start with cubic symmetry and in a stepwise manner test for lower symmetries is retained, because it increases the chance of success. Thus, hexagonal patterns that could not be indexed as such may be indexed as orthorhombic and monoclinic patterns as triclinic. A reduction program can be used to convert the unit cell found to the higher metric symmetry. It is also strongly recommended to check cubic, tetragonal and hexagonal solutions by orthorhombic tests. One reason for this is that it may help to identify geometrical ambiguities that can arise when the metrical symmetry is higher than orthorhombic (Mighell & Santoro, 1975).

The algorithm used in *TREOR* for the triclinic symmetry is more focused on the solution of genuine triclinic patterns than the corresponding algorithms originally used by Kohlbeck & Hörl (1976) for treatment of monoclinic and triclinic lattices by the same procedure. The reason is that a truly triclinic pattern can usually be given lower index limits for the low-order lines than a monoclinic pattern rewritten as a triclinic one.

In the following, the term basis line is used for a diffraction line that is given tentative indices in order to derive a unit cell. Five basis-line sets will generally be sufficient for orthorhombic tests, whereas seven sets even will sometimes be insufficient for a monoclinic pattern. The monoclinic failures have usually been caused by unit cells having one cell axis notably shorter than the two others, thus giving rise to dominant zones. More than the five first consecutive lines of the pattern may then be indexed with a common zero index. We have therefore added a special short-axis test for the monoclinic symmetry. In this test the first three lines are tentatively indexed as a two-dimensional pattern, *i.e.* one index is assumed to be zero. If the following three lines can be indexed by the two-dimensional lattice parameters, then the first unindexable line(s) is/are used to derive the remaining monoclinic cell parameter(s). Depending on whether the two-dimensional lattice is rectangular, *i.e.* a (*hk*0) or (0*kl*) zone, or not, *i.e.* a (*h*0*l*) zone, the program will derive two or one more parameter(s). The trial cells derived by this procedure are then tested for fit with the remaining lines, and the best solutions are saved for later refinement. The solutions are ranked primarily according to the number of indexable lines and secondarily according to the smallest cell volume. The principle used by Taupin (1973), to reject a trial lattice if a line cannot be indexed and this line is not explicitly declared as possibly extraneous, is not used. This is avoided because of the experience that it is often difficult to identify an extraneous line *a priori*.

In order to improve further the indexing of monoclinic lattices we have added a fast procedure suggested by Smith & Kahara (1975) for finding the (020) reflection. The relation

$$2Q_{020} + Q_{h10} = Q_{h30}$$

($Q = \sin^2\theta$) is used as '020 detector'. If relations of this type are found, the monoclinic *b* axis is directly determined and in successive trials ($h = 1$, 2 and 3) the *a* axis is defined. Thereafter, the remaining two parameters *c* and β are defined by trial indexing of two low-order lines. The solutions are tested and ranked as described above.

The short-axis tests and the 020-finding algorithm are useful only for a limited number of problems. Fortunately, however, they have been found to solve many monoclinic

0021-8898/85/050367-04$01.50

Table 1. *Normal basis-line sets and hkl restrictions used by TREOR*

Symmetry	Basis-line sets (Line numbers)	Max. *hkl*	Max. $\|h\|+\|k\|+\|l\|$
Cubic	[(1) (2)]	444	6
Tetragonal	[(1, 2) (1, 3) (2, 3)]	444	4
and		444	4
hexagonal			
Orthorhombic	(1, 2, 3) (1, 2, 4)	222	3
	(1, 2, 5) (1, 3, 4)	222	4
	(2, 3, 4) (1, 2, 6)	222	4
Monoclinic	(1, 2, 3, 4)	222	2
	(1, 2, 3, 5)	222	3
	(1, 2, 4, 5)	222	3
		222	4
Triclinic	(1, 2, 3, 4, 5, 6)	111	1*
	(1, 2, 3, 4, 5, 7)	111	2
	(1, 2, 3, 4, 5, 8)	111	2
	(1, 2, 3, 4, 6, 7)	111	2
	(1, 2, 3, 4, 6, 8)	111	3
	(1, 2, 3, 5, 6, 8)	111	3
	(1, 2, 3, 5, 6, 7)		
	(1, 2, 3, 5, 7, 8)		
	(1, 2, 3, 4, 5, 9)		

*The first triclinic line is always set to be 100 and the second line is only given positive indices.

indexing problems, for which the more time-consuming procedure – to include more basis-line sets than those listed in Table 1 – has been less fruitful.

Thus more than 90% of indexed monoclinic patterns have been indexed without increasing the number of basis-line sets. From this it follows also that the probability is high for the presence of a truly triclinic phase, if monoclinic and higher-symmetry tests fail.

Control parameters

Although the general principles used for trial-and-error indexing are relatively simple and straight-forward, the success of the method is a function of data quality and crystallographic decisions put into the program. An essential part of the *TREOR* program is therefore a standard set of parameter values. They are termed normal values and represent accumulated experience from several hundreds of indexing problems. The parameters are referred to by key-words, and they may easily be changed by the user. In order to make the program user-oriented, key-words may be used in a very free way. An arbitrary number of key-words in arbitrary order may be used, and the data are given in free format.

The normal choice of basis-line sets and their index limits are listed in Table 1. A useful basis-line set must be a function of all cell parameters and must not contain an impurity line. Therefore several sets are always used.

Although the index limits rarely have to be increased, it is sometimes necessary to use more monoclinic basis-line sets than those listed in Table 1. In order to avoid unduly time-consuming tests, however, the basis-line restrictions imposed on the triclinic symmetry cannot be changed by any key-words. Usually only the first 19 lines are used in the trial phase of the calculations. Furthermore, parameters obtained from the first seven lines are automatically adjusted by higher-order lines if such lines are available in the data set.

Table 2. *CPU times used on a VAX 11/750 to index the powder diffraction patterns in NBS Monograph No. 25, section 17 (Morris et al., 1980)*

Symmetry	Min. (s)	Max. (s)	Number of indexed patterns
Cubic	1	2	2
Tetragonal	3	91	5
Hexagonal	18	73	4
Orthorhombic	6	107	19
Monoclinic	16	216	15
Triclinic	2	916	5
			$\sum$ 50

Maximum values for the unit-cell volume and the cell axes should be given by the user. The parameters are generally not crucial for orthorhombic or higher symmetries where they may be given some standard value: $V_{max} = 2000$ Å^3 and cell-axis maximum = 25 Å. For the monoclinic symmetry, however, it has been found that a first trial should be made with $V_{max} = 1000$ Å^3, after which the limit may be increased in steps of 500 Å^3. The reason for this procedure is that a correct unit cell may index fewer lines than false cells of too large dimensions. Unless the correct cell is ranked as one of the five best obtained from a monoclinic basis-line set it will not be saved for eventual least-squares refinement.

Approximations based on the symmetry and the *d* spacing of the 20th line have been proposed for unit-cell volumes (Kohlbeck & Hörl, 1976; Shirley, 1980). Unfortunately, they are usually unreliable for monoclinic and higher symmetries. As long as the space group is unknown, systematic absences cannot be taken into account. The triclinic space groups, however, do not exhibit any systematic absences. Thus the approximation suggested by Smith (1977), $V = 13{\cdot}39\, {d_{20}}^3$, is usually rather close to the correct cell volume and may be used with an addition of a few hundred Å^3 as V_{max}.

The computing time is strongly dependent on the V_{max} parameter. It is therefore important not to allow unrealistically large cell volumes in triclinic calculations, which can be very time consuming. This is another reason for treating the triclinic case separately.

An important part of the program is an extensive comment list, containing not only default values of the parameters and test examples but also recommended strategies for using the program, and a check list for interpretation of the results.

Computing methods

The reciprocal-cell relationship may be written

$$Q(hkl) = h^2x_1 + k^2x_2 + l^2x_3 + hkx_4 + hlx_5 + klx_6.$$

Each trial set of x_i is derived from a system of linear equations $\mathbf{MX} = \mathbf{L}$, where $\mathbf{M}$ is a square matrix containing the Miller indices, $\mathbf{X}$ is the unknown parameter set x_i and $\mathbf{L}$ is the set of observed Q values for the basis lines used. The dimensions of the $\mathbf{M}$ matrix and the $\mathbf{X}$ and $\mathbf{L}$ vectors depend on the symmetry. For monoclinic and higher symmetries the parameters x_i are found by Cramer's rule. In order to minimize the computing time, sub-matrices are saved and

used until elements within them are changed. The procedure is repeated for each new basis-line set. For the triclinic symmetry the inverse matrix $\mathbf{M}^{-1}$ is calculated. Then

$$\mathbf{X} = \mathbf{M}^{-1}\mathbf{L}.$$

Thus new parameter vectors **X** can be derived by changing the elements in **L** (*i.e.* using new basis-line sets) and multiplying the **L** vector by the inverse matrix, $\mathbf{M}^{-1}$. This procedure is used because of the large number of basis-line sets needed for the triclinic symmetry (see Table 1). Thus, matrix inversions are economic for this symmetry.

Program tests

In order to make an objective test of the efficiency of the program we decided to use data sets from one of the NBS monographs for X-ray diffraction patterns (Morris, McMurdy, Evans & Paretzkin, 1980). The monograph contains two cubic, five tetragonal, four hexagonal, 19 orthorhombic, 18 monoclinic and six triclinic patterns. According to our experience monoclinic and triclinic patterns cannot usually be indexed reliably on the basis of powder diffraction data alone, if the cell edges are longer than 25 and 20 Å, respectively. Therefore the monoclinic phenylhydrazine hydrochloride, $C_6H_8N_2.HCl$ with $b = 30{\cdot}641$ Å, and the triclinic clopenthioxol hydrate, $C_{22}H_{25}ClN_2OS.2H_2O$ with $b = 21{\cdot}939$ Å, were not indexed. The monoclinic data set from sodium perchlorate hydrate, $NaClO_4.H_2O$, was of low quality, especially the low-order lines, which are the most important lines for powder indexing by *TREOR*. It is stated in the NBS report that the samples were very unstable, changing readily to the anhydrous phase and back again depending on the atmospheric relative humidity. Also the monoclinic chromium chloride, $CrCl_3$, caused some indexing problems. The reported correct unit cell is *A* centred with the dimensions $a = 6{\cdot}123$, $b = 10{\cdot}311$, $c = 5{\cdot}956$ Å, $\beta = 108{\cdot}64$, $V = 356{\cdot}3$ Å^3.

It was found, however, that also other unit cells with no obvious relation to the correct cell gave acceptable de Wolff figures of merit. For example, the monoclinic cell $a = 11{\cdot}852$, $b = 4{\cdot}664$, $c = 7{\cdot}751$ Å, $\beta = 102{\cdot}26$, $V = 418{\cdot}5$ Å^3 and the triclinic cell $a = 6{\cdot}149$, $b = 7{\cdot}583$, $c = 4{\cdot}871$ Å, $\alpha = 90{\cdot}5$, $\beta = 104{\cdot}72$, $\gamma = 102{\cdot}3$, $V = 214{\cdot}2$ Å^3 can be used to index all the 19 available lines with de Wolff figures of merit equal to 13 and 17 respectively. The small monoclinic cell and the fact that the pattern has no dominant zone make the problem unusual.

All the remaining 50 patterns, 93%, were correctly indexed by the program without using any information about density and formula weights.

Only the first 25 lines were used in the *TREOR* runs, but second-order lines available among the first seven lines, used to define basis-line sets, were omitted. The first-order lines were always corrected by the second-order lines, however. Only basis-line sets as defined by the default values were used.

Patterns indexed by unreduced cells, or containing common factors in the quadratic forms, are here regarded as correct. We have written a separate program for unit-cell reduction and conversion to conventional cells according to the metric symmetry. Thereby all unit cells given in the NBS monograph were easily found from the patterns indexed by *TREOR*. For example, the monoclinic mercury acetate pattern was correctly indexed by the triclinic algorithm within four seconds and was subsequently reduced to the monoclinic cell.

It is true that input of density and formula weight will usually considerably reduce the computing time and make the program more powerful. For example, the incorrect $CrCl_3$ cells mentioned above may be excluded on such grounds. It is our experience, however, that indexing problems usually arise before any accurate knowledge about composition and density is available.

In Table 2 the ranges of computing time on a VAX 11/750 used to index the 50 patterns are given for all symmetries. As can be seen from the table, computing times above 4 min were only necessary for triclinic patterns. Computing times for the monoclinic runs include short-axis tests, but orthorhombic and higher-symmetry tests for these patterns are excluded because they should always be tried in separate runs before any monoclinic test is made.

Patterns for which a short-axis test is necessary to solve the indexing problem are not available in the NBS monograph used (Morris *et al.*, 1980). This part of the program had to be tested on other patterns. For example, in the NBS monograph No. 25, section 16 (Morris, McMurdy, Evans, Paretzkin & Groot, 1979), four monoclinic patterns with a common zero index for the first six lines were successfully indexed by the short-axis test.

The wide range of computing times for the triclinic symmetry (see Table 2) depends on the strategy used by the program to refine the saved unit cells if all the first 19 lines are indexed by a trial cell. The number of triclinic samples tested so far is limited, and although the five patterns included in Table 2 were all successfully indexed, it is expected that the efficiency of the triclinic routine may be increased by further development of the program. The indexing strategy for the triclinic symmetry is based on experience from true triclinic patterns and should not be used unless other symmetry tests have failed.

For comparison a Guinier–Hägg photograph of ammonium persulfate, $(NH_4)_2S_2O_8$, was taken with addition of silicon as internal standard. The exposure time was 30 min, and within 15 min from the moment the photograph was dry it was measured and correctly trial-and-error indexed by the SORD M68 microprocessor used for the Guinier film-scanner system (Johansson, Palm & Werner, 1980). The standard deviations in all parameters and the average difference between observed and calculated diffraction angles were also smaller than those published in the NBS monograph. This is in agreement with the general experience that focusing-camera data are preferable for powder indexing (Shirley, 1980; Smith & Kahara, 1975).

Final remarks

The main parts of the program are written in Fortran IV, but some routines, for example the short-axis tests, are written in Fortran 77. These routines can easily be converted to Fortran IV. Full loading of the program on a SORD M68 microprocessor requires a core memory of 60 000 words, *i.e.* 120 000 bytes. Segmentation of the program is possible. The program is available from this Institute.

The authors wish to thank Professor P. Kierkegaard for his kind interest in this work.

This work has received financial support from the Swedish Natural Science Research Council.

References

JOHANSSON, K. E., PALM, T. & WERNER, P.-E. (1980). *J. Phys. E*, **13**, 1289–1291.

KOHLBECK, F. & HÖRL, E. M. (1976). *J. Appl. Cryst.* **9**, 28–33.

MIGHELL, A. D. & SANTORO, A. (1975). *J. Appl. Cryst.* **8**, 372–374.

MORRIS, M. C., MCMURDY, H. F., EVANS, E. H. & PARETZKIN, B. (1980). *Natl Bur. Stand. (US) Monogr.* No. 25, section 17.

MORRIS, M. C., MCMURDY, H. F., EVANS, E. H., PARETZKIN, B. & GROOT, J. H. DE (1979). *Natl Bur. Stand. (US) Monogr.* No. 25, section 16.

SHIRLEY, R. (1980). *Data Accuracy For Powder Indexing. Natl Bur. Stand. (US) Spec. Publ.* No. 567, pp. 361–382.

SMITH, G. (1977). *J. Appl. Cryst.* **10**, 252–255.

SMITH, G. & KAHARA, E. (1975). *J. Appl. Cryst.* **8**, 681–683.

TAUPIN, D. (1973). *J. Appl. Cryst.* **6**, 380–385.

WERNER, P.-E. (1964). *Z. Kristallogr.* **120**, 375–387.

WOLFF, P. M. DE (1968). *J. Appl. Cryst.* **1**, 108–133.

Computer Program Abstracts

J. Appl. Cryst. (1985). **18**, 370

The category Computer Program Abstracts provides a rapid means of communicating up-to-date information concerning both new programs or systems and significant updates to existing ones. Following normal submission, a Computer Program Abstract will be reviewed by one or two members of the IUCr Commission on Crystallographic Computing. It should not exceed 500 words in length and should use the standard format given on page 189 of the June 1985 issue of the Journal [*J. Appl. Cryst.* (1985). **18**, 189–190].

NEWMAN, program for calculating and plotting Newman projections from atomic coordinates and cell constants.

By H. SCHENK, N. P. BRANDENBURG, B. VAN SANTEN, E. Y. KRAGTEN and B. O. LOOPSTRA, *Laboratory for Crystallography, Nieuwe Achtergracht* 166, 1018 *WS Amsterdam, The Netherlands*

(*Received* 7 *January* 1985; *accepted* 11 *March* 1985)

The crystallographic problem: Crystallographers publish cell constants and coordinates, and in most cases the redundant but more informative bond lengths and angles as well. Coordinates contain all of the structural information, but bond lengths and angles contain no stereochemical information. This information might be presented by listing the torsion angles, but these are not always easy to interpret. Newman projections are stereospecific graphical representations of the torsion angles around a particular bond. These projections are much easier to interpret than angle data. A program to give Newman projections was previously written by Brandenburg (1974).

Method of solution: The atom and cell information is either entered (microcomputer versions) or read from a binary data file (mainframe version). Orthogonal coordinates are calculated, and from the distances a bonding matrix is evaluated within predefined limits on the bond lengths. Then, for a particular bond, all connected bonds are projected in a plane perpendicular to the central bond. All angles between successive projected connected bonds are calculated and printer plotted (batch mode main frames) or shown on a video display (micros such as TRS80 and IBM-PC). Optionally, the Newman projections can be plotted.

Software environment: The main-frame version is written in Fortran IV for CDC Cyber computers under NOSBE control, but is not machine dependent. It is linked without overlay. The plot routines are local. However, since only circles, straight lines and some text are plotted, local adaptations should not be difficult. Microcomputer versions are written in Basic for the TRS80 model I and IBM-PC. Since plot routines are simple, the program can be easily transferred to other micros (for example the CBM64). The Basic microcomputer versions run under NEWDOS (TRS-80 model I) and IBM-DOS 2·0, respectively.

Hardware environment: The main-frame version runs within 32 k 60-bit words, and optionally uses a plotter. The micro versions require a disk drive and a matrix printer. For the IBM-PC a color/graphics card and a printer with dot-graphics facilities are required.

Program specifications: The program does not have any serious restrictions. Run times are a few seconds CPU (main frame). The microcomputer version is interactive. The number of lines of code is approximately 1000 (Fortran) and 500 (Basic). There is no documentation in machine-readable form. Numerous tests have been run for all programs.

Documentation: A manual containing a short description of the program and input instructions is available.

Availability: The program source and documentation are available from the authors. For the main-frame version the cost of magnetic tape, copying, postage, handling and software licence is Dfl 150. For the micro version the cost is Dfl 100 (floppy disk). For developing countries special arrangements may be possible. Users are not entitled to redistribute the program.

Keywords: Newman projection; Stereospecific projection; Plotting routines.

Reference

Brandenburg, N. P. (1974). *J. Appl. Cryst.* **7**, 401.

J. Appl. Cryst. (1985). **18**, 370–371

TYPIST – a program for the tabulation of crystallographic results.

By M. TOMASSINI, *Department of Earth Sciences, Piazza Università, University of Perugia*, 06100 *Perugia, Italy*

(*Received* 2 *January* 1985; *accepted* 22 *March* 1985)

The crystallographic problem: The preparation of tables of relevant structural information is a time-consuming and error-prone task. It is possible to automate the whole process, but standard word processing programs are not well suited to this specialized purpose, as they require input of the information beforehand. We have thus written a set of programs that automatically extracts the relevant items from the output of a cry-

0021-8898/85/050370-02$01.50