Enhancements in Powder Pattern Indexing

By Daniel TAUPIN
Laboratoire de Physique des Solides, associé au C.N.R.S.
Bâtiment 510, Centre Universitaire, 91405 Orsay, France

Abstract

The indexing strategy of our sixteen year old powder pattern indexing programme has been recently revised. Several enhancements lead to an increased reliability and much shorter computing times in indexing difficult patterns, especially monoclinic.

March 27, 1991

I. Introduction.

Half a dozen programmes at least are now available to assign indices to X-ray powder patterns of unknown lattice cell: Werner (1964), Taupin (1968, 1973), Visser (1969), Louër & Louër (1972), Kohlbeck & Hörl (1975), Werner *et al.* (1985), etc.. These programmes and others have been discussed by Shirley (1978) who gives some other references.

Since the problem of retrieving line indices in a powder pattern is not at all trivial, each of these programmes has a distinct strategy. Some of them take account of special relationships between line position (Visser, 1969) and quickly find the solution, when they succeed. Others prefer a systematic strategy, using dichotomy (Louër & Louër, 1972) or testing a tentatively exhaustive set of index combinations (Taupin, 1973).

Although generally successful, our latter programme (Taupin, 1973) appeared to be sometimes defective, namely it sometimes wasted hours of computing time trying hopeless index combinations (e.g. index combinations which would necessarily lead to cell volumes beyond the user's authorized limit).

II. Recalling the bases of our indexing strategy.

Let n be the number of unknown parameters defining the crystal cell (n=1, 2, 2, 3, 4, 6 for cubic, hexagonal, tetragonal, orthorhombic, monoclinic and triclinic systems, respectively); then all possible combinations of n index triplets $\{hk\ell\}$ are tentatively assigned to a given set of n lines of the experimental diagram (usually the n lowest-angle lines) and, for each combination, the following linear system is solved:

$$h_i^2 A + k_i^2 B + \ell_i^2 C + 2h_i k_i D + 2k_i \ell_i E + 2\ell_i h_i F = Q_i \quad ; \quad i = 1, n \tag{1}$$

where

$$Q_i = 1/d_i^2 \tag{2}$$

$$A = \mathbf{a}^{*2} \quad ; \quad B = \mathbf{b}^{*2} \quad ; \quad C = \mathbf{c}^{*2} \tag{3}$$

$$D = \mathbf{a}^* \cdot \mathbf{b}^* \quad ; \quad E = \mathbf{b}^* \cdot \mathbf{c}^* \quad ; \quad F = \mathbf{c}^* \cdot \mathbf{a}^* \tag{4}$$

with additional constraints for all lattices but triclinic, e.g.: $A = B = C$, $D = E = F = 0$ for cubic, $A = B$, $D = E = F = 0$ for tetragonal, etc..[1]

For each set of n triplets $\{hk\ell\}$ there is a solution, i.e. a possible lattice, which obviously indexes the n lines (the *base* lines) taken, but seldom all the

[1] Although the common use is now to call them β and b, we prefer to denote as γ and c the particular angle and the particular axis in the monoclinic lattices.

other lines of the given diagram. If it does, one has found one of the possible solutions of the problem.

If the problem were handled as crudely as described above, one would have to perform a dramatically huge number of trials, and even now – in the last eighties – this would lead to unreasonable computing time spent. It is therefore necessary to introduce constraints and checks which short cut useless, impossible or redundant combinations as soon as possible.

III. The combinatorial strategy and the constraints.

Since the optimization of higher symmetry (i.e. cubic, hexagonal and tetragonal) lattice seeking is not crucial owing to the small amount of trials to be performed, while these lattices require special handling of the indices h, k and ℓ, we shall only discuss below the stategy implemented in the case of orthorhombic, monoclinic and triclinic lattices.

III.1 Varying h, k, ℓ for the base lines.

Since it is definitely impossible to make the indices $\{hk\ell\}$ run from $-\infty$ to $+\infty$, we have been compelled to assign limits to the $\{hk\ell\}$ of the base lines. Rather than assigning separate limits to h, k and ℓ it is obviously wiser to limit the value of $h^2+k^2+\ell^2$ and to make this arbitrary limit proportional to the Q of the *base* lines. In practice, h, k and ℓ do not vary separately for each base line, but we prefer to use a pre-computed table of triplets $hk\ell$ *ordered against* $h^2+k^2+\ell^2$, that is, ordered against decreasing likelihood. Thus, simple lattices with low indices are sought first, and unlikely ones with large cell volumes are examined later, if needed. However, this way of limiting the indices which we have used for years, exhibits some drawbacks when several crystal cell symmetries are tested; in fact, the number of lines satisfying a condition of the type

$$h^2+k^2+\ell^2 \leq K \tag{5}$$

in a triclinic lattice is twice the number in a monoclinic, which in turn is twice the number in an orthorhombic, and so on. Conversely, there is no obvious reason why the probability that the low index lines have intensities below the experimental threshold should be different in any of these systems. Therefore, we think it is wiser not to state a limit for the actual value of $h^2+k^2+\ell^2$ of the base lines, but to state – for example – that the first line should be labelled with a triplet $hk\ell$ choosen within the N first ones (ordered by increasing $h^2+k^2+\ell^2$).

This latter number has a rather obvious probabilistic meaning: for a cell whose parameters are of the same order of magnitude, assigning the label of the first line to be within the N first ones in the table means that we expect that the structural extinction rules or damping effects will not remove all the N first possible labels from the experimental data. Thus a value of N ranging from 15 to 25 is quite reasonable, especially when I, F, A, B and C-centered lattices are handled seperately – i.e. with separate $hk\ell$ tables – as we do now. Concerning the limit of the next base lines, it must be pointed out that, while the $h^2+k^2+\ell^2$ have to be extrapolated proportionally to Q, the value of N (which is akin to a number of reciprocal unit cells in a given ellipsoid) should be extrapolated proportionally to $Q^{3/2}$, in other words, to $1/d^3$.

III.2 The constraints on the reciprocal cell parameters.

In addition to the normal constraints resulting from the symmetry of the cell, except for triclinic lattices, the generality of the programme is not restricted by imposing the following constraints:

$$A \geq B \geq C > 0 \quad ; \quad (orthorhombic) \tag{6a}$$

$$A \geq B \geq 2D \geq 0 \quad ; \quad (monoclinic) \tag{6b}$$

$$A \geq B \geq C > 2E \geq 0 \quad ; \quad B \geq 2D \geq 0 \quad ; \quad C \geq F \geq 0 \quad ; \quad (triclinic) \tag{6c}$$

The positivity of all the unknowns A, B, C, D, E and F leads to a first simple elimination rule: if a line Q_i is tentatively indexed by $h_ik_i\ell_i$ and a line Q_j by $h_jk_j\ell_j$ and if $Q_j > Q_i$, then it is impossible to have:

$$\begin{gathered} h_i^2 \geq h_j^2 \text{ and } k_i^2 \geq k_j^2 \text{ and } \ell_i^2 \geq \ell_j^2 \text{ and} \\ h_ik_i \geq h_jk_j \text{ and } k_i\ell_i \geq k_j\ell_j \text{ and } \ell_ih_i \geq \ell_jh_j \quad . \end{gathered} \tag{7}$$

Therefore, any triplet which violates this condition with respect to the indices assigned to lines of smaller Q must be immediately rejected.

Besides – and this is an important novelty – one does not need to have assigned tentative index triplets to all the first n experimental lines before beginning to solve the system (1). Rather, it is advisable to begin eliminating unknowns between tentative equations, as soon as two lines have been assigned index triplets. Then, due to the high frequency of zeroes in the indices of the low-angle lines, the value of one or several unknowns if often available once only 2 or 3 lines have been tentatively labelled. If this results – for example – in B being greater than A in a monoclinic system, then all trials using these tentative labels have to be rejected (owing to redundancy elimination constraints (6)) and thousands of useless trials can thus be short cut.

Another constraint is of major importance, namely the maximum volume of the unit cell, i.e. the minimum volume of the reciprocal unit cell. A first value of this limit is normally given by the programme user, but it can often be dynamically restricted by information considerations (see Taupin, 1988), especially when at least one good lattice has been found. Since the volume of the reciprocal unit cell is readily deduced from the parameters A to F by

$$V^{*2} = ABC + 2DEF - CD^2 - AE^2 - BF^2 \tag{8}$$

it is sufficient to locate A to F within intervals to get an estimation of the interval within which V^* lies, and reciprocally when taking account of the conditions (6).

Additional efficiency can be also attained by introducing reasonable constraints on the values of the cell parameters a, b and c. They can obviously be stated by the programme user (practically, we ask for constraints on $1/a^*$, $1/b^*$ and $1/c^*$) but, even if he does not want to give them, these upper and lower bounds may be derived from the given set of experimental lines with a good reliability. Practically, one can take account of the fact that indices greater than 2 are very unlikely for the first *recorded* line in orthorhombic, monoclinic and triclinic[2] lattices, and set the lower bound of A, B, C to approximately Q_1 divided by the maximum value allowed for $h_1^2 + k_1^2 + \ell_1^2$, which can be deduced from N, unless explicitely specified.

Conversely, extra-flat tile-shaped cells are also very unfrequent. This can be interpreted in setting the upper bounds of A, B, C to some value between Q_{10} and Q_{20}.

Thus, most of the strategic combinatorial work has to deal with inequalities.

III.3 Tactical considerations.

Each time a tentative triplet is assigned to a new experimental line (the i-th line), the following operations are made:

a) if $i > 1$ the tentative triplet is compared to those previously assigned to the lower angle lines, to see whether it should be rejected according to (7).

b) if this triplet is decided to be eligible at level i (namely with $i-1$ previous triplets already stored), a *formal* equation is built and stored in memory in the form:

$$\alpha_i A + \beta_i B + \gamma_i C ... = p_{ii} Q_i \tag{9}$$

where only α_1, β_1, γ_i and $p_{ii} = 1$ are numerically computed.

[2]The lowest line of a triclinic lattice can always be assessed to be 0 0 1, but not necessarily the first recorded one: if 0 0 1 is missing (too weak or out of range for the detector) and one assesses the label 0 0 1 to the true 0 0 2 line, then it will be impossible to index the 0 0 3 line if it is present.

c) the $i-1$ previously stored equations are *formally* substituted in the new i-th equation (9) which therefore takes a more general form:

$$\alpha_i A + \beta_i B + \gamma_i C ... = p_{i1} Q_1 + p_{i2} Q_2 ... + p_{ii} Q_i \tag{10}$$

thus, the system of equations stored in memory is independent of the choice and of the exact values of the lines Q_1 to Q_i, and the same system can be used to estimate the derivatives of A, B, C, etc., against the Q-values, i.e. to perform error computation.

d) if α_i, β_i, γ_i, ... are all equal to zero – this means that the selected triplet was not independent of the previous ones – the minimum and maximum value of the right hand member of this linear condition are computed, according to the allowed discrepancies stated by the user for these lines, namely ϵ_1, ϵ_2, ... ϵ_i. If these values do not enclose zero, the triplet is rejected and another one is tentatively assigned at this level i. If they do enclose zero, the triplet is considered as eligible for this i-th experimental line, but the equation is dropped since it was not independent of the previous ones; thus other triplets have to be assigned at a higher level to the next higher angle line (note that all this process is recursive). This test is in fact of major importance: if it was not correctly done, the programme would loop indefinitely (or crash) in falling into the trap consisting in assigning – for example – all the $\{hk0\}$ triplets to all the given lines of an orthorhombic material, while the 3rd or the 4th one would have been rejected by this test.

e) if at least one of the α_i, β_i, γ_i, ... is not zero, then the maximum and the minimum possible values of the left hand member of (10) are computed, using the present state of knowledge of the possible ranges of A, B, ..., F and they are compared to the maximum and minimum possible values of the second hand member computed as above. In case of incompatibility, another triplet is tentatively assigned to the same experimental line and the same checking process is performed. In case of compatibility, the resulting equation is kept but further computations are done to try to restrict the intervals where the unknowns are supposed to be located.

f) if, at these stage, the number of equations kept is less than the required number n, then using maximum or minimum of the other unknowns depending upon the sign of α_i, β_i, γ_i, ..., the formal relationskip (10) is used to cyclically update (i.e. restrict) the interval of each unknown. If some update happens, the new intervals are matched to equation (8) and possibly updated once more. Of course, updating intervals may result in contradictions which in turn lead to the rejection of the tentative triplet as above.

g) if the tentative triplet appears to be eligible, but the number of equations kept is less than the required number n, then the whole process is recursively started (step a)), one level higher ($i \to i+1$), in order to assign tentative triplets to the next higher experimental line. The number of levels of this recursive process is, however, limited in order to avoid some pathological cases where – for example – all the experimental lines could be indexed by $hk0$ labels yielded by some high volume cell. In practice, we limit the number of *dropped equations*, also called the *degenerescence index* to a certain number n_{deg}, so that the total number of base lines considered simultaneously never exceeds $n + n_{deg}$. Thus, if the tentative triplet would lead the number of base lines to exceed this limit, this triplet is merely rejected and a new one is tried at the same level (step a)).

h) conversely, if the number of independent equations is equal to n, then a complete indexing work is done, which may or may not succeed in a satisfactory lattice. By "satisfactory" we mean two conditions:

— all given lines have been indexed within their stated accuracy, unless the user authorizes the programme to ignore them by means of a special code in the data;

— the computed *information merit* (see Taupin, 1988) is higher than the threshold choosen by the user and indicated in the input data.

i) if this succeeds, then the minimum *information merit* is updated to its best overall value *minus 10* (this means that any further lattice found whose probability is less than $2^{-10} \approx 0.001$ times the probability of the best one the best one will be rejected). In turn, this often leads to a restriction of the maximum direct cell volume, i.e. an increase of the minimum reciprocal cell volume with all its consequences in shrinking the allowed intervals for the unknowns.

IV. Checking and application.

The new version of the programme has been tested on a set of 112 randomly generated synthetic diagrams, with artificial erratic deviations supposed to simulate experimental imperfections. These diagrams consisted in 16 specimens of cubic, hexagonal and tetragonal lattices, and of 32 orthorhombic and monoclinic lattices.

All the 48 higher symmetry lattices but 3 were indexed at the first trial with a parameter $N = 15$, which corresponds for instance to a first line having indices not higher than 3 0 0 in the P-tetragonal. Two of the generated patterns exhibited abnormally high indices for the first line, and they needed an increase of N (up to 40) to be correctly indexed[3]. For another hexagonal lattice, indexing happened to be possible only by increasing the allowed discrepancies by 40 %[4].

All the 32 orthorhombic lattices but 5 were indexed at the first trial with $N = 15$, which corresponds to the assumption the the indices of the first line are not higher than 2 0 1. Two of them were rather ill-conditioned (bad precision) and spent too much computing time, so that we had to restrict the range of the volume and parameters to achieve the indexing in reasonable times. Two others exhibited first lines with very high indices, so that the parameter N had to be increased. Finally, one orthorhombic lattice needed an increase of the allowed discrepancies to be correctly indexed.

Of the 32 monoclinic patterns, 21 were indexed although most of them required the maximum volume to be tightly restricted within order to make the programme running within reasonable c.p.u. times. At present time (our tests ran almost continuously for 6 months) 5 synthetical patterns are not yet finished. The other 6 patterns definitely failed to be indexed.

The reasons of these failures have been investigated: they are mainly due to excessive imprecision of the generated first lines; this appears to be much more harmful in the monoclinic case than in the higher symmetries. In fact, determining the cell of an orthorhombic lattice requires finding three independent lines and especially at least one with $h \neq 0$, another with $k \neq 0$ and another with $\ell \neq 0$; such a set of lines is usually found in the 4 or 5 first lines. If many lower labels have zero values, the determination of the orthorhombic cell becomes even easier. Conversely, the determination of a monoclinic cell necessarily requires at least one line with $h \neq 0$ **and** $k \neq 0$; this condition often requires to look at upper lines due to the large number of zeroes in the first ones. Even worse, the determination of $\cos\gamma$ always results from differences between lines, so that it is often imprecise.

For monoclinic cells, the computing times ranged from approximately one hour to 245 hours on a HP1000/A900 computer (comparable to a Microvax II or a VAX 750), with a typical c.p.u. time of 10 to 20 hours.

The programme also works and gives sensible results when trying to index patterns with triclinic lattices, but we have not yet done systematic trials to

[3] Whether an hexagonal pattern starting with the 2 0 2 and 1 1 3 lines is a realistic situation, is another question...

[4] One of the new features of our programme allows the user to increase or reduce the maximm deviations by a common factor, without having to change them individually in the record corresponding to each given line.

assess that it always finds the correct cell.

Besides, the VAX version of the programme was sent to G. Johnson who used it for trial indexing of the materials recorded in the Powder Diffraction File recorded as "unindexed". In fact, according to G. Johnson (1988), 25 % of the unindexed materials were indexed by our programme, namely, all the 20 first lines were indexed within estimated experimental accuracy with an overall *information merit* greater than the standard threshold of 80.

V. Parameters and features.

Although the programme is intended to work with all default values for its parameters, the user may change many of them to fit his special purposes:

- cubic, hexagonal, tetragonal, orthorhombic, monoclinic and triclinic lattices may be separately selected;
- the user can either search F, I, A, B, C together with P-lattices or only the P-lattices;
- input data may be d, $1/d$, $1/d^2$, θ, 2θ or 4θ;
- each given line may be given its own precision;
- each given line may be allowed to be ignored by the programme instead of rejecting the whole trial;
- the total number of skipped lines can be choosen by the user;
- the minimum *information merit* can be optionally changed;
- a global error scaling factor can be introduced to relax or to tighten the estimated precision of all lines;
- a zero angle shift can be introduced to take account of possible systematic errors;
- a general wavelength is given, but individual lines may have been recorded with different radiations;
- the maximum volume of the unit cell can be modified;
- the maximum values of a, b and c can be specified;
- the information about specific weight and molecular mass can be introduced;
- for long trials, the user is provided with a restart facility and output file swapping which allows him to look at previous results without stopping the process.

VI. Availability.

The indexing programme is written in standard Fortran 77, with a few assembler or specific supervisor calls for timing controls. It runs presently on VAX/VMS, HP1000/RTE-A, IBM/MVS. If needed, specific versions can be derived for UNISYS/1100, IBM/VM/CMS and PC.

VII. References.

Kohlbeck, F. & Hörl, E.M. (1976). *J. Appl. Cryst.*, **9**, 28-33.

Louër, D. & Louër, M. (1972). *J. Appl. Cryst.*, **5**, 271-275.

Shirley, R. (1978). *Indexing Powder Diagrams* in *Proc. of 1978 Summer School*, Delft Univ. Press and Oosthoeks, 1978.

Taupin, D. (1968). *J. Appl. Cryst.*, **1**, 178-181.

Taupin, D. (1973). *J. Appl. Cryst.*, **6**, 380-385.

Taupin, D. (1988). *J. Appl. Cryst.*, **21**, *in the press.*

Visser, J.W. (1969). *J. Appl. Cryst.*, **2**, 89-95.

Werner, P.E. (1964). *Z. Kryst.*, **120**, 375-387.

Werner, P.E., Eriksson, L. & Westdahl, M. (1985). *J. Appl. Cryst.*, **18**, 367-370.

Appendix[5]
NOT to be published

Practical use of the Powder Diagram Indexing Programme

A.1 Invoking the programme.

Typical invocation of the programme is of the form:

`PWDCDS` <*input file*>`,`<*output file*>`,`<*breakpoint file*>`,`<*log file*>
(VMS, RTE-A).
`// EXEC PGM=PWDCDS,PARM='`<*input file*>`,`<*output file*>`,`<*breakpoint file*>`,`<*log file*>`'`
(IBM-MVS).

<*input file*> is the name of the file containing the input data (this is a DDNAME in the case of IBM-MVS). If omitted, the user either prompted to give it, or standard input is used, depending of the implementation. <*output file*> contains the printout after execution (standard output is used when omitted). On VAX-VMS, RTE-A, and other executive systems which allow dynamical allocation of new files (i.e. not on IBM-MVS) <*output file*> is always a *new* file: if the name already exists, a new version or a new derivated file name is created.

If specified, <*breakpoint file*> is used by the programme to store restart informations which enable it to resume its operations at the point where it has been stopped because of exceeded c.p.u. time conditions (see parameters below). If it does not exist, it is created (with restrictions for IBM-MVS); if it exists, it is re-used for input and for output. Therefore, it must be deleted or erased when performing a new trial, i.e. performing a trial which is not the continuation of a previous one.

If specified, <*log file*> contains a very short summary of the execution, namely the title of each indexed lattice, its cell parameters and its information merit, and the c.p.u. time already consumed for that pattern. If it does not exist, it is created (with restrictions for IBM-MVS); if it exists, it is re-used, but the new information is *appended* to the existing file which, therefore, contains the summary of all trials referring the the same <*log file*>.

[5]This appendix is only given here for the referrees' need of information.

A.2 Input data (the <*input file*> structure).

In order to avoid problems with editors (e.g. TSO) which use columns 73:80 for record numbering, significant data are supposed to be stored only in columns 1:72 of the records of the <*input file*>.

A.2.1 First record

Columns 1:12 contain a set of *option characters* which specify which operations are to be done, and which kind of data are input below.

`'A'` means that input data will be 2θ angles, rather than d values.

`'A1'` means that input data will be θ angles.

`'A4'` means that input data will be 4θ angles.

`'C'` means that cubic lattices have to be sought.

`'H'` means that hexagonal lattices have to be sought.

`'L'` disables seeking of S, I, F lattices. If specified, only P-lattices are looked for.

`'M'` means that monoclinic lattices have to be sought.

`'O'` (letter O, not zero) means that orthorhombic lattices have to be sought.

`'P'` causes special "punch" output to happen on Fortran unit 7, which has to be specially allocated before execution (this is a rather obsolete feature).

`'Q'` means that input data will be $1/d^2 = Q$ rather than d.

`'S'` suppresses the normal redundancy test which eliminates lattices which index lines whose angles are lower than some base lines.

`'T'` means that tetragonal lattices have to be sought.

`'U'` means that input data will be $1/d = q$ rather than d.

`'V'` enables a check that the cell volume is approximately a multiple of the volume of the elementary molecule. This option implies that the user gives the density of the sample and its molecular weight.

`'X'` suppresses the updating of the minimum information merit to the best one minus 10.

`'Z'` produce an extensive printout. To be used for maintenance or checking only.

`'3'` means that triclinic lattices have to be sought.

Remark: if none of the options `C`, `H`, `M`, `O`, `T` and `3` are specified, then all lattice symmetries are tried.

Columns 13:72 contain a title which will be reproduced as a heading at the beginning of each apparently successful indexing, and in the *<log file>*. It is usual to indicate the name of the sample in this field.

A.2.2 Second record

While the first record mainly contained *boolean* parameters, the second one contains numerical parameters. These parameters are in free format, separated by commas, and blanks are meaningless. The user can choose to give these parameters as *positional parameters* in which case he is welcome to make no errors in counting the commas when specifying the 16-th parameter (for example). He can also use *keywords* like `WLTH=1.54018,` to specify the wavelength.

Keyword specified parameters have precedence against positional ones.

These parameters are specified as follows:

No.	*keyword*	*Description.*
1	`NLIN=`	Number of experimental lines given. May be omitted, then the given linesare counted.
2	`NEXT=`	Number of lines the programme is allowed to skip if it cannot index them within the stated maximum deviations.
3	`NDEG=`	Degenerescence parameter: this is the maximum number n_{deg} of *base lines* which are allowed to be dependent of the previous ones. For instance, if this parameter is equal to 5, this means that the programme will not try to introduce the 9-th line as a base line when the 8 first ones (8=3+5) are not sufficient to determine a trial cell in an orthorhombic lattice. The default value is 2; the maximum value is 20-n, but it is not wise to use values greater than 5.
4	`INDX=`	If not negative, this parameter is the maximum value of $h^2 + k^2 + \ell^2$ for the first line. If negative, its absolute value gives the number N of missing lines before the first experimental one, when line triplets are ordered against $h^2 + k^2 + \ell^2$. Default value is -20, i.e. $N = 20$.
5	`CPUX=`	Maximum c.p.u. time, in minutes. If positive, the programme will stop when this limit is reached. Restart informations will be stored into *<breakpoint file>* if this file name is present in the command. If negative, the programme will continue after

this limit is reached, but the <*output file*> will be closed, and the output will continue in another file whose name is derived from the previous one. Thus, the user will be able to look at the recently closed files in order to see whether something is pgoing wrong and decide whether the programme should continue running.

6 `VMAX=` Maximum volume of the unit cell. Default value is 1000 Å^3.

7 `INFM=` Minimum information merit. Only the cattices whose information merit is greater than this parameter are printed and recorded in the <*log file*>. This minimum is eventually increased to the best merit found, minus 10. This updating is disabled by option `X`. Default value is 80. It is not wise to set it below 40, unless for checking purposes. Usual good lattices have a merit ranging from 60 to 300.

8 `MOLM=` Molecular mass of the sample. Useful only with `V` option.

9 `DENS=` Density of the sample in g/cm^3. Useful only with `V` option.

10 `DERR=` Relative error on that density. Useful only with `V` option.

11 `WLTH=` Wavelength used when recording the pattern. Necessary if input data are angles. Useful to extrapolate deviations when not given in the specific record of any line.

12 `ERSC=` Error scale factor. If not equal to 1, this parameters is used to increase the given maximum deviations of all lines, without having to modify each record. Thus, if an indexation seems to fail and no other obvious reasons appear, it may be wise to re-run the programme with `ERSC=1.5` or more. Conversely, if a great number of bad merit cells appear in the output, it might be wise to ty indexing with `ERSC=0.7` or less. Note that this factor is internally restricted when trying large cell volumes with high minimum merits.

13 `CHIM=` The minimum value of the expected reduced χ^2. Default value is 0.1. It is not wise to increase it; decreasing it could help finding weird cells, but it could cost much c.p.u. time.

14 `AMIN=` The minimum value of $1/a^*$, $1/b^*$ and of $1/c^*$.

15 `AMAX=` The maximum value of $1/a^*$, $1/b^*$ and of $1/c^*$.

16 `BMMX=` The number of "bad merit" cells to be accepted before the programme stops as if the maximum c.p.u. time condition was reached. The default value is 128. This parameter is a security

which is supposed to warn the user that his data, or their quality he states through his allowed deviations, are so bad that a great number of cells fit them, without reliability. The corresponding counter is reset at each partial execution, just like c.p.u. consumption registers.

17 `AOFF=` An offset to be added to the angular values of all the lines. The purpose of this feature is to permit shifting all the input angles by a constant offset in order to account for a possible offset of the zero angle of the diffractometer. This feature is effective only if the `A` option has been specified.

Remark: using keywords instead of positional values usually leads to a second record exceeding the 72 column limit. This is circumvented by terminating this record by a `'-'`; this last character is then dropped and the next record is considered as a continuation of the previous one. The total limit is 144 characters.

A.2.3 Experimental line records.

There is one record per experimental line. Each record is in free format, with positional values separated by commas. The values are as follows:

1) The line position, namely d. If options `A`, `A1`, `A2` are present, this position is the corresponding angle, if option `Q` is present, the value given is $1/d^2$, if option `U` is present the value given is $1/d$.

2) The maximum deviation of this line, expressed in the same range as the first value.

4,5,6) The values of h, k, ℓ of that line, if they are known. One should give the complete triplet, or nothing.

7) The particular wavelength with which this line was recorded. The maximum deviation is necessary only for the first line. If further maximum deviations are omitted, they are extrapolated from the most recently stated, assuming that deviations are constant in the θ scale.

The presence of the letter `'P'` or of the question mark `'?'` anywhere in the record means that the programme is allowed to drop this line, provided that the number of dropped lines does not exceed the value of the second parameter of the second record (`NEXT=`).

A.3 Some advices.

The default values are fit for the most common trials. It is wise to change them only in case of some certitude (e.g.: cell volume, maximum and minimum parameters) or when the trial with standard parameters actually failed.

The user should remember that multiplying N (`INDX=`) by 2 results in multiplying the c.p.u. time by approximately 16 in the orthorhombic case, and by approximately 50 in the monoclinic case... The precision (actual and stated) of the input data is also a very critical parameter: a factor of 2 in the precision usually results in factors of 30 to 100 in c.p.u. times.

Therefore, the user is strongly advised to make the first trials with rather restrictive conditions, and to progressively extend the error scale factor (`ERSC=`), the degenerescence parameter (`NDEG=`) and the value of N (`INDX=`) in case of failure.