E X P O

A package for full pattern decomposition and
for solving crystal structures by direct methods

Authors:

A. Altomare, B. Carrozzini, G. Cascarano, C. Giacovazzo, A. Guagliardi, A.G.G. Moliterni, R. Rizzi

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche, CNR,

c/o Dipartimento Geomineralogico, Campus Universitario, Via Orabona 4, 70125 Bari, Italy.

M.C. Burla, G. Polidori

Dipartimento di Scienze della Terra, Universita', 06100 Perugia, Italy.

M. Camalli

Istituto di Strutturistica Chimica 'G. Giacomello', CNR,

Via Salaria Km 29.200 - Monterotondo stazione - 0016 Roma, Italy.

U S E R' S   M A N U A L

Support : sirware@area.ba.cnr.it

Web site : http ://www.ba.cnr.it/IRMEC/SirWare.html

Reference:

Altomare A., Burla M.C., Cascarano G., Giacovazzo C., Guagliardi A., Moliterni A.G.G. & Polidori G.: "*Extra: a Program for extracting Structure-Factor Amplitudes from Powder Diffraction Data*" *(J. Appl. Cryst. 1995, **28**,842-846).*

Altomare A., Cascarano G., Giacovazzo C., Guagliardi A., Burla M.C., Polidori G. & M. Camalli: "*SIRPOW.92 - a program for automatic solution of crystal structures by direct methods optimized for powder data*" *(J. Appl. Cryst. 1994, **27**,435-436).*

1- General Information

The background of the system

EXPO is the integration of the EXTRA package with the SIRPOW package.

EXTRA (Altomare, Burla, Cascarano, Giacovazzo, Guagliardi, Moliterni, & Polidori, 1995) is a program designed to extract integrated intensities from a powder pattern.

The Le Bail algorithm (Le Bail, Duroy & Fourquet, 1988) is combined with least-squares cycles to calculate optimized integrated intensities.

Neutron or synchrotron or home x-ray diffractometer data collected in a step scan mode can be used.

At the end of the run, the program generates a file of reflections containing the extracted intensities which can be used as input file for SIRPOW (Altomare, Cascarano, Giacovazzo, Guagliardi, Burla, Polidori & Camalli, 1993). Reflections are generated taking into account the input cell and the space group.

SIRPOW (Semi-Invariants Representation applied to powder data) has been developed for solving crystal structures from powder data by direct methods.

The representation theory, proposed by Giacovazzo (1977,1980) allowed the derivation of powerful methods for estimating structure invariants (s.i.) and structure seminvariants (s.s.).

The mathematical approach makes full use of the space group symmetry. SIRPOW exploits symmetry in a quite general way allowing the estimation and use of s.i. and s.s. in all the space groups. The range of options available to experienced crystallographers for choosing their own way of solving crystal structures is rather wide. However scientists untrained in direct methods or experienced people trustful in the SIRPOW default mode often can solve crystal structures without personal intervention.

In a default run of EXPO, SIRPOW uses the integrated intensities extracted by EXTRA, the user can skip the extraction routine supplying reflection list by means of an external file.

EXPO works, in a completely automatic way, from the extraction process to the structure determination. At the end of the run a structural model is supplied, which can be analysed and, eventually, improved by means of a useful and user-friendly graphic interface; the user intervention is allowed also in the extraction step. The main feature of EXPO is to produce and exploit several kinds of prior information, potentially useful for improving the pattern decomposition process. As a result a more accurate list of structure factors and a more straightforward application of direct methods procedures are obtained. An initial Rietveld refinement is also available.

EXPO creates the following files (in the work directory):
 - a direct access file to preserve data and results (.bin),
 - a direct access file to preserve the Fourier/Patterson map (.map),
 - an ascii file containing (h k l) and integrated intensities of the generated reflections (.hkl)
- an ascii input file for Sirpow (.sir),

- an ascii file to store peak coordinates for refinement (.xyz),
- an ascii file to store peak coordinates for graphics (.plt).
A .bmp file can be obtained.

2 - Notes on the implementation

All modifications to the defaults of the program have to be applied to the file *specif97*.f*. Search in this file for the string "CHANGE" in block data "BLOCK DATA DEPEND" (Only for Unix and VMS versions).

The default dimensions of the graphic window are stored in the variables *kscrx* and *kscry* (720x525).The user can only increase these values depending on screen resolution.

The default use of the graphic window is determined by the value of the variable *kscreen*.
If *kscreen* is set to zero the window will not be displayed unless the %WINDOW command is used.

Graphic interface in EXPO

This feature is available, at the moment, on UNIX, VMS, Windows'95 and NT (workstation) platforms.

To be able to implement the graphic part the user has to compile the FORTRAN files and one C-file and then link together using the library of X-Window. Examples of Makefile are supplied with program files.

If a graphic device is not available then remove all C files and rename the file xmr97.txt as :

xmr97.f	for Risc/6000, HP
xmr97sgi.f	for Silicon Graphics, SUN, Linux, Digital Alpha (Unix)
xmr97vms.f	for Digital VMS

and then compile and link.

3 - Installing the program.

Please refer to README.TXT file for last modifications.

On Unix platform
- extract files in expo.tar, i.e.:
  - tar xvf expo.tar
  - uncompress *.Z

- copy the files SirWare.* in your HOME directory
- create the executable file using the command
    make -f *file_name*

where *file_name* must be one of the following:

makefile.aix	for Risc/6000
makefile.hp	for HP
makefile.sgi	for Silicon Graphics
makefile.sun	for SUN
makefile.linux	for Linux
makefile.axp	for Digital Alpha (Unix)

- run the test :

    expo cime

- WARNING : the first time the program will take some minutes to create the configuration file SirWare.cfg@Local ; if errors occur or you are not satisfied of the fonts used by the program you have to create the SirWare.cfg@Local :

  - edit the file sirware.cfg.all and extract the part you are interested in ; call it

   SirWare.cfg@Local (upper and lower case are mandatory)

  - copy the file SirWare.cfg@Local in your HOME directory

  - copy the file SirWare.cfg@Local in your HOME directory

  - run the test

    expo cime

On VMS platform

- copy the files SirWare.* in your default directory (sys$login)
- create the executable file using the command

    @build_expo

- edit the file RUN_EXPO.COM in order to customize it (follow hints in the file) ;
- define in your LOGIN.COM file a symbol, i.e. :

    EXPO == "@DKA100 :[USER.EXPO]RUN_EXPO" ;

- run the test :

    expo cime

  - WARNING : the first time the program will take some minutes to create the configuration file SirWare.cfg; if errors occur or you are not satisfied of the fonts used by the program you have to create the SirWare.cfg:

  - edit the file sirware.cfg.all and extract the part you are interested in ; call it

    SirWare.cfg

  - copy the files SirWare.cfg in your default directory (sys$login)

  - run the test :

    expo cime

On Windows '95 and Windows NT (workstation) platform

- copy the file EXPO.ZIP in a new directory (i.e. c :\expo)

- unzip EXPO.ZIP

- install Expo on your DeskTop using the standard link mechanism of Windows '95 or NT

- to run the program click on Expo icon

- Description of files used by EXPO

channel

5	card reader
6	line printer (or video)
30	direct access structure file
31	sequential scratch file
32	sequential scratch file
33	sequential scratch file
34	profile counts file/reflection input file
35	atomic parameters file for refinement
36	atomic parameters file for graphic
41	scattering factors file

If PATTERN/REFLECTIONS FOLLOW directive is used, then the channel associated to counts/reflections input file is set equal to the channel number of the card reader.

4 - Description of the EXPO program

The main modules of the program are :
EXPO, LIST, DATA, EXTRACTION, NORMAL, SEMINVARIANTS, INVARIANTS, PHASE, FOURIER/LEAST-SQUARES, MENU/REFINE, PATTERSON, RIETVELD.

EXPO module
It interprets commands and calls desired routines.

LIST module
It is the software interface between EXPO and the direct access file on which data and results are stored.

DATA module
This routine reads the basic crystallographic information like cell parameters, space group symbol, wavelength, unit cell content and the pattern (or reflections). It includes a modified version of the subroutine SYMM Burzlaff & Hountas (1982). Symmetry operators and information necessary to identify structure seminvariants (estimated in SEMINVARIANTS routine) are directly derived from the space group symbol.

An external file (containing counts or reflections) is supplied by the user.

Diffraction data are checked in order to find out equivalent reflections or systematically absent reflections (which are then excluded from the data set) and, eventually, reflections (weak) not included in the data set (Cascarano, Giacovazzo & Guagliardi, 1991). If the overlapping is supplied by the user using the directive REF 1 and some equivalent or systematically absent reflections with overlapping code different from one are found in the reflection list the program stops. If reflections (weak) are omitted in the data set, indices of these reflections are printed for user usefulness.

EXTRACTION module
In this module diffraction integrated intensities are extracted from the powder diffraction pattern.

To calculate the integrated intensities, the program uses the Le Bail algorithm (Le Bail, Duroy & Fourquet, 1988) and combines it with a least squares procedure minimizing the quantity:

$$\sum w_i \, (y_i - y_{ci})^2 \qquad\qquad (1)$$

The summation in (1) goes over all the data points,

$w_i = 1 / y_i$

y_i = observed intensity at the i-th step

y_{ci} = calculated intensity at the i-th step

$$y_{ci} = s\sum_{\mathbf{h}} LP_{\mathbf{h}}\, m_{\mathbf{h}}\left|F_{\mathbf{h}}\right|^2 P(2\theta_i - 2\theta_{\mathbf{h}})A(2\theta_i - 2\theta_{\mathbf{h}}) + y_{bi}. \tag{2}$$

The summation over **h** includes neighbouring Bragg **h** reflections,

s = scale factor

$LP_{\mathbf{h}}$ = Lorentz - polarization factor

$m_{\mathbf{h}}$ = reflection multiplicity

$F_{\mathbf{h}}$ = structure factor

P = profile function

A = asymmetry function

y_{bi} = background intensity at the i-th step.

The background in (2) is automatically estimated and is described by the polynomial function (Young, 1993)

$$y_{bi} = \sum_{m=0}^{N} bkg_m \left[\frac{2\theta_i}{2\theta_M} - 1\right]^m,$$

N, the polynomial degree, is automatically defined by the program as the number which provides the best fit of the background interpolation points.

$2\theta_M$ is the 2θ associated to the minimum intensity in the range in which the background is estimated.

From the pattern a standard peak is automatically selected. The standard peak fitting is performed to provide the starting values of the variables to refine in the least squares procedure.

The full pattern is automatically divided into intervals. The refined parameters derived by the fitting of each interval will be taken into account for the final full pattern fitting.

The following profile shape functions are available for user choice:

$$\frac{\Gamma(\beta_{\mathbf{h}})}{\Gamma(\beta_{\mathbf{h}} - 0.5)} \cdot \frac{C_{2_{\mathbf{h}}}}{\sqrt{2}} \cdot \frac{2}{H_{\mathbf{h}}}\left(1 + 4C_{3_{\mathbf{h}}} X_{i_{\mathbf{h}}}^2\right)^{-\beta_{\mathbf{h}}} \qquad \text{Pearson VII}$$

$$\frac{C_o^{1/2}}{\sqrt{\pi}H_{\mathbf{h}}} \exp(-C_o X_{i_{\mathbf{h}}}^2)^2 \qquad \text{Gaussian}$$

$$\frac{C_1^{1/2}}{\pi H'_{\mathbf{h}}}(1 + C_1 X_{i_{\mathbf{h}}}^2)^{-1} \qquad \text{Lorentzian}$$

$$\eta_{\mathbf{h}} \frac{C_1^{1/2}}{\pi H_{\mathbf{h}}}(1 + C_1 X_{i_{\mathbf{h}}}^2)^{-1} + (1 - \eta_{\mathbf{h}})\frac{C_o^{1/2}}{\pi^{1/2}H_{\mathbf{h}}} \exp(-C_o X_{i_{\mathbf{h}}}^2) \qquad \text{Pseudo-Voigt}$$

where

$$C_o = 4\ln 2 \qquad C_1 = 4 \qquad C_{2_h} = \sqrt{2^{1/\beta_h} - 1} \qquad C_{3_h} = C_{2_h}^2$$

$$X_{i\,h} = \left(2\theta_i - 2\theta_h\right)$$

$$H_h = \left[w + v \cdot \mathrm{tg}\,\theta_h + u \cdot \mathrm{tg}^2\theta_h\right]^{1/2} \qquad H_h' = Z \cdot \mathrm{tg}\,\theta_h + T/\cos\theta$$

$$\eta_h = \eta_o + \eta_1 \cdot \mathrm{tg}\,\theta_h + \eta_2 \cdot \mathrm{tg}^2\theta_h$$

$$\beta_h = \beta_o + \frac{\beta_1}{2\theta_h} + \frac{\beta_2}{\left(2\theta_h\right)^2} \quad .$$

The default is the Pearson VII function. In this case

$$A\left(2\theta_i - 2\theta_h\right) = \frac{\alpha \cdot \left(2\theta_i - 2\theta_h\right)^3}{\left(\dfrac{H_h^2}{C_{2_h}} + \left(2\theta_i - 2\theta_h\right)^2\right)^{3/2}} + 1 \quad .$$

otherwise

$$A\left(2\theta_i - 2\theta_h\right) = 1 - \alpha \cdot \mathrm{sign}\left(2\theta_i - 2\theta_h\right)\frac{\left(2\theta_i - 2\theta_h\right)^2}{\mathrm{tg}\,\theta_h} \quad .$$

The peak shape, the asymmetry, the cell and the background parameters are all refined by the program.

After a first default run of EXTRA, it is possible to run it again starting from the refined variables in the previous run. In this case the preliminary standard peak procedure and the interval by interval extraction procedure are skipped. The related command is %RECYCLE.

NORMAL module

In this module diffraction intensities are normalized using the Wilson method. Statistical analysis of the intensities is made in order to suggest the presence or absence of the inversion centre and to identify the possible presence and type of pseudotranslational symmetry (Cascarano, Giacovazzo & Luic', 1988 a,b; Fan, Yao & Qian, 1988). Possible deviations (of displacive type) from ideal pseudotranslational symmetry are also detected. All the above information is used as prior information in the integrated intensities extraction process, unless the directive NOPSEUDO is given to the program.

When some additional prior information, besides positivity and atomicity of electron density, is available, then a suitable renormalization of structure factors is made.

EXPO is able to deal with the following types of prior information:

a) pseudotranslational symmetry, identified by normalization routine or other source;

b) a well oriented and well positioned molecular fragment (Camalli, Giacovazzo & Spagna, 1985; Burla, Cascarano, Fares, Giacovazzo, Polidori & Spagna, 1989).

SEMINVARIANTS module

This routine is not used in the default run of EXPO.

One-phase s.s. are estimated by means of their first and second representation as described by Giacovazzo (1978), Cascarano, Giacovazzo, Calabrese, Burla, Nunzi, Polidori & Viterbo (1984). As a default the second representation is calculated.

Two-phase s.s. are estimated by means of their first representation as described by Giacovazzo, Spagna, Vickovic & Viterbo (1979), Burla, Giacovazzo & Polidori (1989).

The estimated s.s. are stored in the direct access file; those evaluated with highest reliability will be actively used in the phasing process while the others will contribute to compute, with other phase relationships, the figure of merit CPHASE.

INVARIANTS module

Up to 10000 triplets relating reflections with normalized E values greater than a given threshold (strong triplets) are stored for active use in the phasing process. Also triplets (PSI-ZERO triplets) relating two reflections with large E and one with E close to zero are generated: they are used to define a special figure of merit (PSCOMB). Special types of triplets (PSI-E triplets) based on two strong and one intermediate reflections (just below the threshold of strong reflections) are calculated and used in the FOURIER/LEAST-SQUARES module in order to extend phase information (Altomare, Cascarano, Giacovazzo & Viterbo, 1991).

Negative quartets are generated by combining the psi-zero triplets in pairs, and those with cross-magnitudes smaller than a given threshold are estimated by means of their first representation, as described by Giacovazzo (1976). These quartets are used to provide an important contribution to the FOM CPHASE.

Active triplets may be estimated according to Cochran's (1955) P_3 distribution: the concentration parameter of the von Mises distribution is then:

$$C = \frac{2 \cdot |\, E_h \cdot E_k \cdot E_{h-k}\, |}{\sqrt{N}} \qquad (3)$$

Triplets can also be estimated according to their second representation (i.e. P_{10} formula, as described by Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi and Polidori, 1984). The concentration parameter of the new von Mises distribution is given by

$$G \ = \ C \,(1 + q) \qquad (4)$$

where q is a function (positive or negative) of all the magnitudes in the second representation of the triplet. The G values are rescaled on the C values and the triplets are ranked in decreasing order of G. The top relationships represent a better selection of triplets with phase value close to zero than that obtained when ranking according to C. These triplets will be actively used in the phase determination process.

Triplets characterized by a negative G represent a sufficiently good selection of relationships close to 180 degrees, to be used both for active use in the phasing process (Giacovazzo, Burla & Cascarano, 1992) and for the calculation of a powerful FOM (CPHASE). Triplets with G close to zero are expected to have values widely dispersed around 90 or 270 degrees and are used to compute an enantiomorph sensitive FOM. A similar FOM is also computed using quartets estimated with a very small concentration parameter.

As a default, triplets are estimated according to P_{10} formula.

The parameter C of the Cochran distribution (say P_3) is suitably modified when prior information, such as that described above in the section 'NORMAL module' is available. Then triplet phases are no longer expected to be around zero (see quoted references) and may lie anywhere between 0 and 2π.

The method of the joint probability distribution functions (Cascarano, Giacovazzo, Guagliardi & Steadman, 1991; Carrozzini, Giacovazzo, Guagliardi, Rizzi, Burla & Polidori, 1997) can be applied to obtain an estimate of normalized structure factors, that will be used as a source of supplementary prior information in the extraction process (if directive ESTIMATION is used).

PHASE module

In the EXPO program the most reliable one-phase s.s. are treated as known phases. Besides triplets, also the most reliable negative quartets and two-phase s.s. may be actively used.

Each relationship is used with its proper weight: the concentration parameter of the first representation for quartets and two-phase s.s., and C or G for triplets.

- CONVERGENCE/DIVERGENCE PROCEDURE -

The convergence procedure (Germain, Main & Woolfson, 1970) is a convenient way of defining an optimum starting set of phases to be expanded by the tangent formula or by any other algorithm.

When the P_{10} formula is used, as a default, a special convergence process is devised which chooses the starting set according to

$$\langle \alpha_{\mathbf{h}} \rangle = \sum_{j} \left[G_j \cdot D_1(G_j) \cdot D_1(\langle \alpha_{\mathbf{k}} \rangle) \cdot D_1(\langle \alpha_{\mathbf{h-k}} \rangle) \right] \qquad (7)$$

as suggested by Giacovazzo (1979) and by Burla, Cascarano, Giacovazzo, Nunzi & Polidori (1987), with

$$D_1(G_j) = \frac{I_1(G_j)}{I_0(G_j)} \qquad (8)$$

I_1 and I_0 represent modified Bessel functions of order one and zero respectively. The summation in (7) is over all relationships defining the reflection $\mathbf{h}$. If P_3 formula is used the default choice is

$$\langle \alpha_{\mathbf{h}} \rangle = \sum_{j} \left[C_j \cdot D_1(C_j) \right] \qquad (9)$$

Once the starting set has been defined, a good pathway for phase expansion is determined by a divergence procedure. In the divergence map, starting from the reflections in the starting set, each new reflection is linked to the preceding ones with the highest value of $< \alpha >$.

- PHASE EXTENSION AND REFINEMENT -

The starting set defined by the preceding step is usually formed by the origin (and enantiomorph) fixing reflections, a few one phase s.s. and a number of other phases which may be obtained:

a) by magic integer permutation (White & Woolfson, 1975; Main, 1978),

b) by a random approach (Baggio, Woolfson, Declercq & Germain, 1978; Burla, Cascarano & Giacovazzo, 1992).

The option a) is the default, b) runs if the directive RANDOM is used. In this last case a large number (depending on the available computer time) of trials can be requested.

If a partial structure is available (the directive PARTIAL should have been used in the normalization routine) the PHASE routine is automatically able to take that information into account. No further directives are strictly necessary. However directives SYMBOLS, SPECIALS, MAXTRIAL may be used to change default values.

Phase expansion and refinement are carried out by means of a tangent formula using triplets, negative quartets, psi-zero triplets and the most reliable two-phase structure seminvariants. In the weighting scheme the experimental distributions of the alpha parameters are forced to match with the theoretical ones (Burla, Cascarano, Giacovazzo, Nunzi & Polidori, 1987).

For each phase set, several FOM's are computed using all invariants and seminvariants estimated by means of the representation method. Their meaning and an optimized way of combining all the computed FOM's to give a highly selective combined figure of merit (CFOM) is described in the papers by Cascarano, Giacovazzo & Viterbo (1987) and by Cascarano, Giacovazzo & Guagliardi (1992).

All FOM's, as well as the combined CFOM, are expected to be equal to 1.0 for correct solutions. CFOM larger than 0.5 can be considered encouraging.

If pseudotranslational symmetry is present then CFOM > 0.3 may characterize the correct solution.

FOURIER/LEAST-SQUARES module

The sets of phases generated by the tangent routine are first expanded through psi-E relationships and then passed to the fast fourier transform routine written by L.F.Ten Eick (1977) and subsequently modified by the MULTAN team (Main, Fiske, Hull, Lessinger, Germain, Declercq & Woolfson, 1980).

Several additional features have been introduced in the present version.

a) From the solution with the highest CFOM value a Fourier map may be calculated by involving a supplemental number of reflections which are phased trough the so-called psi-E triplets (Altomare, Cascarano, Giacovazzo & Viterbo, 1991)

b) Special positions are handled, peaks very close to symmetry elements are moved onto symmetry elements, the site symmetry is defined and the atomic occupancy factor is calculated; key numbers for

designating free, coupled or fixed positional parameters for least-squares subroutines are also calculated together with symmetry conditions on the thermal ellipsoid.

c) Cycles of structure factor calculation -least squares- ($2F_o$ -F_c) Fourier synthesis are automatically performed in order to provide a more complete and refined crystal structure. Due to reflection overlapping the results of the refinement must carefully be considered by the user who will take always into account the ratio 'number of parameters to refine/number of independent observations'.

The program provides the following R-factors which can be used as reliability parameters:

$$R(f) = \frac{\sum |F_o - F_c|}{\sum |F_o|} \tag{10}$$

$$Rw(f) = \frac{\sum |F_o - F_c| \cdot F_o^2}{\sum |F_o| \cdot F_o^2} \tag{11}$$

$$R(f^2) = \frac{\sum |F_o^2 - F_c^2|}{\sum F_o^2} \tag{12}$$

MENU/REFINE module

Atomic parameters produced by the preceding module, stored in the direct access file, can be exported in ASCII file in a format suitable for other programs such as MOLDRAW, SCHAKAL. If a graphic interface is available, it is possible to delete, to relabel some atoms, to apply some geometrical restrains and restart the FOURIER/LEAST-SQUARES procedure, by means of the REFINE option (geometrical restrains are not actively used in RIETVELD refinement).

If the graphic interface is available, the user can decide to exploit the information provided by a well localized structure fragment (eventually modified) in the integrated intensity extraction procedure. The list of structure factors calculated by the fragment can be used as starting point in the Le Bail algorithm for a new run.

PATTERSON module

In EXPO it is possible to compute a PATTERSON map using various coefficients.

Using a special directive (INVERSION) it is possible to exploit the positivity of the electron density (in the direct space) in the extraction routine. The so-modified Patterson map is inverted providing a set of structure factor values to use in a new extraction process.

RIETVELD module

In this module an initial Rietveld refinement is performed starting from the profile model supplied by EXTRA and the structural model supplied by SIRPOW (eventually modified by the user) or supplied to the program by an external file (using the directive FRAGMENT in the FOURIER command).

The procedure can be activated by means of the corresponding option of the graphic menu in SIRPOW.

As soon as the option is selected, EXPO switches from SIRPOW to EXTRA graphic environment ; the observed, calculated and difference patterns are shown on the screen and the following options become available: MENU, RIETVELD OPTIONS, REFINE, FOURIER, SAVE, RESTORE (see HELP on line).

The routine adopts from EXTRA the subdivision of the full pattern into intervals and the same profile shape function. The initial pattern for each interval is calculated as follows:
1) zero correction, peak shape, asymmetry, FWHM and unit cell parameters are the final full-pattern refined values from EXTRA;
2) The background model is that supplied by EXTRA at the end of the interval refinement;
3) the intensity values of each reflection are calculated using the structural available model and corrected by Lorentz-polarization factor. An isotropic thermal factor for each atom is considered.

An automatic diagonal least-squares refinement can be performed (20 cycles) in which the profile and background parameters are refined for each interval; the atomic and cell parameters are refined considering the whole pattern. After each cycle the minimized quantity

$$S = \sum w_i (y_i - y_{ci})^2$$

is calculated using the refined parameters and compared with that of the previous cycle: if it is decreased, the shifts are accepted, otherwise they are rejected and the corresponding damping factors are reduced.

At the end of the refinement cycles, the following agreement indices are printed on the output file (Hill & Fischer, 1990; Young, 1993):

$$R_p = \frac{\sum_i^N |y_i - y_{ci}|}{\sum_i^N y_i}$$

$$R_{pw} = \left[\frac{\sum_i^N w_i (y_i - y_{ci})^2}{\sum_i^N w_i y_i^2} \right]^{1/2}$$

$$R'_p = \frac{\sum_i^N |y_i - y_{ci}|}{\sum_i^N (y_i - y_{bi})}$$

$$R'_{pw} = \left[\frac{\sum_i^N w_i (y_i - y_{ci})^2}{\sum_i^N w_i (y_i - y_{bi})^2} \right]^{1/2}$$

$$R_p{'}_p = \frac{\sum_i^{N_p} |y_i - y_{ci}|}{\sum_i^{N_p} (y_i - y_{bi})}$$

$$R_{pw}{'}_{pp} = \left[\frac{\sum_i^{N_p} w_i (y_i - y_{ci})^2}{\sum_i^{N_p} w_i (y_i - y_{bi})^2} \right]^{1/2}$$

13

$$R_{\exp} = \left[\frac{(N-P)}{\sum_i^N w_i y_i^2} \right]^2 \qquad\qquad R_{\exp}' = \left[\frac{(N-P)}{\sum_i^N w_i (y_i - y_{bi})^2} \right]^2$$

$$GoF = \left[\frac{\sum_i^N w_i (y_i - y_{ci})^2}{(N-P)} \right]^2 = \frac{R_{pw}}{R_{\exp}}$$

where:

N is the total number of counts in the pattern

N_p is the number of counts belonging to the peak region

w_i is $1/y_i$ (default) or $1/y_i^2$

5 - Commands and their use

The input consists of a sequence of comments, commands and directives. The commands are headed by '%' character and directives must follow the related command.

EXPO recognizes the following commands:

Command	Description
%INITIALIZE	Initialize the direct access file (to override previous results and data)
%DATA	Data input routine
%EXTRACTION	Integrated intensities extraction routine
%NORMAL	Normalization routine
%SEMINVARIANTS	Seminvariants routine
%INVARIANTS	Invariants routine
%PHASE	Converge-tangent routine
%FOURIER	Fourier/Least-Squares routine
%MENU	
%END	End of the input file
%PATTERSON	Patterson map calculation routine
%JOB string	A caption is printed in the output
%CONTINUE	The program runs in default conditions from the last given command up to the end
%STRUCTURE string	This command is used to specify the name of the structure to investigate. The program creates the name of the files needed adding the appropriate extension to the structure name. The file names are: string.bin -> direct access file string.xyz -> final coordinates file string.plt -> file for graphics If this command is not used the default string "STRUCT" (instead of the name of the structure) is used to create file names.
%WINDOW [x y]	Graphic window is required. Optionally it is possible to increase the dimension using x and y. Default values are 720, 525.
%NOWINDOW	Graphic window is suppressed
%REFINE	Fourier/Least-Squares restart routine
%RECYCLE	Extraction recycle routine

Directives are described below, in the sections dedicated to the various routines.

All commands and directives are in free format (between columns 1-80) and are case independent. Only the first four characters are significant. The keywords can start in any position. If the first non-blank character is ">", then the record is interpreted as a comment.

EXPO preserves intermediate results. For example, if invariant estimates have been already obtained during a previous run of EXPO, in a new run the command %INVARIANTS can be omitted.

Commands can be given in any order, under the following conditions:

- first routine used must be DATA, if it has not been used in a previous run;
- EXTRACTION routine must be used to extract integrated intensities from the diffraction pattern, if a reflections file is not supplied by the user;
- INVARIANT and SEMINVARIANT routines have no meaning if diffraction moduli are not normalized;
- PHASE routine has no meaning if no triplets have been calculated;
- FOURIER routine cannot run before PHASE routine (unless a fragment information is supplied by the user; in this case only DATA, EXTRACTION and FOURIER commands are needed).
- PSEUDO and PARTIAL directives require their own normalization. If one renounces to use PSEUDO or PARTIAL information, EXPO must be restarted from %DATA command.

The minimal information needed by EXPO is constituted by (see Example at page 35):

- ♦ cell parameters
- ♦ space group symbol
- ♦ cell content
- ♦ diffraction pattern (or reflections)
- ♦ wavelength
- ♦ type of radiation (if neutron or synchrotron data have been used; home diffractometer data is the default).

The following limitations hold for commands and directives:

- PTEN cannot be used if the directive PSEUDO is on;
- seminvariants cannot be requested if PSEUDO is on;
- if PARTIAL is on, PSEUDO and RANDOM cannot be used;
- the directive SYMBOLS is of no use if RANDOM is on.

How to run EXPO

There are two ways to run the program:

1) EXPO "*input file name*" "*output file name*"

The default extension of the input file is ".exp".

Output file name can be optional. If it is not supplied the default name is "structure name".out.

2) EXPO (or directly clicking on the icon on Windows95 or NT environment)

An interactive window will be open by the program with the following requests:

 - *work-files directory*

 - *input file name*

 - *output file name*

Additional options allow to select (BROWSE) the input file among all the *.exp files in the work directory and/or to edit the specified input file.

6 - Directives

Preparation of data for DATA routine

CELL a b c, α β γ

      cell dimensions a, b, and c are in angstroms, α β and γ in degrees.

SPACEGROUP string

      String is the symbol of the space group, according to International Tables (1974). Blanks are necessary among the terms constituting the space group symbol (see examples at the end of this manual).

SHIFT sx sy sz

      Origin shift components if desired by the user. They must be integral multiples of 1/24

CONTENTS El_1 n_1 El_2 n_2 El_3 n_3 ....

      Unit cell contents. El_i is the chemical symbol of atomic species *i*, n_i is the corresponding number of atoms in the unit cell (to a maximum of 8 atomic species). For each chemical element up to Cf (Z=98) X-ray scattering factor constants are stored, together with information on the atomic number and weight, covalent and Van-der-Waals radii., etc. in a file (see notes on implementation). Neutron scattering factors are coded in the program for elements up to Pu (Z=94).

ISOTOPE El sf

      To modify the value of the neutron scattering factor for element El using the value sf.

WAVELENGTH w

      w is the wavelength (in Angstrom) used to collect the powder pattern.

NEUTRON

      This directive is used to specify that neutron radiation has been used.

SYNCHROTRON

      This directive is used to specify that synchrotron radiation has been used.

RANGE thmin thmax thstep

      thmin, thmax and thstep are the minimum 2θ, the maximum 2θ and the step width respectively of the powder pattern (thmin must be equal to the 2θ corresponding to the first count in the pattern file).

RHOMAX x

Maximum value of $(\sin\theta / \lambda)^2$ accepted for reflections to be used. The default is that all the data are accepted.

FWHM x

It is supplied the value of D for the overlapping calculation. Two reflections assumed to be in overlapping if $2\theta_1 - 2\theta_2 < D * FWHM$. Default is 0.1.

PATTERN filename

Filename is the name of the profile counts file (up to 40 characters). It is used to change the default. [Default is name.pow where name is defined using the command %STRUCTURE].

FORMAT string

String is the run time format to read profile counts (default value per string is 10f8.0) or to read reflections (default value is 3i4,2f15.4).

RECORD n

It specifies the number of profile counts per record, when $n \neq 10$ (max n value is 10) or the number of reflections per record, when n>1 (only if REF1 directive is used).

JUMP

It is used to skip the reading of the first line (comment line) in the input pattern file (or reflection file).

DOUBLE

It is used when the reading of the odd lines (or, of the even lines, if JUMP is activated) in the input pattern file (or reflection file) must be skipped.

ALPHA

It is used to start the alpha2 stripping.

HISTOGRAM n

n is the number of bars in the histogram for the alpha2 stripping procedure (only if ALPHA directive is used).

EQUIPORTION

If this directive is given the overall intensity of a group of overlapping reflections is equiportioned among the reflections in the group (equal E-values are obtained).

REF1 string

This directive is used to supply an external reflection file.

If string is equal to FOLLOW then the program expects reflections on card reader after the %END or %CONTINUE command. Otherwise string is the name of the reflections file (it must be up to 40 characters).

Records have n reflections, each with h, k, l, F_{obs}, OV, where all are integer except F_{obs}. OV is equal to 1 if the reflection is single, equal to n if the reflection is in overlapping with the following n-1 reflections, equal to 0 if the reflection is in overlapping with a previous one. h,k,l cannot be larger than 63. The run time format is modified by the program as (3i4, f8.2, i2).

The end of reflections is detected using one of the following:

- blank record.
- end of file.

The program expects a complete set of reflections (observed and unobserved). Failure to observe these points will result in inaccurate calculation of the E's.

<u>REF2</u> string

This directive is used to supply an external reflection file.

EXPO reads directly the output provided by any profile fitting program. H K L   FWHM and F**2 must be supplied [default format is (3i4,2f15.4)]. If a comment line is at the beginning of the reflection file, the first character must be '>'.

Preparation of data for EXTRACTION routine

INTERVAL n

        n is the number of intervals in which the powder pattern is divided (max 8). n is only given if the user wants to change the automatic division of the pattern.

POLYNOMIAL k

        k is the background polynomial degree (max 10). k is given if the user wants to use a fixed degree in each interval and not the degree automatically selected by the program.

GAUSSIAN

        To use only if the gaussian function is preferred.

LORENTZIAN

        To use only if the lorentzian function is preferred.

PVOIGT

        To use only if the Pseudo-Voigt function is preferred.

DAMPING x

        x is the damping factor used in the least squares routine (default value 0.1). (It is convenient to use a small x value).

FSTARTING   filename

        If the "filename" file is available from any source, containing h, k, l and the corresponding F-squared values, Le Bail algorithm is started again for obtaining new estimates.

FRECYCLE   filename

        If the "filename" file is available from a previous run of EXPO (or EXTRA), containing h, k, l and the corresponding F-squared values, Le Bail algorithm is started again for obtaining new estimates.

FORMAT   string

        String is the run time format to read F-squared file if the directive FSTARTING is used. Default value for string is (3i4,15x,f15.0).

STANDARD

        To activate the procedure of selection of the profile function type (among Pearson VII, Gaussian, Lorentzian, Pseudo-Voigt) from the best fit of the standard peak .

FIRST

To activate the procedure of selection of the profile function type (among Pearson VII, Gaussian, Lorentzian, Pseudo-Voight) from the best fit of the first interval .

ALL

To activate the procedure of selection of the profile function type (among Pearson VII, Gaussian, Lorentzian, Pseudo-Voight) from the best fit of each interval.

NOSINGLE

This directive is used if the preliminary fitting of the standard peak (automatically selected by the program) is omitted. In this case a preliminary fitting of the first interval is carried out.

NOCELL

To skip the cell refinement.

ZEROPROFILE x

x is the 2θ shift value supplied by the user (to be substracted from the 2θ observed positions). In this case the automatic calculation of the 2θ shift using the preliminary fitting of the standard peak is skipped by the program).

NCYCLE n

n is the number of cycles for each interval (default value is 60).

FAST

This directive is used if the graphic interface has been requested by the "window" command. In this case, the user graphic interaction in a default run is avoided.

SLOW

This directive can be used if the graphic interface has been requested by the "window" command. In this case the program stands by after some relevant calculations and waits for the user graphic interaction to go on.

WIDTH x

x is the factor which, multiplied by the FWHM of each peak, gives the total peak width. (In this case the automatic calculation of x is not taken into account).

VERBOSE

This directive is used to have a more detailed print in the output file. In this case, the R_p and R_{pw} values and the values of the refined variables are printed interval by interval.

Preparation of data for NORMAL routine

NREFLECTIONS n

        Number of active reflections with largest E-values subject to a minimum value of E(obs) = 1.0.
Default = 4 * number of independent atoms + 50 if centrosymmetric + 50 if triclinic + 150 to a maximum of 499. If PSEUDO directive is on, the program tries to use all the E's down to 1.0.

NZRO n

        Number of smallest E-values (maximum E-value 0.3) to calculate psi-zero triplets. The default (and the maximum) is 1/3 of the strong reflections. (The directive allows to decrease that number).

BFACTOR x

        Temperature factor if the user wants to supply it. (The scale factor is assumed equal to 1).

PSEUDOTRANSLATION n(1,1) n(2,1) n(3,1) n(4,1) n(1,2) . . . n(4,3)

        If parameters are equal to zero, or omitted, the normalization routine performs statistics for the pseudotraslation effects and renormalizes the reflections on assuming the most probable pseudotranslation as prior information.

        If user wants to supply a specific pseudotranslational symmetry, and this corresponds to a class given in Table 1 the n(1,1) can be set to the class number and the other values must be omitted.
If the pseudotranslation is more complex, it should be specified by up to 3 sets of 4 values n(i,j) so that

$$(n(1,j)*h + n(2,j)*k + n(3,j)*l = n(4,j)*m) \quad j=1,3$$

where m is an integer number.

NOPSEUDO

        If this directive is given the program does not recycle the PSEUDO information.

NOPREFERRED

        If this directive is given the program does not apply the correction for preferred orientation.

PARTIAL

        To perform the procedure which aims at recovering the complete structure from a partial one. This instruction is followed by the list of atomic information (one per line) consisting of atomic species, x, y, z and optionally occupancy factor. See Example 10. Please consider the directive FRAGMENT in the module FOURIER/LEAST-SQUARES for a different approach.

<u>NPLA</u>

The number of candidate planes for preferred orientation analysis (default 10; maximum 30)
The number includes the equivalent planes which will not be considered.

<u>PLANE</u> h k l G

Preferred orientation direction supplied by the user; G is the correction factor (see NORMAL module)

<u>THRP</u> x

Threshold value of the mean fractionary scattering power (M.F.S.P.) to recycle the pseudo-translational symmetry information in a new extraction process.

Preparation of data for SEMINVARIANTS routine

FIRST

> To perform the estimation of one-phase seminvariants via the first representation.

NRS1

> The number of the strongest reflections to use for one-phase seminvariants search.

Default value = number of strongest reflections chosen by normalization routine

LIST

> Print the list of the strongest reflections to use for one-phase seminvariants search.

NUMK n

> The number of free-vectors used by the second representation formula for the one-phase

seminvariants. The default value is computed according to the structural complexity:

> $n = min (natom/msym + 40, 100)$

where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

NRS2 n

> n is the number of the strongest reflections to use for two-phase seminvariants search.

Default value = number of strongest reflections chosen by normalization routine (max 350).

Preparation of data for INVARIANTS routine

NRTRIPLETS n

      The number of strongest reflections for the sigma2 search. (Up to $E_{min} = 1.0$)

Default = number computed by normalization routine (nstrong).

The value will substitute the default one in the direct access file.

Value less than nstrong/2 is not accepted.

GMIN x

      Positive triplets with G < x are not actively used.

Default value x=0.6 (in any case x > 0.2 and x $\leq$ 1.0 is activated only if the P_{10} formula is applied).

EMIN x

      Minimum value for the normalized structure factors in the sigma2 search. x < 1.0 not accepted.

Default = minimum value of E for the strongest reflections selected by normalization routine.

NRPSIZERO n

      The number of weakest reflections for the psi-zero search.

Default = number computed by normalization routine.

The value will substitute the default one in the direct access file.

NLARGE n

      The number of largest reflections for the psi-zero search (default = number of strongest reflections used for sigma2 search).

nlarge > nrtriplets is not accepted.

EMAX x

      Maximum value of E for the weak reflections for the psi-zero search (up to a maximum of NRPSIZERO reflections). x > 0.3 not accepted.

COCHRAN

      To use the P_3 formula. (By default P_{10} formula is used).

NUMK n

      The number of free-vectors used by P_{10} formula.

Default value is provided by EXPO according to the structural complexity:

        n = min (natom / msym + 30, 70)

where natom is the number of atoms in the cell and msym is the number of symmetry operators of the space group.

CORRECTION x

        If a PSEUDOTRANSLATION procedure is on, a correction factor depending on x is applied to the probabilistic argument of weak-weak-weak triplets when they are present. The default value is calculated by program.

                x = -1 eliminates weak-weak-weak triplets,

                x = 1 corresponds to the largest confidence in them.

ESTIMATION

        To run the procedure which estimates the $|E|$'s for all the reflections and use them as a prior information in a new extraction process.

The procedure will not be applied if PARTIAL directive (NORMAL module) has been used.

The procedure will not be applied if PSEUDOTRANSLATIONAL information has been recycled.

YPSIE

        To calculate PSI-E triplets (Default doesn't use them).

NQUARTETS

        To not calculate negative quartets (Default uses them).

NPSI0

        To not calculate PSI-0 triplets (Default uses them).

Preparation of data for PHASE routine

LIST n

      The number of reflections at the top of the divergence map for which a list of the map is obtained. Default = no printout of the map.

ORIGIN n(i) phi(i)

      Codes and phases of the reflections chosen by the user to fix the origin. The total number of reflections (max 3) needed to fix the origin must be given (by default the origin is chosen by the program).

ENANTIOMORPH n

      Code of the reflection chosen by the user to fix the enantiomorph (by default the enantiomorph is chosen by the program).

SYMBOLS n

      The number of permuted phases (max 12). Default = 5.

PERMUTE n(i)

      Codes of the permuted reflections chosen by the user (i.e. included in the number of required symbols, max 12).

SPECIALS n

      The number of reflections with restricted phase to be permuted (i.e. included in the number of required symbols, max 12).

PHASE n(i) phi(i) wt(i)

      Codes, phases and weights of reflections with known phase (max 200).
 If one card is not sufficient, the directive PHASE must be repeated in the other card(s).

TABLE

      To print the table of alpha values (at the beginning of PHASE procedure).

TRIAL n(i)

      To print the final phases for the specified trial.

MINFOM x n

      x: the program automatically stops when a solution is found with CFOM > x . Default value = 1.00.
      n: is the maximum number of phase sets (with the largest combined figure of merit - CFOM) to retain (default value = 10).

<u>RANDOM</u> n

To use random phases starting sets.

n: is the number of random phases. If n is omitted, it is equal to one half of the strong reflections (nstrong).

Maximum value allowed for n = nstrong - 4

Minimum value allowed for n = nstrong / 4

<u>SEED</u> n

Random generation seed. n must be an integer large value and must be odd. (Default 67543).

<u>NOREJECT</u>

Used to retain equivalent sets of phases.

<u>TEST</u>

Used to restart the random number generator using the default seed.

<u>MAXTRIALS</u> n

Maximum number of trials when random approach is used. (Default 100).

<u>YNEG</u>

To actively use negative triplets in the phasing process (Default does not use them).

<u>YNGQ</u>

To actively use negative quartets in the phasing process (Default does not use them).

Preparation of data for FOURIER/LEAST SQUARES routine

SET n

  The serial number of the set from which to calculate the E-map. If 'SET n' is not specified, the set with the highest combined figure of merit will be used, and then if FOURIER routine is re-run, the set with the next highest combined FOM will be used.

GRID x

  The grid on which the Fourier-map is evaluated will have a spacing of approximately x angstroms. The default spacing is computed by the program (1/3 Å).

MAP

  If MAP is specified, the E-map will be printed on the lineprinter. The printing of the E-map is rarely used. The output is in sections of constant y with x across the page and z running down it on a grid of about three points per angstrom. The maximum number of grid points across the page is 36 (i.e. about 12 angstroms in length). If more are needed, a new page is started.

LEVEL n

  All numerical values in the E-map greater than n will be underlined with **** in the lineprinter output to facilitate contouring. The default is 100. The map is automatically scaled between ± 999 (approximately) at the grid points at which it is calculated. Note, however, that it is printed between ± 99. (This directive can be used only if associated to MAP directive).

LIMITS x_1 x_2 x_3

  The E-map will be printed from 0 to x_1 along x, from 0 to x_2 along y, from 0 to x_3 along z. (This directive can be used only if associated to MAP directive).

LAYX

  This directive can be used only if MAP directive is given. The Fourier-map will be printed in section of constant x.

LAYY

  This directive can be used only if MAP directive is given. The Fourier-map will be printed in section of constant y.

LAYZ

  This directive can be used only if MAP directive is given. The Fourier-map will be printed in section of constant z.

PEAKS n

The number of peaks to be searched for in the map. The default is the number of peaks for which

$$\sum(occ(i)*m) = 1.3 * n$$

where occ is the crystallographic site occupancy factor, m is the number of symmetry operators, n is the number of non-hydrogen atoms in the unit cell.

FWHM x

The value of D parameter for the condition $2\psi_1 - 2\psi_2 < D*FWHM$, may be changed. Default value is approximately 0.9 depending on the ratio between the number of observations and the number of parameters to refine.

RADIUS El x

The value x of the radius used to compute the connectivity is assigned to element El.

COORDINATION El dmin dmax n

The minimum and maximum value of the bond distanced are assigned to element El. n is the coordination type; Allowed values for bond angles are:

n= 1	55-180 Degrees
n= 2	85-145 Degrees

RECYCLE n

Used to stop the recycling procedure after cycle number n. If n is equal to zero only the E-map will be calculated.

FRAGMENT string

Used to supply a known fragment to be completed. String is the name of the file in which, for each atom, are stored: Element X Y Z

SOGF x

The percentage of the largest |F| value reflections actively used in the Fourier recycling routine.
The reflections with the smallest |F| values corresponding to the rejected percentage have a zero weight value in the refinement process. (Default value 100.0).

SOGI x

The percentage of the most intense groups of overlapping reflections actively used in the Fourier recycling routine. (Default value 91.0).

PSE

To actively use PSI-E triplets in these procedures (Default doesn't use them). (This directive can be used only if associated to YPSIE directive of INVARIANTS routine).

INTEGRATION x n

x is the radius of integration in the peak search routine. (Default value 0.5). n may be 1 to use the gaussian fitting of the fourier peak in the map or 2 to use the peak average density. (Default value 2).

Preparation of data for MENU routine

The export module provides a graphic interface to interact with the model, complete and refine it. Please refer to the On-Line Help to get information about the available features.

In the following directives, string is used to supply the output file name. If string is omitted the default file name will be used.

MOLDRAW string

A file containing peak coordinates is created in MOLDRAW format.

SCHAKAL string

A file containing peak coordinates is created in SCHAKAL format. This is the default.

XYZ string

A file containing peak coordinates is created in XYZ format (Cartesian coordinates).

Preparation of data for PATTERSON routine

MAP, LAYX, LAYY, LAYZ, PEAKS, LIMIT, GRID: See FOURIER directives.

E**2 (or E*E)

    Coefficients to be used in the Patterson synthesis. This is the default.

F**2 (or F*F)

    Coefficients to be used in the Patterson synthesis.

E*F (or F*E)

    Coefficients to be used in the Patterson synthesis.

INVERSE

    To activate the procedure of calculating the $|F|$ values from an inverted suitably modified Patterson map to use as starting point in the Le Bail algorithm in the extraction routine.

NPAT n

    The number of cycles of successive inversions of the Patterson map.(Default values: 6 for X-ray data, 5 for neutron data).

7 - When default EXPO fails: strategies

Crystal structures are solved by EXPO with only default settings. In case of failure the user can modify the default pathway in order to improve the success rate.

The number of options is quite large and some of them are suggested here.

The CFOM figure of merit is in practice an estimate of the probability that a trial solution is correct. If the highest CFOM does not correspond to the correct solution, subsequent CFOM's may be successful.

A special problem arises when several nearly equivalent CFOM's are available for different trials. Then the individual FOM's have to be carefully considered by the user: e.g., if a very small number of negative triplets and quartets is occasionally available then CPHASE gives weak indications, while ALCOMB may be more useful in these conditions. All relevant information about FOM's may be obtained through the keyword ALL.

Viceversa, if pseudotranslation symmetry is present, the FOM ALCOMB is rather poor while CPHASE may be more selective. Magic integers procedure is an effective and simple way for starting a multisolution process. Second representation formula for triplet invariants proved much more powerful than first representation formula therefore it is used as a default for estimating triplets. In case of failure, the user can choose P_3 formula, and/or conveniently increase the number of symbols, or require special types of reflections as symbols.

A check for possible pseudotranslational symmetry is always made by EXPO: if PSEUDO keyword is not activated then information is given but not used in the phasing process. If a pseudotranslational symmetry has been found by the normalization routine a new run of EXPO may be started including the keyword PSEUDO.

Perhaps because of a faulty data collection strategy, often weak reflections may not be included in diffraction data. This lack of information influences both the normalization process (scale and overall thermal factors are affected by systematic errors; the experimental E-distribution is often non-centric even when the crystal structure is centrosymmetric) and the estimation of triplet and quartet invariants: in particular, a reduced number of negative triplets (via P_{10} formula) and of negative quartets is calculated. Success in the structure solution is often obtained if weak reflections are also used.

Even if there are still a large number of available options, three of them are quoted for special cases:

a) the value of NREF (number of reflections actively used in the phasing process) is fixed by EXPO. For some special structures the ratio "number of active triplets/NREF" is too small (less than 10). Larger values of NREF may improve the phasing procedure.

b) high (or low) resolution reflections may occasionally play a too important role in the first steps of convergence and divergence process. Fixing a thermal factor lower (or larger) than that provided by the normalization routine may successfull change convergence and divergence process.

c) an alternative space group should be carefully considered.

Example of "command input file" for a default run (home diffractometer data)

%window
%structure crox
%job data from a home diffractometer
%init
%data
pattern crox.pow
range 6.0 80.0 0.02
cell 5.447 6.5576 12.1147 106.382 95.715 77.970
content Cr 8 O 21
wavelength 1.3922
space p -1
%continue

Default "profile counts file" name: crox.pow

crox.pow content:

```
 1166.  1179.  1224.  1172.  1221.  1159.  1204.  1347.  1295.  1146.
 1353.  1188.  1214.  1250.  1253.  1240.  1285.  1344.  1314.  1324.
 1438.  1444.  1463.  1522.  1664.  1548.  1836.  1865.  1913.  2160.
 2367.  2412.  2856.  3186.  3681.  4280.  4941.  5883.  7298.  8978.
11248. 14375. 20187. 25377. 23771. 15690. 8496.  6385.  5553.  5022.
 4643.  4371.  4202.  3840.  3684.  3296.  3199.  3047.  2833.  2710.
 2567.  2334.  2412.  2156.  2163.  2134.  1981.  1981.  2033.  1901.
 1849.  1768.  1761.  1709.  1654.  1616.  1580.  1629.  1561.  1609.
 1658.  1609.  1493.  1334.  1298.  1331.  1305.  1289.  1292.  1214.
 1149.  1182.  1081.  1146.  1094.  1065.  1030.  1055.   981.  1020.
 ........ ....... ....... ....... ....... ...... ...... ....... ...... ......
```

Examples of non-default input command file

Example 1 [Data are from neutron source and the profile counts file format is not the default one].

```
%window
%struct cfcl
%job cfcl.dat from neutron source
%data
pattern  cfcl.pow
range  5.0 150.0 0.05
cell  10.168 14.964 5.100 90.000 90.000 90.000
content  C 8 F 16 Cl 16
space  f d d 2
wave  1.595
neutron
record  1
format  (f8.0)
%continue
```

"cfcl.pow" content:
```
    391.
    403.
    422.
    427.
    .................
```

Example 2 [The profile counts file name is not the default one and the α_2 stripping with a 3 bars histogram is requested].

```
%window
%struct  zno
%job standard zno
%data
pattern  zno.cou
range  5.0 100.0 0.02
cell  3.249 3.249 5.205 90.0 90.0 120.0
contet  zn 2 o 2
space p 63 m c
wave 1.54056
alpha
histog 3
```

%continue

Example 3 [A Pseudo-Voigt function will be used and the background in each interval will be described by a constant. The counts file format is not the default one and the graphic window is suppressed].

%nowindow
%struct mes
%job mes.dat
%data
pattern mes.pow
range 5.0 88.0 0.02
cell 8.588 9.931 11.105 90.0 93.754 90.0
content C 24 N 4 O 20 S 4 H 52
space p 21/c
wave 1.5406
double
%extraction
pvoigt
polynomial 0
%end

Example 4 [Tthe starting values of the integrated intensities are supplied by the user in the file "yono.ext" and the number of cycles is set to 40].

%window
%struct yono
%job yono.dat fron synchrotron data
%data
pattern yono.pow
range 7.0 80.0 0.01
cell 9.385 16.394 3.627 90.000 101.067 90.000
content Y 8 O 26 N 2 H 18
space p 21
wave 1.2323
synchrotron
%extraction
fstarting yono.ext
ncycle 40
%end

Example 5 [The preliminary fitting of the standard peak is skipped and the starting 2θ shift value is supplied by the user (the program refines it)].

```
%window
%struct agpz
%job agpz.dat
%data
pattern  agpz.pow
range  5.0 80.0 0.02
cell  6.526 20.059 6.464 90.000 90.000 90.000
space  p b c a
content  Ag 8 N 16 C 24 H 24
wave  1.54056
%extract
nosing
zeroprof  0.02
%continue
```

Example 6 [Neutron data and a deuterium atom in the cell content are used. Fast graphic interaction is activated].

```
%window
%structure baco
%job  structure BaC2O4
%init
%data
neutron
isotope H 0.67
range 20.7  150.0  0.05
pattern baco.pow
cell  10.0632  7.9330  6.8487   90.0  122.3423  90.0
space C 2/m
cont Ba 4  C 8  O 20  H 8
wave 1.909
%extract
fast
%continue
```

Example 7 [The program restarts from the PHASE procedure by increasing the number of symbols after a previous run using the default 5 symbols].

%structure yono
%phase
symbols 8
%continue

Example 8 [A random approach is used and the best 100 sets, over 250 trials, of phases are retained in the direct access file. No Fourier is required.]

%nowindow
%structure nizr
%phase
random
maxtrials 250
minfom 1.0 100
%end

Example 9 [The E-map of the set number 7and the complete FOURIER/LEAST-SQUARES procedure for set number 10 are requested by the user].

%window
%structure nbpo
%fourier
set 7
recyc 0
%fourier
set 10
%continue

Example 10 [The program restarts from the NORMALIZATION procedure by using the PARTIAL directive].

%window
%structure PBS
%normal
  partial   Pb   .9789   .3384   .3803
          S   .5010   .3318   .3916

%continue

Example 11 [To view only the structure previously produced by FOURIER routine.]

%window
%structure crox
%menu
%end

Example 12 [Use of the pseudotranslational symmetry (if it is present) as prior information for a new extraction process is inhibited.]

%nowindow
%struct agpz
%job agpz.dat
%data
pattern agpz.cnt
range 5.0 80.0 0.02
cell 6.526 20.059 6.464 90.000 90.000 90.000
space p b c a
content Ag 8 N 16 C 24 H 24
wave 1.54056
%extract
%normal
nopseudo
%continue

Example 13 [Use of the triplet estimation as prior information for a new extraction process.]

%nowindow
%struct mes
%job mes.dat
%data
pattern mes.cnt
range 5.0 88.0 0.02
cell 8.588 9.931 11.105 90.0 93.754 90.0
content C 24 N 4 O 20 S 4 H 52
space p 21/c

wave 1.5406

double

%extraction

%normal

nopseudo

%invariants

estimation

%continue

Example 14 [Use of the Patterson map inversion as prior information for a new extraction process.]

%nowindow

%struct lasi

%job lasi.dat

%init

%data

pattern lasi.pow

range 10.0 115.724 0.0529

cell 5.4059 8.7934 14.2754 90.000 112.731 90.00

content La 8 Si 8 O 8

space p 21/c

wave 2.3400

neutro

format (10f8.0)

%extraction

%normal

nops

%patterson

inversion

%continue

Example 15 [A known fragment is used to complete the structure by appling the FOURIER/LEAST-
　　　　　SQUARES procedure. The binary file "tipor.bin" must exist].

%structure sapo

%fourier

fragment SAPO.FRG

%continue

Coordinates are in the file "SAPO.FRG" which contains

Si .00000   .00000   .00000

Table 1

Classes of reflections corresponding to the low index pseudo-translational symmetry.

1) h+k+l= 1n	2) h= 2n	3) k= 2n
4) l= 2n	5) h+k+l= 2n	6) h+k= 2n
7) h+l= 2n	8) k+l= 2n	9) h= 3n
10) k= 3n	11) l= 3n	12) h+k= 3n
13) h+l= 3n	14) k+l= 3n	15) h+k+l= 3n
16) h+k+2l= 3n	17) h+2k+l= 3n	18) 2h+k+l= 3n
19) h+2k= 3n	20) h+2l= 3n	21) k+2l= 3n
22) l= 4n	23) k= 4n	24) h= 4n
25) h+k= 4n	26 h+l= 4n	27) k+l= 4n
28) h+k+l= 4n	29) 2h+2k+l= 4n	30) 2h+k+2l= 4n
31) h+2k+2l= 4n	32) 2h+k+l= 4n	33) h+2k+l= 4n
34) h+k+2l= 4n	35) h+2k= 4n	36) h+2l= 4n
37) k+2l= 4n	38) 2h +k= 4n	39) 2h+l= 4n
40) 2k+l= 4n	41) 3h+3k+l= 4n	42) 3h+k+3l= 4n
43) h+3k+3l= 4n	44) h+2k+3l= 4n	45) h+3k+2l= 4n
46) 3h+k+2l= 4n	47) h+3k= 4n	48) h+3l= 4n
49) k+3l= 4n	50) 3k+2l= 6n	51) 2k+3l= 6n
52) 2h+3k= 6n	53) 3h+2k= 6n	54) 3h+2l= 6n
55) 2h+3l= 6n	56) 2h+2k+3l= 6n	57) 3h+2k+3l= 6n
58) 3h+3k+2l= 6n	59) 4k+3l=12n	60) 4h+3l=12n
61) 4h+3k=12n	62) 3k+4l=12n	63) 3h+4k=12n
64) 3h+4l=12n		

65) h = 2n &   k = 2n
66) h = 2n &   l = 2n
67) k = 2n &   l = 2n
68) h = 2n & k+l = 2n
69) k = 2n & h+l = 2n
70) l = 2n & h+k = 2n
71) h+k = 2n & h+l = 2n
72) h= 2n & k= 2n & l= 2n

References

1 - Altomare A., Cascarano G., Giacovazzo C. & Guagliardi A. (1993). J. Appl. Cryst. **26**, 343-350.

2 - Altomare A., Cascarano G., Giacovazzo C. & Viterbo D. (1991). Acta Cryst. **A47**, 744-748.

3 - Altomare A., Cascarano G., Giacovazzo C., Guagliardi A., Burla M.C., Polidori G. & Camalli M. (1994). J. Appl. Cryst. **27**,435-436.

4 - Baggio R., Woolfson M.M., Declerq J.P. & Germain G. (1978). Acta Cryst. **A34**, 883-892.

5 - Burla M.C., Cascarano G. & Giacovazzo C. (1992). Acta Cryst. **A48**, 906-912.

6 - Burla M.C., Cascarano G., Fares E., Giacovazzo C., Polidori G. & Spagna R. (1989). Acta Cryst. **A45**, 781-786.

7 - Burla M.C., Cascarano G., Giacovazzo C., Nunzi A. & Polidori G. (1987). Acta Cryst. **A43**, 370-374.

8 - Burla M.C., Giacovazzo C. & Polidori G. (1987). Acta Cryst. **A43**, 797-802.

9 - Burla M.C., Giacovazzo C. & Polidori G. (1989). Acta Cryst. **A45**, 94-104.

1 - Burla M.C., Nunzi A., Giacovazzo C. & Polidori G. (1981). Acta Cryst. **A37**, 677-684.

2 - Burzlaff H. & Hountas A. (1982). J. Appl. Cryst. **15**, 464-467.

3 - Camalli M., Giacovazzo C. & Spagna R. (1985). Acta Cryst. **A41**, 605-613.

4 - Carrozzini B., Giacovazzo C., Guagliardi A., Rizzi R., Burla M.C. & Polidori G. (1997). J. Appl. Cryst. **30**, 92-97

5 -   Cascarano G., Giacovazzo C. & Guagliardi A. (1991). Acta Cryst. **A47**, 698-702.

6 - Cascarano G., Giacovazzo C. & Guagliardi A. (1992). Acta Cryst. **A48**, 859-865.

7 - Cascarano G., Giacovazzo C. & Luic' M. (1988-a). Acta Cryst. **A44**, 176-183.

8 - Cascarano G., Giacovazzo C. & Luic' M. (1988-b). Acta Cryst. **A44**, 183-188

9 - Cascarano G., Giacovazzo C. & Viterbo D. (1987). Acta Cryst. **A43**, 22-29.

10 - Cascarano G., Giacovazzo C., Burla M.C., Nunzi A. & Polidori G. (1984). Acta Cryst. **A40**, 389-394.

11 - Cascarano G., Giacovazzo C., Calabrese G., Burla M.C., Nunzi A., Polidori G. & Viterbo D. (1984).   Z. Kristallogr., **167**, 34-47.

12 - Cascarano G., Giacovazzo C., Camalli M., Spagna R., Burla M.C., Nunzi A. & Polidori G. (1984). Acta Cryst. **A40**, 278-283.

13 - Cascarano, G., Giacovazzo, C., Guagliardi, A. & Steadman, N. (1991). Acta Cryst. A**47**, 480-484

14 - Cochran W.(1955). Acta Cryst. **8**, 473-478.

15 - Fan H., Yao J. & Qian J. (1988). Acta Cryst. **A44**, 688-691.

16 - Germain G., Main P. & Woolfson M.M. (1970). Acta Cryst. **B26**, 274-285.

17 - Giacovazzo C. (1976). Acta Cryst. **A32**, 958-966.

18 - Giacovazzo C. (1977). Acta Cryst. **A33**, 933-944.

19 - Giacovazzo C. (1978). Acta Cryst. **A34**, 562-574.

20 - Giacovazzo C. (1979). Acta Cryst. **A35**, 757-764.

21 - Giacovazzo C. (1980). Acta Cryst. **A36**, 362-372.

22 - Giacovazzo C. (1983). Acta Cryst. **A39**, 685-692.

23 - Giacovazzo C., Burla M.C. & Cascarano G. (1992). Acta Cryst. **A48**, 901-906.

24 - Giacovazzo C., Spagna R., Vickovic I. & Viterbo D. (1979). Acta Cryst. **A35**, 401-412.

25 - Hill R.J. & Fischer R.X. (1990). J.Appl.Cryst, **23**, 462-468.

26 - Le Bail A., Duroy H. & Fourquet J.L. (1988). Math. Res. Bull. **23**, 447-452.

27 - Main P. (1977). Acta Cryst. **A33**, 750-757.

28 - Main P. (1978). Acta Cryst. **A34**, 31-38.

29 - Main P., Fiske S.J., Hull S.E., Lessinger L., Germain G., Declercq J.P. & Woolfson M.M. (1980) - MULTAN80, a system of computer programs for the automatic solution of crystal structures from x-ray diffraction data - Univ. of York, England.

30 - Ten Eick L.F. (1977). Acta Cryst. **A33**, 486-492.

31 - White P.S. & Woolfson M.M. (1975). Acta Cryst. **A31**, 53-56.

32 - Young R.A. (1993). The Rietveld Method, Oxford University Press.