

MOSFLM 7.0.4 User Guide

This describes MOSFLM version 7.0.4 for processing image plate and CCD data

(17th October 2008)

Andrew G.W. Leslie
MRC Laboratory of Molecular Biology
Hills Road,
Cambridge CB2 2QH
UK

E-mail: andrew@mrc-lmb.cam.ac.uk
Tel (+44) (0) 1223-248011

Any constructive comments on this User Guide would be very welcome.

Index

Major Changes

Help Library

Important notes

1: Overview

- 1.1 Programs covered in this guide
- 1.2 Input and Output files
- 1.3 Allowed detector types
 - 1.3.1 Using the DETECTOR keyword
 - 1.3.2 Using the SITE keyword
- 1.3 Allowed detector types
- 1.4 Inspection of images
 - 1.4.1 Writing JPEGs

2: A Quick Guide

- 2.1 Startup keywords
- 2.2 Masking "bad" regions of the detector or setting the direct beam co-ordinates
- 2.3 Autoindexing

- 2.3.1 DPS autoindexing
 - 2.3.1.1 Unknown symmetry
 - 2.3.1.2 Space group information already given
- 2.3.2 REFIX autoindexing
 - 2.3.2.1 Unknown symmetry
 - 2.3.2.2 Space group information already given
- 2.4 Estimating mosaic spread
- 2.5 Running the Strategy option
- 2.6 Determining oscillation angles with the TESTGEN option
- 2.7 Integrating the first image to determine if the exposure time is OK
- 2.8 Interpreting those "WARNING" messages
- 2.9 Getting accurate cell parameters
- 2.10 Integrating a block of images
- 2.11 Integrating the dataset

3: Determination of crystal orientation, cell parameters and spacegroup

- 3.1 Autoindexing Interactively
- 3.2 Autoindexing when running the program in background
 - 3.2.1 REFIX and general notes
 - 3.2.2 DPS Indexing in background
 - 3.2.2.1 Autoindexing using different images from the same crystal

4: Running the STRATEGY and TESTGEN options

- 4.1 Overview of the STRATEGY option
- 4.2 Some Examples of the STRATEGY options
- 4.3 Determining the oscillation angle for each image (TESTGEN option)

5: Determining Accurate Cell parameters

- 5.1 Using Post-refinement to refine the cell

6: Collecting data and processing the images

- 6.1 Overview
- 6.2 Special MOSFLM features
 - 6.2.1 Accumulating profiles over several images

- 6.2.2 Addition of partials (ADDPART)
- 6.2.3 Post-refinement of orientation and cell parameters
- 6.2.4 Optimisation of measurement box parameters
- 6.3 Running a processing job
 - 6.3.1 Running MOSFLM interactively
 - 6.3.2 Processing the first block of data) (Non-interactively)
 - 6.3.3 Finally, processing the dataset

7: Interpreting the output

- 7.1 The log files
- 7.2 The summary file
- 7.3 Checking the quality of the data

8: General tips.

- 8.1 Estimating the GAIN of a detector
- 8.2 Processing images with no (or very few) fully recorded reflections
- 8.3 Processing images when the spots are not fully resolved
- 8.4 Processing data from other detectors, or standard detectors with different rotation axis orientation.

9: Example command files

- 9.1 Autoindexing an initial image (interactively)
- 9.2 Determining an accurate cell
- 9.3 Integrating a series of images

Appendix I

Setting the measurement box parameters manually

Appendix II

Overview of the MOSFLM program

Appendix III

Definition of coordinate systems

MAJOR CHANGES

Changes since 7.0.3

Enhancements to the iMosflm GUI

- A greatly improved and extended version of the iMosflm GUI. This now has almost all the functionality of the old X11 based ipmosflm GUI and is much easier to use.

We strongly recommend users switch to the new iMosflm GUI as the X11 version is now inferior and will no longer be supported.

For expert users, iMosflm now has the option of specifying ipmosflm keywords for the very specialised options that are only rarely used and are not yet included in the iMosflm GUI explicitly.

- A "wait" option allows images that have not yet been collected when the iMosflm session is started to be included in an integration run (previously this was not possible). This is useful when processing data as it is collected at a synchrotron (or on a home source).
- The "testgen" option, which calculates the maximum oscillation angle possible while avoiding spot overlaps, or reports the number of overlaps for provided oscillation angles, has been fully implemented.
- It is possible to run both POINTLESS and SCALA directly from the iMosflm GUI. Results are reported in a web browser using Baubles. In both cases, the programs can only be run in default mode (no user input is possible).
- A new iMosflm User Guide is available and the iMosflm Tutorial has been updated.

Enhancements to ipmosflm

- Significant improvements in the spot finding routine when dealing with very weak images or very long unit cells resulting in poorly resolved spots. The improved spot finding increases the success rate of the autoindexing. and some changes to the indexing itself has further improved the reliability of indexing.
- New detectors such as the Pilatus, Saturn, Jupiter, Oxford Diffraction and Bruker CCDs can all be handled better than previously.
- Many bugs fixed

Changes since 7.0.2

This is a bug-fix release which addresses some minor (but annoying) issues which have come to light since the release of version 7.0.2.

The following bugs have been fixed in Mosflm itself. Other issues that arose in iMosflm have been addressed separately:

- Image display for R-Axis detectors was reversed in iMosflm (the new interface).
- The presence of extremely large pixel values (>262128) for Mar IP scanners caused Mosflm to shut down.
- Direct beam co-ordinates read from Mar CCD image headers not sited at ESRF had X and Y swapped.
- Direct beam co-ordinates were not used from Ed Westbrook NOIR images by iMosflm (although they were used by Mosflm).
- Reversephi check box in iMosflm did not function as intended.

Changes since 7.0.1

- New detectors - Mar555 Flat Panel. Pilatus "miniCBF" images added.
- Better support for Bruker 100 Format images (converted with *frm2frm*).
- Change IP address for communication with iMosflm to the loopback address (127.0.0.1 or "localhost").
- Rectangular non-square detectors added to iMosflm display.
- Bug involving R-Axis detectors writing d*Trek style headers fixed.
- Many small bugs fixed.

Changes since 7.0.0

- New detectors - "native format" Pilatus 6M (i.e. *not* "miniCBF") and ADSC Q270
- Better support for CBF files produced by ADSC detectors.
- Indexing made more robust for difficult cases. New keywords added for command-line indexing.
- Native MS-Windows version now available (not relying on Cygwin environment).
- Native Mac Intel build. Statically linked so it is not dependent on a specific version of libgfortran.*.dylib being installed on the user's machine.
- Check added to make sure the current image has finished being written.
- Many small bugs fixed.

Changes since 6.2.6

- Many changes which allow processing information to be communicated between Mosflm and the new TclTk-based Graphical User Interface, which is now called iMosflm. Mosflm itself can be run as before as a standalone program, but can also be run from iMosflm. The original X11 GUI, based on John Campbell's XDL_VIEW libraries is still included and can be used if desired.
- Now checks to see if an image file is still being written to before attempting to process it.
- More improvements to the autoindexing to make it more robust in cases where it failed previously.
- Build developed which will run (without the X11 graphics) on MS-Windows, allowing iMosflm to run properly on that operating system. It works on both XP and Vista, and may well work on 2000 and 98.

- The output directory can now be specified so that it does not have to be in the same directory as that in which Mosfilm is started.
- Many small bug fixes which address rare but annoying problems.

Changes since 6.2.5

- Improvements to spot finding for autoindexing, including a new mode "FINDSPOTS LOCAL", which uses a locally (rather than radially) calculated background.
- Many changes for use with the new Graphical User Interface.
- numerous small bug fixes and minor changes.

Changes since 6.2.4

- Detector contribution to uncertainty in $\sigma(I)$ has been improved.
- More types of excluded area can now be specified, either via the command-line (see major command EXCLUDE) or using mouse clicks on an image in the GUI. The new areas are
 - ◆ generalized convex quadrilateral
 - ◆ *inside* an arc region
 - ◆ *outside* an arc region
 - ◆ Ice or other powder rings can also be picked with mouse clicks.

This has been accompanied by a tidying up of the "Beam/backstop" button in the GUI.

- Autoindexing has been further improved, and is also more flexible and gives (optionally) more feedback. All options are available via either the GUI or comand line indexing. These include:
 - ◆ choice of minimum cell edge
 - ◆ L.S. refinement of basis triclinic solution prior to symmetry tests
 - ◆ Distortion index in addition to Penalty for each solution
- The user can now choose either hexagonal (H3, H32) or rhombohedral (R3, R32) axes for rhombohedral cells.
- New detectors added include MarMosaic 325mm and Ed Westbrook's lens coupled CCD detector (NOIR at APS).
- output files can now be written for input directly to BEST, Sasha Popov & Gleb Bourenkov's improved strategy program.
- minor improvements to STRATEGY option
- numerous small bug fixes and minor changes.

Changes since 6.2.3

- CCP4 version 4 no longer supported. Mosfilm version 6.2.4 requires the prior installation of CCP4 version 5.
- Revamped build, which should be easier to customize for different compilers.
- Beam centres now read from MarCCD; improved processing of Oxford Diffraction and Bruker images.
- Output option for Sasha Popov's intelligent strategy program BEST (available from him at sasha@embl-hamburg.de) - invoked by keywords BEST ON.
- The bug which occasionally produced the message "QOPEN (no streams left)" has been fixed.
- Many other small bug fixes and minor changes.

Changes since 6.2.2

- The refined value of the mosaic spread is now used by default (previously it was NOT used by default). To make the refinement more stable, the refined value is actually a running average of the refined value over several images. A small "safety factor" is also added. Additional subkeywords **MOSSMOOTH** and **MOSADD** on the **POSTREF** keyword control this. If the mosaic spread is large (more than 0.7 degrees) the refined value may be too small. Check the prediction to test this. If it is too small, either use a larger "safety factor" (**MOSADD**), or do not use the refined value by giving keywords **POSTREF USEBEAM OFF**.
- The detector type is recognised automatically from the image for all common detectors. The **DETECTOR** keyword is therefore no longer required in most cases.
- Further changes to post-refinement to make it more robust.
- New detectors added: Mar Mosaic 225, Rigaku Jupiter, Saturn, Mercury and new-style R-Axis. Old- and new-style R-Axis HTC images. Bruker SMART and Proteum, but only if written "unwarped" and with 2-bytes per pixel. 1 and 4 byte per pixel images are not supported. Oxford Diffraction Sapphire.
- Autoindexing improved:
 - ◆ Further changes to improve robustness.
 - ◆ solutions can now be refined even if the user has already input one.
 - ◆ RMS error in spot positions is also taken into account by the "autopicking" if all solutions are refined.
 - ◆ spots near ice-rings can be excluded from autoindexing.
- Circle fitting now works for ice-ring ellipses on a "swung out" detector.
- Messages printed on unexpected shutdowns now give more information for debugging, and for interaction with CCP4i Graphical User Interface.
- SPOTOD file is now opened as "scratch", so is deleted when the program finishes.
- Reading and writing the SPOTOD file is now buffered to give an immense improvement in performance on Linux boxes using NFS mounted disks. This may be disabled using the keywords **NUSPOT OFF**.
- Mosaicity estimation improved so that it does not crash the program if the cell is large and diffracts to high resolution.
- Ability to write out JPEG files for the current image using code originally developed for a new GUI (see keyword "XGUI").

Some improvements to command-line (or batch) processing;

- Spot file used for autoindexing can be written with a user-defined name (**FINDSPOTS FILE <filename>**) and can be used again in command-line autoindexing (**AUTOINDEX DPS FILE <filename>**).

Changes since 6.2.1

- The conflict of space groups introduced by CCP4 version 4.2 has been resolved; Mosflm now conforms to the situation where, *e.g.* 'H3' refers to the hexagonal setting of the space group formerly called 'R3'.
- The DPS autoindexing routine has been made more robust, especially for those cases where one cell edge is much greater than at least one other.
- The option for refining each solution from DPS autoindexing prior to making a choice has been modified slightly to give better statistics.

Changes since 6.2.0

- Command line spot finding independently of autoindexing (**FINDSPOTS** **FIND** command).
- Duplicate directory names supplied by **DIRECTORY** command are now detected and removed.
- **DATASETNAME** and **PROJECTNAME** are now set by the program if omitted (can be over-ridden with **DNAME** and **PNAME** commands, respectively).
- Maximum length of command line increased to 400 characters.

Changes since 6.11

- Change in the use of the variable **CCOMEGA** in refinement.
- Multiple runs of **DPS** autoindexing using different images from the same crystal modified to obtain the matrix closest to the original.
- Many small changes to allow more automated processing without use of the GUI, *e.g.* improved command line autoindexing (**AUTOINDEX** **DPS**) and mosaicity estimation (**MOSAIC** **ESTIMATE**).
- Autoindexing can now use a spot file written in Mosflm format by external procedures (**AUTOINDEX** **IMAGE** **<n>** **FILE** **<filename>**)
- Mosaicity estimation improved.
- Postrefinement for fine-phi slicing (**POSTREF** **MULTI**) has been improved and is now the default.
- **WAIT** command modified to allow minutes and seconds, and a delay to allow completion of image writing.
- **ADSC** Q210 (binned and unbinned) and Q315 (binned mode) detectors added to default distribution; unbinned mode for Q315 requires a rebuild.

Changes since 6.10

Multiple mosflm.lp files via environment variable **MOSFLM_VERSION_NUMBERS**; background **DPS** autoindexing; many small modifications and bug fixes. Installation cleaned up (now using **CCP4** distributed **XDL_VIEW** libraries for all platforms).

Changes since 6.01

Automatic mosaicity estimation (via GUI). New post refinement option to allow post refinement when the sum of the mosaicity and beam divergence is more than twice the oscillation angle (**POSTREF** **MULTI**). Improvements to autoindexing and spot finding. Improved circle fitting, and option to define backstop shadow interactively. Anisotropic resolution limits allowed. New detectors: Brandeis 2x2 CCD (**B4**), R-Axis-V and **DIP2040**. Reads (basic) **CBF** format images. Data harvesting (not fully implemented, requires latest **CCP4** library). Many small bug fixes.

Changes since 6.00

Cell refinement added to the new **DPS** autoindexing, plus refinement of direct beam position. Manual spot-finding option. Option to write multiple **MTZ** files (if processing while collecting data). Fit circles option. Bug fix for non-zero two-theta indexing. Many small bug fixes.

Changes since 5.51

New FFT based autoindexing algorithm from DPS added (Steller,Bolotovskiy and Rossmann (1998) J. Appl. Cryst. 30, 1036-1040.) New detector types LIPS (large image plate scanner at ESRF) and SBC1 (Westbrook detector at APS) added.

Changes since 5.50

New detector type MARCCD for 135mm circular CCD detector from Mar Research. Allows partials to lie on up to 100 images (previous limit was 10) New keyword TEMPLATE for more general definition of image filenames. Generalise direction of two-theta axis (previously had to be parallel to fast changing direction in image). More changes to STRATEGY algorithm. Standardise FORTRAN so that code will compile under Linux (this needs a new version of the autoindexing code...post 5/5/98)

Changes since 5.40

Improved strategy algorithms. New detector types (Mar345, ADSC CCD) A host of minor bug fixes, and changes to make the program easier to run.

Changes since 5.30

The major change to version 5.40 is that the code for the spot-finding program (IMSTILLS) and autoindexing (REFIX) has been incorporated into MOSFLM. A new menu for the X-window interface has been introduced, which allows the user to find spots, autoindex images, run the strategy option, refine cell parameters (using post-refinement) and integrate images interactively. (All these features are of course still available when running a background job). The new menu is invoked by using the "IMAGE" keyword to read in an initial image. The data collection strategy option has been available since version 5.30, but has been improved in the current version, particularly by the option to speed up the calculation by any desired factor and optimise anomalous data. Additional image formats have been added. The program can now handle images from Mar Research, R-Axis (II or IV), Mac Science, Molecular Dynamics, Fuji and ESRF CCD detectors.

The following keywords are no longer necessary (but can still be given to override program defaults) :

RASTER
SEPARATION
GENFILE
HKLOUT
PIXEL

and for Mar Research, ADSC, Mac Science and R-AxisIV detectors:

WAVELENGTH
DISTANCE

and oscillation angles.

(Note however that the header information is not always correct for Mar detectors at synchrotron sites, because the software controlling the spindle axis (and/or distance) does not communicate with the Mar software controlling the detector. Check this with the station manager)

See [Major Changes](#) for a more detailed list of major changes from earlier versions.

Help Library

This User Guide is not exhaustive in describing all options available. However all possible keywords are described in the help library file (mosflm.hlp) which is part of the distribution. The help library is an ascii file, and can therefore be read with an editor, or invoked by typing "HELP" at the mosflm prompt (==>) when running the program interactively. Note that the environment variable "CCP4_HELPDIR" should point to the directory containing the help library. The bugs that prevented the online help working correctly have now been fixed.

Important notes

The source code as distributed contains the code for the FFT based autoindexing but NOT for REFIX. I can distribute the REFIX code by E-mail to academic institutions only, or to those who already have the Mar XDS software. Please send an E-mail to the address given above to get the REFIX autoindexing code.

[Back to Contents Page](#)

1: Overview

- 1.1 [Programs covered in this guide](#)
 - 1.2 [Input and Output files](#)
 - 1.3 [Allowed detector types](#)
 - 1.3.1 [Using the DETECTOR keyword](#)
 - 1.3.2 [Using the SITE keyword](#)
 - 1.3 [Allowed detector types](#)
 - 1.4 [Inspection of images](#)
 - 1.4.1 [Writing JPEGs](#)
 - 1.5 [Example input](#)
-

1.1 Programs covered in this guide

Data processing falls naturally into three sections:

- 1) Determining the crystal orientation, cell parameters and possible space group.
- 2) Generating the reflection lists and integrating the images.
- 3) Scaling and merging the resulting data.

These notes will be restricted to topics (1) and (2), which are now both present in the MOSFLM program alone. The CCP4 program SCALA is strongly recommended for the third step (scaling and merging).

1.2 Input and Output files

There are several input and output files and it is crucial that the output files are given unique filenames when two (or more) processing jobs are being run from the same directory, or the results are very unpredictable!

Input files

- 1) The image file
- 2) [The file containing the crystal orientation matrices]

NAMING CONVENTION FOR IMAGES

It is assumed that the images conform to a naming convention where the image name is made up of three parts, an identifier, a three digit number and an extension. The identifier can be up to 40 characters long, and should be separated from the three digit number by a hyphen (-) or an underscore (_). The extension can be up to 8 characters long and should be separated from the three digit number by a period (.). Note that the identifier can contain underscores or hyphens.

Examples of valid image filenames are:

```
lysozyme_cryst1_021.image
catx1_001.img
fl_tray42_wellb6_001.osc
```

ALTERNATIVE NAMING CONVENTION

If the image filenames do not conform to the specification given above, the **TEMPLATE** keyword can be used to define a very general format for the image filenames. If the **TEMPLATE** keyword is to be used, it *MUST* precede the **IMAGE** keyword in the input, and the image **NUMBER**, not the filename, should be given. See **TEMPLATE** for more information.

eg

```
TEMPLATE fred ###
IMAGE 23 PHI 22 TO 23
```

will read the file "fred_023" (no filename extension).

Output Files

- 1) The output MTZ file containing the integrated intensities. Set with keyword **HKLOUT** or on the command line using **HKLOUT**. When not given, a default filename made up from the crystal identifier (or **TEMPLATE**) and the first image number is used. It is now possible to specify that multiple MTZ files are written during one integration run (subkeyword **MULTIPLE** on **HKLOUT** keyword. In this case, a separate MTZ file will be written for each block of images processed (see **BLOCK** subkeyword of **PROCESS** keyword). The filenames will be distinguished by _001, _002, _003 etc being appended to the specified (or default) filename. *e.g.* lysx1_1to50_001.mtz, lysx1_1to50_002.mtz etc. This allows some of the data to be scaled and merged prior to the data processing finishing. (Binary)
- 2) If autoindexing or refining cell parameters a file containing the refined crystal orientation matrices is written. Filename set with keyword **NEWMAT**, defaults to **NEWMAT**. (ASCII)
- 3) The summary file. Contains a summary of processing results. Can be assigned on the command line (**SUMMARY**), defaults to **SUMMARY**. This file can be input to loggraph for graphical representation.(ASCII)
- 4) When running interactively, all output written to the terminal window is also written to the file "mosflm.lp" (ASCII). If the environment variable **MOSFLM_VERSION_NUMBERS** has been assigned, the program will

write sequentially numbered output files `mosflm_*.lp` (* = 01 - 99) for successive runs of the program.

Temporary Files

- 1) The "Generate" file. Assigned with keyword **GENFILE**, defaults to the same as the MTZ filename but with the extension ".gen" instead of ".mtz". (Binary)
- 2) The measurement boxes file. Assigned on the command line using **SPOTOD** defaults to **SPOTOD**. This file can be very large, and should normally be assigned to a scratch disk and deleted as part of the command procedure. From version 6.2.3, this file is automatically assigned to be scratch so does not need to be deleted by a command procedure.
- 3) A reflection coordinate list, assigned using **COORDS** on the command line. This is only produced when the "SEPARATION CLOSE" option is being used to process images with very closely separated spots.

Example of a command file:

```
ipmosflm HKLOUT lyso_srs.mtz SUMMARY lyso_srs.sum \
                                SPOTOD /scr0/andrew/lys.spotod \
                                COORDS /scr0/andrew/lys.coords << eof-ipmos
GENFILE /scr0/andrew/lys.gen
NEWMAT lyso_srs.mat
....
....
eof-ipmos
```

1.3 Allowed detector types

The type of detector is specified either by the **DETECTOR** keyword, or by a **SITE** keyword, with the latter generally being used for synchrotron sites that use detectors that are not commercially available. At present, the follow detectors are allowed.

1.3.1 Using the **DETECTOR** keyword

- Mar Research (18cm (**SMALLMAR**), 30cm or 34.5 cm plate (**MAR**));
- Mar Research CCD detectors (both 135 and 165 mm) as well as the Mar Mosaic detector (**MARCCD**);
- ADSC Quantum 4, Q210 and Q315 CCD detectors (**ADSC**);
- Mac Science Dip2000 (**DIP2000**), 2020 (**DIP2020**), 2030 (**DIP2030**) or 2040 (**DIP2040**) (horizontal or vertical rotation axis);
- R-Axis II (**RAXIS**) (horizontal or vertical rotation axis),
- R-Axis IV (**RAXIS4** or **RAXISIV**) (Horizontal or vertical rotation axis),
- R-Axis V (**RAXIS5** or **RAXISV**),
- Rigaku Jupiter, Saturn, Mercury CCD detectors,
- ESRF Image intensifier CCD (**ESRF CCD**),
- Molecular dynamics (offline) (**MD**),
- FUJI scanners (offline) (**FUJI BAS2000** or **FUJI BA100**),
- Ed Westbrook or Oxford Instruments 3x3 CCD detector (**SBC1**),
- Brandeis 2x2 CCD detector (**B4**) (**ADSC**),
- ESRF Large Image Plate Scanner (**LIPS**),

- There is limited support for the Oxford Sapphire CCD and Bruker SMART and Proteum detectors.

From Mosflm version 6.2.3, most of the commercial detectors listed above are recognized automatically by Mosflm, so the **DETECTOR** keyword is no longer required in many cases.

Note that no special input is required to distinguish between the different types of Mar Research image plate scanner (18,30 or 34.5cm (Mar345)). The image size is read from the header record and the appropriate limits and pixel size are set up automatically.

Both unpacked and packed image formats are supported for the Mar345 scanners (no **DETECTOR** keyword required to distinguish these).

For Mar, R-Axis and Mac Science scanners it is not necessary to specify the size of the image, as it is determined from the image header.

For offline scanners (FUJI and MD) it will also be necessary to define the orientation of the image relative to the X-ray beam and rotation axis, also using the **DETECTOR** keyword. See the help library (Subsection Novel detectors of **DETECTOR**) for details on how to do this.

If the rotation axis is reversed (usually a peculiarity of synchrotron sites) this can be dealt with by specifying: **DETECTOR REVERSEPHI** ..again, see the help library.

1.3.2 Using the **SITE** keyword

For CHESS, the station (A1, F1 or F2) and the detector must be specified. Possible detectors are the Gruner CCD detector working in 1K, 2K or 2K binned modes, the ADSC single module CCD detector (**ADSC**) the ADSC 2x2 CCD detector (**QUANTUM4**) and FUJI image plates. *eg*

```
SITE CHESS [A1 F1 F2] [FUJI [CCD [1K 2K 2KBINNED ADSC QUANTUM4]]]
```

For SSRL and ALS, the 2x2 ADSC detector is allowed: *eg*

```
SITE SSRL ADSC, SITE ALS ADSC
```

1.4 Inspection of images, invoking the new menu

It cannot be emphasised strongly enough that images must be examined closely to check the following:

- 1) Does the crystal diffract ?
- 2) What is the effective resolution limit...should the detector be moved further back to take advantage of the full active area of the detector?
- 3) Is the crystal twinned, split, disordered etc ?
- 4) Is the exposure time long enough ?

Images can be displayed using the **IMAGE** keyword followed by the full filename of the image (including the directory if the image is not in the current directory, but it is recommended that the **DIRECTORY** is set

explicitly). The only other keyword required specifies the type of detector (default is Mar Research image plate scanners). This will bring up the new menu interface which allows autoindexing, integration etc.

Note that MOSFLM displays the image viewed from the detector looking towards the source (cameraman's view), and also that the "fast changing" direction in the image is *ALWAYS* vertical in the display, regardless of whether it is vertical or horizontal in the actual detector. Thus some images will be rotated by 90 degrees.

example

```
DIRECTORY /fred/images
IMAGE lysox1_001.image
GO
```

or

```
IMAGE /fred/images/lysox1_001.image
GO
```

or

```
DETECTOR RAXISIV
DIRECTORY /fred/images
IMAGE lysox1_001.image
GO
```

In order to measure the resolution of individual spots on the images or display the resolution circles the wavelength and crystal to detector distance need to be given. For Mar Research, ADSC, R-AxisIV and Mac Science detectors the wavelength and distance will automatically be taken from the header records in the image file, for other types of image the **DISTANCE** and **WAVELENGTH** keywords should be given, or the values set interactively using the X-windows interface.

Parameters that may need to be defined (and the appropriate keywords) are:

- 1) crystal to detector distance (**DISTANCE**)
- 2) Wavelength (**WAVE**)
- 3) Direct beam coordinates (**BEAM**)

1.4.1 Writing JPEGs

JPEG files can be written for diffraction images read by MOSFLM using code which was originally written for a new Graphical User Interface. This can be accessed from the command-line or a batch file thus;

```
XGUI ON
GO
CREATE_IMAGE BINARY TRUE FILENAME <filename>
RETURN
EXIT
```

This may be useful for creating movies of images from a data collection or allowing third parties to view copies of the images.

If an orientation matrix has already been determined, the "CREATE_IMAGE" line can also include "PREDICTION TRUE" in order to display predicted positions of spots.

1.5 EXAMPLE INPUT

NOTE: Input within square brackets is optional for Mar, ADSC, R-AxisIV and Mac Science images. If a PHI keyword is given, this will override the phi values in the image header, and phi values in the header will be ignored for any subsequent image that is read in. The default DETECTOR type is MAR, so this need not be given for Mar images.

```
IMAGE catx1_001.img [PHI 0.0 TO 1.0]
BEAM 149.5 151.0
DETECTOR MAR (or SMALLMAR, MARCCD, ADSC, RAXIS, RAXISIV, SBC1, DIP2000,
              DIP2030, DIP2040, ESRF CCD, FUJI, MD)

[NEWMAT test_001.mat]    ! Defines the name of the file in which the results
                        ! of autoindexing or postrefinement will be written.
[WAVELENGTH 1.542]
[DISTANCE 250.0]
GO
```

This will invoke the X-window display, and a Menu list as shown below:

Read image	Read in another image.
Find spots	Find spots on the current (displayed) image.
Edit spots	Allows manual rejection of spots.
Clear spots	Deletes spots from display or list of stored spots.
Select images	If spots have been found on several images, allows selection of images to be used in autoindexing.
Autoindex	Invokes autoindexing (DPS or REFIX).
Estimate mosaicity	Gets an initial estimate of mosaic spread.
Predict	Predicts spot pattern.
Clear prediction	Deletes predicted pattern from display.
Adjust	Adjust the fit between observed and predicted patterns.
Refine cell	Invokes a POSTREF SEGMENT run to refine cell parameters.
Integrate	Allows integration of images.
Strategy	Run the strategy option.
Keyword input	Allows keyworded input.
Find hkl	Allows a specified reflection to be identified.
Pick	Display pixel values
Measure cell	Measure cell parameters.
Circles	Display resolution circles.
Beam/ backstop	Allows interactive definition of beamstop shadow.
Exit	Close down X-windows display.

The various options in this menu list are discussed in the sections below. Note that there are some on/off and yes/no toggle boxes at the bottom of the "Processing parameters" window. These are described below:

Prompts On/Off (default on)

When "prompts" is "on", additional information is given when some of the menu options are chosen. For experienced users, this additional information can be suppressed by turning the prompts "off".

Update display:

After refinement	No/Yes
After integration	No/Yes

By default, the display is updated each time a new image is read, and at no other time. By setting the "After refinement" toggle to "Yes", the display will be updated after refinement of the detector parameters, so that it is possible to check how well the predicted pattern matches the image. If the "After integration" toggle is set to "Yes", each image will be display after it has been integrated, with "Bad spots" indicated and residual vectors (between observed and predicted spot positions) for fully recorded spots also shown. It is possible to reject additional reflections, or reclassify Bad spots, at this point.

Note that because images are integrated in "Blocks", during the actual integration of all images in a block the image that is displayed will be that of the last image in the block, unless the "After integration" toggle has been set to yes.

Timeout mode: Off/On

The Timeout mode is now set to "On" during an Integration or Refine Cell run, so that when each image is displayed the program will wait for 1 second for the user to select a menu option (it is best to start by turning the Timeout mode off if you want to do this). After this period (which can be changed with keyword "TIMEOUT") the program will just carry on. With the timeout mode "On" it is therefore possible to integrate a series of images without directly interacting with the program. This can be very useful if one just wants to keep an eye on the processing but do not want to keep hitting the "Continue" menu option.

[Back to Contents Page](#)

2: A Quick Guide

- 2.1 [Startup keywords](#)
- 2.2 [Masking "bad" regions of the detector or setting the direct beam co-ordinates](#)
- 2.3 [Autoindexing](#)
- 2.3.1 [DPS autoindexing](#)
- 2.3.1.1 [Unknown symmetry](#)
- 2.3.1.2 [Space group information already given](#)
- 2.3.2 [REFIX autoindexing](#)
- 2.3.2.1 [Unknown symmetry](#)
- 2.3.2.2 [Space group information already given](#)
- 2.4 [Estimating mosaic spread](#)
- 2.5 [Running the Strategy option](#)
- 2.6 [Determining oscillation angles with the TESTGEN option](#)
- 2.7 [Integrating the first image to determine if the exposure time is OK](#)
- 2.8 [Interpreting those "WARNING" messages](#)
- 2.9 [Getting accurate cell parameters](#)
- 2.10 [Integrating a block of images](#)

2.11 Integrating the dataset

This is a brief guide on how to process data using the new menu options. For more details on each step in the procedure, see sections 3-6.

2.1 Startup keywords

Use the following keywords to bring up the image display and menu. These should be given at the MOSFLM => prompt:

```
=> TITLE My lysozyme data           ! This title is transferred to the
                                     ! MTZ file

=> IMAGE lyso_001.image [PHI 0 TO 1] ! Filename of first image. For Mar,
                                     ! ADSC, Mac Science and R-AxisIV images
                                     ! the phi values will be taken from the
                                     ! image header if not given here.
                                     ! If phi values are specified here,
                                     ! the values in the header will be
                                     ! ignored for this an all subsequent
                                     ! images read in.

=> BEAM 150.0 149.0                 ! Direct beam coordinates
```

If not processing Mar Research, ADSC, R-axis or Mac Science images:

```
=> WAVE 0.91                         ! For most modern detectors,
=> DISTANCE 300                     ! this information is taken from the he
                                     ! but can be overwritten using the
                                     ! keywords.

=> SYMM p43212                       ! If known, give cell and symmetry
=> CELL 79 79 38                     ! otherwise omit completely.
```

Not essential for first stages, but needed for integration:

```
=> DIVERGENCE 0.1 0.03              ! If isotropic, the beam divergence
                                     ! can be included in the mosaic spread.
=> SYNCHROTRON POLARIZATION 0.9     ! Defaults to 0.95 (SRS, Daresbury UK)
=> GAIN 1.7                          ! See section 8.1 for a way to
                                     ! estimate the gain if not known.

=> GO
```

At this point, the image will be displayed with a list of "Processing parameters" on the far left (these can be changed by the user), a "Main menu" and beneath the Main menu a table of "Output" parameters.

2.2 Masking "bad" regions of the detector or setting the direct beam co-ordinates

Parts of the detector which are obscured, *e.g.* by a Cryostream head or a backstop arm can be masked off using the options obtained by pressing the "**Beam/Mask areas**" button.

The beam can be set either by clicking on a single point (*e.g.* if the backstop is "leaky") or by clicking around points in a circle (*e.g.* on an ice ring).

A circular backstop can be masked by clicking on several points around its periphery.

Different shaped areas can be omitted from spot finding (and hence indexing), refinement and integration. These include rectangular areas (aligned with edges horizontal and vertical), general convex quadrilaterals, and regions defined by arcs (as either the internal or the external boundary).

Circular regions (*e.g.* ice or other powder rings) can be defined by clicking on their inside and outside edges.

All excluded areas are marked on the image with red lines.

2.3 Autoindexing

Select the menu option "**Autoindex**". The program will locate and display spots on the image. Parameters governing the spot finding are listed under ***SPOT SEARCH*** in the Processing Parameters table, but the program automatically sets suitable values for these parameters and they will not normally have to be changed. The user is asked if they want to try the new autoindexing. The new autoindexing is the FFT based DPS indexing, which is usually more successful, so this should be the first choice. The alternative is the REFIX style autoindexing.

If no spacegroup information has been given, for both algorithms the user will be presented with a list of choices, sorted on a "PENALTY" parameter (the lower the PENALTY the better). The user must select a spacegroup, and the cell is refined imposing that symmetry.

When using the new DPS indexing, if a spacegroup and cell have been given, the cell parameters determined by the autoindexing will be permuted to best match the input values, but the user must still select the solution from the list provided. If using REFIX, the same information will force the image to be autoindexed with this cell and no alternatives will be listed.

The success of the autoindexing can be checked by predicting the spots for the current image using the menu option "**Predict**". If not successful, try adjusting the intensity threshold "**Min I/sig(I)**" or the maximum cell length (for the FFT based algorithm) or read in another image (**Read image** menu option), find spots on it (**Find spots**) and repeat autoindexing (**Autoindex**). Spots from a satellite crystal can be removed using the **Edit spots** option.

2.3.1 DPS autoindexing

2.3.1.1 unknown symmetry

A number of questions follow; in most cases the default answer should be accepted, since it will have been chosen to provide a reliable answer in most cases.

The program first asks if the detector distance is to be fixed. The default is "yes", since the distance refinement can be unstable if a poorly defined solution is chosen.

Then the program asks if spots close to ice-ring positions should be excluded; unless there are obvious ice-rings, the default answer ("no") should be chosen. Next, the user will be prompted to supply a filename for the output orientation matrix. Next, the user will be asked to enter a value for the maximum expected cell edge; the program makes an estimate of this, based on the current spot list and the current detector parameters. If the value given here is significantly (>25%) less than the true largest cell edge, the indexing will fail. Equally, if the true cell is VERY much smaller than the default value, it may also select an incorrect cell, with one or more parameters a multiple of the true cell. The default value will work in the great majority of cases.

From version 6.2.5 onwards, there are some optional questions, based on whether the user wants extra output to help decide between similar solutions. First, the user is asked whether they want to try enhanced solution picking (default "no"); if the user replies "yes", the following four questions are asked. Usually, the best option is to take the default answer in each case.

- *Do you want to pre-refine the solutions (N):* If this is chosen, each solution with a penalty ≤ 200 will be refined after the appropriate symmetry has been imposed. This makes the indexing take somewhat longer but may help in difficult cases.
- *Do you want to print the distortion index? (Y):* This is a quality indicator based on that used in Denzo; it is calculated in a different way to the normal Mosflm penalty, so it may give more information about the quality of a solution.
- *Do you want to prerefine the triclinic solution? (Y):* This option refines the triclinic solution *before* applying symmetry. In some cases (particularly with long cell edges), it can help indexing succeed.
- *Do you want to specify a minimum celledge? (N):* If there are very few spots found, the Fourier Transform can have peaks at very low values; this option can help to filter these out.

The program then reports the number of reflections which will be used for indexing and the sigma level chosen (20, by default), and asks if the user wants to proceed.

2.3.1.2 Space group information already given

If space group information has already been supplied to Mosflm, this is reported in the top line of the indexing dialogue box. Consequently, the dialogue changes. After choosing DPS indexing, the user is asked if they want to change the space group to 0 (*i.e.* undefined); the default is "no". Then the user is asked if they want to fix the cell (*i.e.* impose the current cell edges; the default is to allow similar dimensions) and the detector distance, and to exclude spots close to ice-ring positions, and for the name of a matrix file.

Finally, the current cell and space group is printed and the user asked if they want to proceed. When the window is refreshed, the target cell is printed with the nearest found cell and the user asked if they want to accept the result. If they don't, the dialogue reverts to that for the "unknown cell" option.

2.3.2 REFIX autoindexing

2.3.2.1 unknown symmetry

In this case, the user is only asked if they want to fix the detector distance, what the name of the output matrix file should be and if they want to proceed.

2.3.2.2 Space group information already given

The same questions as for DPS indexing (above) are asked.

2.4 Estimate mosaic spread

An estimate of the mosaic spread can be obtained by choosing the menu option "**Estimate mosaicity**". This works quite well for mosaic spread up to about 0.6 degrees. The resulting value should always be checked by predicting the pattern and seeing if all observed spots are predicted.

2.5 Run Strategy option

Select the "**Strategy**" menu option. Input for this option has to be given at the MOSFLM => prompt in the terminal window.

```
MOSFLM => STRATEGY
MOSFLM => GO
```

This will generate a reflection list, a unique reflections list, merge them and tell you what rotation range to use to get a maximally complete dataset.

If you then want to reduce the total rotation range (to save time) and still get a maximally complete dataset type the following at the STRATEGY => prompt:

```
STRATEGY => ROTATE 60 SEGMENTS 2
STRATEGY => GO
```

This instructs the program to find two 30 degree segments that give maximum completeness. You can try 3 segments (of 20 degrees) if you like, but this rarely (in my experience) gives significantly greater completeness and will take significantly longer. (Also don't forget that the more segments you have, the more unmatched partials you will get).

For orthorhombic space groups, you should also try STRATEGY ALTERNATE if the predicted completeness is not as high as expected.

2.6 Determine oscillation angles with the TESTGEN option

Having determined what rotation range needs to be collected, you can check what the (maximum) rotation angle is to avoid getting (too many) spatial overlaps on the images. Remember you must have realistic estimates of the mosaic spread and minimum spot separation for this to be meaningful.

At the STRATEGY => prompt type:

```
STRATEGY => TESTGEN
```

This will describe the possible keywords. If your data collection was in two segments of -15 to 15 degrees and 45 to 75 degrees and you want no overlaps type :

```
STRATEGY => TESTGEN START -15 END 15
```

```
STRATEGY => G0
```

and the program will calculate the MAXIMUM possible rotation angles for this range, at intervals of 5 degrees.

then type:

```
STRATEGY => TESTGEN START 45 END 75
STRATEGY => G0
```

for the second segment.

To test for overlaps using a particular oscillation angle (*e.g.* 1.5 degrees) type:

```
STRATEGY => TESTGEN START 45 END 75 ANGLE 1.5
STRATEGY => G0
```

2.7 Integrate the first image to determine if the exposure time is OK

The best way to get an indication of the data quality is to integrate the first image and see how the mean $\langle I \rangle / \sigma \langle I \rangle$ varies with resolution. These values are always slightly optimistic, so you should aim to have a ratio of at least 3.0 in the outermost resolution bin. If it is lower than this, consider collecting data to a lower resolution (and moving the detector further back) or using a longer exposure time.

To get back to the Menu, type EXIT at the STRATEGY prompt:

```
STRATEGY => EXIT
```

First, set the centre and radius of the backstop shadow, by selecting the "**Beam/backstop**" menu option. Click with the mouse around the backstop shadow. The program will fit the best circle to these points. If the fit looks OK, the backstop centre and radius are updated. This can also be used to update the direct beam position, but it is rare that the beamstop shadow is accurately centred on the direct beam position !

An alternative way of dealing with backstop shadows (particularly for an extended backstop) is to use the NULLPIX keyword. This is used to define a minimum pixel value, and any spot that has a pixel within its measurement box with a value lower than (or equal to) this minimum will be rejected. Be sure that the value given is *not* BIGGER than the background at the edge of the image, or all the high resolution data will be rejected !

Then select the "**Integrate**" menu option, and answer the questions (most can be answered by entering carriage return).

The program will put up the predicted pattern and then wait for 1 second before continuing. If the pattern is not a good fit, set the timeout mode "Off". Choose menu option "**Continue**". If the pattern is not aligned with the spots, choose the option "**Adjust**" and follow the instructions to align the pattern with the spots. (The usual reason for poor alignment is an error in the direct beam coordinates, but as these are refined as part of the autoindexing this should not normally be a problem).

The image will then be integrated. Check the $\langle I \rangle / \sigma \langle I \rangle$ values (either in the terminal window or in the "mosflm.lp" file), and the Rsym if you have symmetry related fully recorded reflections on this image. The

value of **SDRATIO** is a better guide than the actual **Rsym**, as the latter will depend on the intensity of the reflections. The **SDRATIO** should lie in the range 1 to 3.

2.8 Interpreting those "WARNING" messages

After the integration, the program will usually print a list of "WARNING" messages to the terminal window (or `mosflm.lp`). Don't worry about messages about the standard profiles at this stage, or large positional residuals (because the cell has not yet been accurately determined). However if the " **OVERALL BACKGROUND RATIO (BGRATIO)** " message is present, this suggests the detector **GAIN** may be wrong, and the input value (or default value if not input) needs to be multiplied by the square of the value of the **BGRATIO** (get the default value from the `mosflm.lp` file, where all parameters are printed prior to the integration step). Beware, however, that images showing diffuse scatter will give a high **BGRATIO** even when the **GAIN** is correct (*e.g.* up to 1.5).

2.9 Getting accurate cell parameters

Accurate cell parameters are essential to obtain the best data quality. MOSFLM uses a post-refinement procedure to determine accurate cell parameters. For trigonal or higher symmetry, an accurate cell can usually be determined from a single "wedge" of data (typically 3-5 degrees), unless the unique axis is approximately along the X-ray beam direction in which case either a different **phi** value should be used or two segments of data. For orthorhombic or lower symmetry two "wedges" of data widely separated in **phi** will give the best results. In the latter case, one can either wait until a large rotation range has been collected before refining the cell, or one can start by collecting a few degrees at (say) **phi** 85 to 90, then start collecting from **phi** = 0.

When the appropriate data (images) are available, select the "**Refine cell**" menu option. Answer the queries. Although the default number of images to use is 2 (in each wedge), this is in fact the minimum number and better results will often be obtained by using 3 or 4 images in each wedge.

It is important to have a realistic estimate of the mosaic spread before refining the cell.

Post-refinement yields very accurate cell parameters, but has a relatively small radius of convergence. If the shift in cell parameters is more than 2.5 times the estimated error, the integration of the images and the actual refinement will be repeated. This will happen up to a maximum of 5 times. It is not unusual for 2 or 3 complete rounds to be required if the initial cell parameter estimates came from auto-indexing a single image.

2.10 Integrating a block of images

The next step is normally to integrate a block of between 5 and 10 images. Use the "**Integrate**" menu option as before. In this case, pay particular attention to the list of warning messages to see if any parameters or options need to be reset. It is also a good idea to check the appearance of the standard profiles (these are output to the terminal window but also to the file "`mosflm.lp`"). Make sure that adjacent spots are being adequately resolved, and that the peak is not spilling into those pixels marked as background. The **PROFILE TOLERANCE** parameters are crucial in determining the appearance of the standard profiles. Also try to ensure that *NO* profiles are being averaged. If necessary, change the minimum rms variation in the background (**PROFILE RMSBG**) or the number of different profiles (by defining **PROFILE XLINES** and **PROFILE YLINES**) to avoid profile averaging. Check that there are not too many reflections being rejected as "BAD SPOTS". If a significant number of strong reflections are being rejected for "Poor profile fit", if the cell has been accurately determined, the mosaic spread appears correct and the **GAIN** is correct, consider increasing the rejection value (**REJECTION PKRATIO**) from its default of 3.5 to 4.0. It should *NOT* be necessary to increase this above 4.0, as rejection of the strongest reflections may have a serious effect on structure

determination.

2.11 Integrating the dataset

Once a block of images has been successfully integrated, the complete dataset can be integrated. If data processing is started before data collection is complete, use the **WAIT** keyword to make the program wait for an image to be completed before it tries to integrate it.

*e.g. WAIT 15 for 15 minute exposures,
or WAIT 0 12 for 12 second exposures,
or WAIT 1 17 5 for 1 minute 17 second exposures, and waits for another 5 seconds once the file is there to make sure it is fully written.*

You may also wish to specify multiple MTZ files (one for each "block" of images) so that some data can be scaled and merged in SCALA before data collection/processing has completely finished (HKLOUT MULTIPLE option).

Set the "Prompts" toggle to Off.

[Back to Contents Page](#)

3: Determination of crystal orientation, cell parameters and spacegroup

- 3.1 [Autoindexing Interactively](#)
 - 3.2 [Autoindexing when running the program in background](#)
 - 3.2.1 [REFIX and general notes](#)
 - 3.2.2 [DPS Indexing in background](#)
 - 3.2.2.1 [Autoindexing using different images from the same crystal](#)
-

The crystal orientation, cell parameters and possible spacegroups are normally determined from a single rotation image (although 2, 3 or more can be used also). This will typically be a rotation of between 0.5 and 2 degrees, the value being chosen to avoid generating a significant number of spatially overlapped reflections. Autoindexing can be performed interactively or in background mode; spacegroup selection can **ONLY** be done with REFIX indexing interactively as the user is required to select a cell and spacegroup from a number of possibilities, but the DPS indexing can pick a cell and spacegroup itself. However, users should be aware that this can sometimes be unreliable and they should check the results carefully themselves.

3.1 Autoindexing Interactively

3.1.1 Finding Spots

The first step in autoindexing an image is to locate the positions of diffraction spots. This can be done with the "**Find spots**" menu option, but if only one image is to be used for autoindexing one can go straight to the "**Autoindexing**" menu option.

3.1.1.1 Parameters used in Finding Spots

Parameters associated with spotfinding are listed in the "Processing parameters" window:

```

Threshold
Rmin
Rmax
X offset
Y offset
Min X size
Max X size
Min Y size
Max Y size
Min no of pix
X splitting
Y splitting

```

All of these parameters have "sensible" defaults and normally they do not need to be changed.

Pixels are considered to be part of a spot if the pixel value is more than $(\text{Threshold} \times \sigma)$ above the local background at that radius. The threshold is determined automatically by the program and will normally be appropriate, but for images with very low background (less than 20) it may be necessary to increase the threshold.

The program searches for spots lying with radial limits of Rmin and Rmax (mm) from the direct beam position. The first step is to determine a radial background. The direction of this radial background is chosen to be at right angles to the rotation axis (to avoid any backstop shadow). It is normally centred on the direct beam position, but can be offset to one side by the "X offset" or "Y offset" parameters. If the radial background is along Y (the "fast" changing direction in the stored image), then use the "X offset" to change its position. If the default direction is along Y, and a value is entered for the "Y offset", this automatically changes the direction of the background strip to be along the X axis. Entering a negative value for either Rmin or Rmax will switch the background strip to the opposite side of the direct beam position.

For tiled detectors with an even number of tiles, the background strip is offset automatically from the blank region between the tiles.

The position of the background strip is shown as a red rectangle on the display. If necessary its position should be changed to avoid any shadows or other unusual features on the image.

Note that the **LIMITS EXCLUDE** keywords can be used to exclude rectangular regions of the detector from spot finding (and integration).

The minimum and maximum spots sizes (in X and Y) are expressed as a multiple of the median spot size. If the image is very strong and the threshold is too low, then two adjacent strong spots may be treated as a single spot (because the pixel values do not go down to the threshold in between them). This problem can be avoided by either increasing the Threshold, or by decreasing Max X and Y sizes, as these spots will be almost twice as large as the average spot.

"Min no of pix" sets the minimum number of pixels that constitute a proper spot.

Split spots will be treated as a single spot if they are less than "X splitting" and "Y splitting" mm apart in X and Y. Normally this is not a problem with image plate data.

In tricky cases (*e.g.* very weak spots on a high background) it may be necessary to add spots manually. The program asks if you want to find spots manually both before and after the automatic spot search. This is done by clicking on spots with the mouse. A red cross will be drawn for each spot found. You do not need to position the mouse exactly on the spot, the program will search the area around the mouse position and find the centre of gravity which will be used as the spot position. Manually selected spots are assigned a value of 1000 for $I/\sigma(I)$. Select the **"End add spots"** menu option to finish.

3.1.1.2 Displaying found spots

The positions of the spots found are displayed as red crosses. Note however that **ONLY** spots which will be used for autoindexing (*i.e.* those with $I/\sigma(I)$ greater than the threshold) are displayed. This is determined by the only parameter associated with autoindexing, the "Min $I/\sigma(I)$ " parameter which follows the spot finding parameters in the "Processing parameters" window. This value defaults to 20.

It is a good idea to check that program is correctly locating the spots, and that in particular if the spots are very close and the image is strong, it is not treating two neighbouring strong spots as a single spot (in which case the red cross will come half-way between the two spots). If this is a problem, try increasing the "Threshold" or decreasing the "Max X size" and "Max Y size".

3.1.1.3 Editing found spots

If the crystal is not single, and the program finds spots that do not lie on the major lattice, it is a good idea to remove these spots. If the second lattice is much weaker than the main lattice, it may be possible to do this just by increasing the "Min $I/\sigma(I)$ " parameter. If this does not work, select the **"Edit spots"** option from the main menu. Identify spots to be deleted by clicking on them with the mouse...this will result in an "X" being written over that spot. Be careful, as the mouse must be quite close to the spot position in order to reject the spot. When editing is finished, click on the **"End edit"** in the main menu.

The autoindexing algorithm is quite sensitive to the presence of "rogue" spots, so it is usually a good idea to reject them if the autoindexing is not successful.

3.1.2 Finding spots on other images

If you want to use more than a single image for the autoindexing (and this can provide successful autoindexing when using a single image fails) then read in another image using the **"Read image"** option in the Main Menu. The phi values will be read from the header (if they were not given on the original **IMAGE** keyword) or set automatically, assuming the image is part of a contiguous series, but the phi limits can be reset.

Then choose the **"Find spots"** option as described above. Note that there is a limit to the total number of spots that can be stored internally, which may place a limit on how many images can be used. Raising the spot finding "Threshold" will reduce the number of spots found if this causes problems. Note also that the REFIX autoindexing algorithm itself will use a maximum of 2000 spots, while the DPS indexing can use up to 5000.

IMPORTANT All found spots are stored and will be used in autoindexing. If changing to a new crystal, spots found previously *MUST* be deleted using the **"Clear spots"** menu option.

3.1.3 Selecting images for autoindexing

If spots have been found on several images, then by default all of these spots will be used for autoindexing. If, however, you only want to use the spots from selected images, use the "**Select images**" menu option. The spots found on each image are stored in a separate "slot", and the "slot" numbers (rather than the image numbers) must be given when selecting the images (so that images with the same image number can be used). If you wish to make a fresh start, use the "**Clear spots**" menu option.

3.1.4 Running the autoindexing

Autoindexing uses either the FFT based algorithm from DPS (Steller, Bolotovskiy and Rossmann, (1998) J. Appl. Cryst. 30, 1036-1040) or Wolfgang Kabsch's REFIX program (Kabsch, 1988, 1993) both of which have been incorporated into the MOSFLM program.

Autoindexing is performed by selecting the "**Autoindexing**" menu item. For DPS indexing or REFIX indexing if the spacegroup is not known (or set to zero) the program will present a list of possible unit cells and space groups, sorted on the PENALTY of each solution, and the user has to select the appropriate choice. (When using REFIX, if a crystal symmetry AND unit cell have been applied then only solutions for this symmetry will be listed).

In this list, the first number is the number for that solution, the second number is a score for that solution (headed "PENALTY"). This is followed by the lattice type, the cell parameters and a list of possible spacegroups. Normally one would choose the solution with the highest possible symmetry, but which still has a reasonably low "PENALTY" (The LOWER the PENALTY the better).

example:

18	150	cI	103.13	103.36	103.01	62.8	62.8	62.9	I23, I213, I432, I4132
17	64	tP	74.44	74.54	74.56	92.6	92.2	92.4	P4, P41, P42, P43, P422, P4212, P4122, P41212, P4222, P42212, P4322, P4321
16	63	oP	74.44	74.54	74.56	92.6	92.2	92.4	P222, P2221, P21212, P212121
15	63	tP	74.54	74.56	74.44	92.2	92.4	92.6	P4, P41, P42, P43, P422, P4212, P4122, P41212, P4222, P42212, P4322, P4321
14	62	hR	107.32	103.13	131.17	92.0	89.8	121.4	H3, H32 (hexagonal settings of R3 and
13	41	oC	103.01	107.80	74.44	90.2	93.3	90.0	C222, C2221
12	41	mP	74.54	74.44	74.56	92.2	92.6	92.4	P2, P21
11	41	mP	74.56	74.44	74.54	92.4	92.6	92.2	P2, P21
10	40	oC	103.01	107.80	74.44	89.8	93.3	90.0	C222, C2221
9	40	mP	74.54	74.44	74.56	92.2	92.6	92.4	P2, P21
8	26	cP	74.54	74.56	74.44	92.2	92.4	92.6	P23, P213, P432, P4232, P4332, P4132
7	24	mC	103.36	107.32	74.54	89.8	93.6	90.1	C2
6	22	mC	103.36	107.32	74.54	89.8	93.6	90.1	C2

5	19	aP	74.44	74.54	74.56	87.4	92.2	87.6	P1
4	6	hR	107.32	107.52	123.58	89.9	90.0	119.8	H3,H32 (hexagonal settings of R3 and
3	4	mC	103.36	107.32	74.54	90.2	93.6	89.9	C2
2	2	mC	103.01	107.80	74.44	89.8	93.3	90.0	C2
1	0	aP	74.44	74.54	74.56	92.6	92.2	92.4	P1

In this case H3 or H32 is an obvious choice.

If the direct beam coordinates are inaccurate (or the detector distance or wavelength) there may not be a clear separation between solutions with low penalties and those with much higher penalties.

Once you have made a selection the autoindexing is repeated automatically and the cell is refined imposing the appropriate cell constraints. Reflections whose calculated position differs by more than 2.5 standard deviations from the observed one are rejected from the refinement, but this cutoff can be changed. You are given the choice of accepting or rejecting the refined cell parameters and the refined direct beam position. Finally, you are then given the choice of accepting that solution or trying another one from the list.

REMEMBER that the true spacegroup can only be determined from reflection intensities, NOT from unit cell parameters.

BEWARE of monoclinic spacegroups with beta angles close to 90 being misclassified as orthorhombic etc.

The final orientation (A matrix) and cell parameters are written to a file, which can be defined when the autoindexing procedure is initiated, or with the keyword NEWMAT (defaults to NEWMAT).

The file can be read in (MATRIX keyword) in future processing jobs.

If you want to change (permute) the order of the cell axes, simply include a CELL keyword giving the cell that you would like to fit. For example, in orthorhombic space groups the autoindexing will select the cell with $a < b < c$. If the cell dimensions are 50, 100, 150 and you want to have $a=150$, $b=50$, $c=100$ give the keyword:

```
cell 150 50 100
```

before running the autoindexing.

3.1.4.1 How do you know if it has worked correctly?

The most obvious test of the success of the autoindexing is to predict the pattern using the "**Predict**" menu option and see if it matches the observed pattern.

If there was a large error in the input direct beam coordinates, with the REFIX autoindexing this is sometimes apparent in a shift of the predicted pattern relative to the observed spots. This shift can be corrected using the "**Adjust**" menu option. With the DPS indexing, the direct beam coordinates are automatically updated, so it should not be necessary to "**Adjust**" the pattern. If the shift is significant, it is probably worth repeating the autoindexing with the updated direct beam parameters (they are updated automatically by using "**Adjust**") as this will give more accurate cell parameters.

The single most important number by which to judge whether autoindexing has succeeded is the positional residual (standard deviation of spot position). This value should be below 0.2-0.3mm. If it is above 0.3mm the solution is highly suspect, and if above 0.4mm it is almost certainly wrong. Values of 0.08mm to 0.12mm are typical for a correct solution (Note that the positional residual will depend on the size of the diffraction spots. The values given here are for a spot size of about 6x6 pixels with a pixel size of 0.15mm, larger spots will give slightly larger residuals).

3.1.4.2 Errors reported by the autoindexing

Possible options if the autoindexing fails are:

- 1) Make sure the direct beam coordinates are correct! (The autoindexing is quite sensitive to these). If necessary, record a powder pattern (*e.g.* from bee's wax or paraffin wax) and display this image and work out the coordinates of the centre of the rings.
- 2) Try changing the intensity threshold "Min(I)/sig(I)" in the "Processing Parameters" menu (up or down).
- 3) For the new DPS indexing, change the maximum cell edge.
- 3) Include data from other images (this could also give a more accurate cell).
- 4) Try to avoid images looking down a principal zone.

3.2 Autoindexing when running the program in background

3.2.1 REFIX and general notes

(Note that as far as MOSFLM is concerned, a job that directs output to a file rather than to a terminal is considered to be a background or batch job).

The REFIX autoindexing can be used to autoindex in a background job if the spacegroup and cell are known; an approximate cell and the spacegroup must be given.

Autoindexing is invoked by including the keyword AUTOINDEX. If no images are specified, the first image to be integrated (specified on the PROCESS keyword) will be used for autoindexing.

Thus:

```
CELL 107 107 123 90 90 120
SYMM H32
AUTOINDEX
PROCESS 1 TO 30 [ ANGLE 1.0 START 0.0 ]
```

will autoindex using image 1 and then integrate images 1 to 30.

Note that the cell derived from the autoindexing, rather than that given by the CELL keyword, will be used during integration. If the cell is known accurately it is usually better to override the cell derived from autoindexing by using the KEEP keyword:

```
CELL KEEP 107.73 107.73 123.59 90 90 120
```

If you want to include more than one image in the autoindexing they can be specified explicitly:

```
AUTOINDEX IMAGES 1 2 3
```

will use the first three images. In this case it is assumed that the phi values are read from the image header, or that these images form part of a contiguous rotation in phi. If this is not the case, the phi values can be specified explicitly:

```
AUTOINDEX IMAGE 1 PHI 0 1 IMAGE 20 PHI 50 51
```

If the image identifier (used to form the template for the image filename) is not the same for all images, it can also be specified explicitly:

```
AUTOINDEX IMAGE 1 PHI 0 1 IMAGE 20 PHI 50 51 IDENT test_2
```

Note that if PHI or IDENT are given, then only ONE image can be specified on each IMAGE keyword so that:

```
AUTOINDEX IMAGE 1 PHI 0 1 IMAGE 20 21 IDENT test_2 is NOT allowed
```

The "Min I/sig(I)" threshold can be set also:

```
AUTOINDEX THRESHOLD 30
```

Parameters associated with spot finding can be set with the FINDSPOTS keyword:

e.g.:

```
FINDSPOTS THRESHOLD 10 RMIN 25 RMAX 75 SPLIT 0.5 0.5
FINDSPOTS MINX 0.5 MAXX 1.5 MINY 0.5 MAXY 1.5 XOFFSET 25
```

3.2.2 DPS Indexing in background

The DPS autoindexing can be used to autoindex in a background job whether or not the spacegroup and/or cell are known. The algorithm used, however, will choose the highest symmetry characteristic lattice consistent with the solution unless instructed otherwise. If the image header does *not* contain oscillation angle information, this must be supplied on the command line. Thus:

```
AUTOINDEX DPS IMAGE n PHI 0 1
```

Will autoindex using image *n* and assign the space group to the lowest symmetry space group available for the characteristic lattice with the highest symmetry for an acceptable solution, *e.g.* if the lattice is primitive orthorhombic, the space group will be P222. Adding space group information, *e.g.* with the following command line:

```
SYMM P212121
```

forces the program to accept *not only* the primitive orthorhombic solution but also sets the space group to P212121. If the cell has been given, the program tries to find a close solution to the known cell.

The program can be forced to choose a solution M from the list of 44 generated thus;

```
AUTOINDEX DPS IMAGE  $n$  SOLU  $M$ 
```

This should only be used when the user knows which solution number is correct; the penalty obtained in the autoindexing is ignored in this case.

The maximum cell length (in Angstroms) used in the search can be set manually;

```
AUTOINDEX DPS IMAGE  $n$  MAXCELL xxxx
```

Sometimes it can help to discriminate between close solutions if they are pre-refined before making a choice;

```
AUTOINDEX DPS IMAGE  $n$  REFINE
```

This gives two extra columns of output (rms SD and fraction of reflections used in the refinement).

The keywords used for REFIX autoindexing can all be used for DPS indexing.

3.2.2.1 Autoindexing using different images from the same crystal

If a separate autoindexing matrix is derived from different images from the same crystal, it is possible that the autoindexing will select different, but symmetry related, orientations for the crystal.

In the particular case where the cell parameters are being refined using two (or more) segments of data (*e.g.* separated by 90 degrees in ϕ), and the crystal orientations for the first image in each segment are sufficiently different (due to slippage) that different orientation matrices are required to get a good prediction, this could lead to an error message in earlier versions of the program.

To avoid this difficulty, the orientation matrix derived from a second or subsequent autoindexing operation is permuted, according to the symmetry operations of the space group, to obtain the matrix that is closest to that obtained from the first autoindexing run. Previously this was implemented for the old (REFIX) style autoindexing but not for the new (DPS) indexing.

[Back to Contents Page](#)

4: Running the STRATEGY option

- 4.1 [Overview of the STRATEGY option](#)
- 4.2 [Some Examples of the STRATEGY options](#)
- 4.3 [Determining the oscillation angle for each image \(TESTGEN option\)](#)

Having determined the crystal orientation, one then needs to know what rotation range is necessary to collect a complete (or essentially complete) dataset. The STRATEGY option provides a very rapid and convenient means for doing this.

4.1 Overview of the STRATEGY option

The STRATEGY option allows the design of a data collection strategy in a semi-automatic way for a single axis rotation camera. It requires all the parameters normally used to process a set of images (crystal symmetry, orientation, crystal to detector distance, wavelength, detector type, direct beam position).

The rotation range (PHITOT) required to collect a complete dataset is determined from the crystal symmetry and orientation (*e.g.* 180 degrees for Laue group P2/m if rotating about the b axis, 90 degrees if rotating about a or c).

The phi value (PHIZONE) which, for orthorhombic (or lower) symmetry places an axis in the XZ plane (containing the X-ray beam and the rotation axis), or for trigonal or higher symmetries places the unique symmetry axis in the plane normal to the X-ray beam and containing the rotation axis, is determined. A reflection list corresponding to a total rotation of PHITOT starting at phi=PHIZONE is generated.

For orthorhombic spacegroups the algorithm used to calculate PHIZONE is not foolproof ! It works in approximately 90-95% of cases. When it does not work, the predicted completeness may be up to 3-4% less than what could be achieved using a different value of PHIZONE. If the predicted completeness is less than expected, try giving the ALTERNATE keyword as part of the STRATEGY command. This will use a different value for PHIZONE which may (rarely) give a higher completeness. As a rule of thumb, it should be possible to get at least 90% completeness for a total rotation of 60 degrees in two segments.

To save time, the true unit cell will be "shrunk" when generating the reflection lists. This can be controlled by the SPEEDUP subkeyword, but the program will calculate a sensible default SPEEDUP if none is specified.

This reflection list is then compared to a list of all unique reflections for this spacegroup and the completeness and multiplicity is calculated, both as a function of rotation and resolution.

It is assumed that all possible reflections are measured (*i.e.* none are lost because of spatial overlaps or because they extend over too many images). However, some reflections may be unobserved because they lie in the cusp region. The percentage of reflections within the cusp will depend on the wavelength, crystal symmetry and crystal orientation, and can be minimised by trying to orient the crystal so that the crystal axis closest to the rotation axis is at least THETAMAX degrees AWAY from the rotation axis, where THETAMAX is the maximum Bragg angle.

It is often possible to collect data with a very high percentage completeness with a total rotation significantly less than PHITOT. This will inevitably result in a lowering of the overall multiplicity, but if data collection time is limited (for example at a synchrotron source) it is preferable to obtain a dataset with high completeness and less than optimal multiplicity rather than an incomplete dataset with higher multiplicity! Equally, if radiation damage is a serious problem, it is best to get a complete dataset first, and then collect additional images to increase the multiplicity.

If the total rotation angle to be collected is specified, and the number (up to 3) of discontinuous segments to be used, the program will determine the start and end phi values for each segment that will give the highest possible completeness. For example, a total rotation of 60 degrees in 2 segments for an orthorhombic spacegroup will result in the identification of two 30 degree segments which give the highest completeness.

If some data has already been collected from one (or more) previous crystals, the program will determine the starting phi value for the "current" crystal that will give the maximum completeness, with the assumption that the phi rotation for this crystal is such that the TOTAL rotation for ALL the crystals is PHITOT (This assumes that all crystals are mounted about the same axis). The user may also define the total rotation angle

for the current crystal.

4.2 Some Examples of the STRATEGY option

Once an image has been autoindexed, select the "**Strategy**" option from the menu. The input for the STRATEGY option has to be given in the I/O window, initially at the MOSFLM => prompt.

4.2.1 No previous data have been collected

Enter the following keywords:

```
STRATEGY
GO
```

The program will determine the phi angle PHIZONE (see above), and generate a reflection list starting at that phi angle, for a total rotation determined by the Laue group. It will then generate a list of all unique reflections and merge the two lists. Finally it will give the completeness of the data for the rotation range generated:

```
Optimum rotation gives  98.0% of unique data
This corresponds to the following rotation ranges for the final run
From  20.0 to 110.0 degrees
Type "STATS" for full statistics
....
STRATEGY =>
```

Typing STATS at the prompt will give a breakdown as a function of rotation angle and resolution, and a breakdown of the anomalous data.

If insufficient time is available to collect the full rotation range required, one can determine the best segments to collect to achieve maximum completeness. Type at the STRATEGY => prompt:

```
STRATEGY => ROTATE 50 SEGMENTS 2
STRATEGY => GO
```

(*** The ROTATE keyword *MUST* be given before the SEGMENTS keyword **) The program will then give the best phi ranges to collect for two segments, each of 25 degrees, giving a total rotation of 50 degrees:

```
Optimum rotation gives  96.1% of unique data
This corresponds to the following rotation ranges for the final run
From  20.0 to  45.0 degrees
From  65.0 to  90.0 degrees
```

One can try using 3 segments of data instead of two:

```
STRATEGY => rotate 50 segments 3
STRATEGY => go
```

In this case, the result is:

```
Optimum rotation gives  98.0% of unique data
```

This corresponds to the following rotation ranges for the final run

From	20.0	to	37.0	degrees
From	52.0	to	69.0	degrees
From	74.0	to	90.0	degrees

The effect of using other total rotations and different numbers of segments can also be tested (but using more than 3 segments is very time consuming and in fact there is an absolute limit of 4 segments). Alternatively, the completeness of specified segments can be tested:

```
START 0 END 20
START 65 END 90
GO
```

Note that the phi ranges specified on the START and END keywords *MUST* lie within the phi range generated by the program when it first starts. Thus if, for example, the program has generated reflections from phi=10 to phi=100 then it not possible to try:

```
START 0 END 30
GO
```

or

```
ROTATE 100
```

at the STRATEGY prompt (The program will complain).

4.2.2 When some data have already been collected

It is also possible to deal with the case where some data have already been collected (from the same or from other crystals).

a) Data from the same crystal

Consider the case where 30 degrees of data have been collected (from phi = -10 to phi = 20 say), and we want to determine how best to complete the dataset with an additional rotation of 40 degrees.

Select the "Strategy" menu option, and enter the following keywords:

```
STRATEGY START -10 END 20 PARTS 2
GO
STRATEGY ROTATE 40 SEGMENTS 2
GO
```

The program will then find the phi values for the two segments (each of 20 degrees) which when combined with the 30 degrees of data already obtained will give the maximum completeness.

b) Data from different crystals

Imagine that data have been collected from phi = -20 to 15 from an orthorhombic crystal with an orientation matrix "xtal_1.mat"

A second crystal is mounted, an image collected and it is autoindexed to give an orientation matrix "xtal_2.mat". The STRATEGY option can now determine the best phi range for this second crystal to complete the data.

First, specify the orientation of the first crystal using "Keyword Input" and the keyword MATRIX:

```
MATRIX xtal_1.mat
```

Then autoindex the first image of the second crystal. This *MUST* be done AFTER the orientation matrix for the first crystal has been specified, because the orientation of the second crystal has to be referred to that of the first.

Then select the "Strategy" menu option, and enter the following keywords:

```
MATRIX xtal_1.mat
STRATEGY start -20 end 19 PARTS 2
GO
MATRIX xtal_2.mat
STRATEGY AUTO
GO
```

Normally the first crystal will have been collected starting at a zone. If this is NOT the case, it will probably be necessary to collect two segments of data from the second crystal to get complete data. This can be done by specifying "STRATEGY AUTO SEGMENTS 2" for the second crystal, and it may be advantageous to specify the sizes of the two segments. Thus if the first crystal was collected starting at 15 degrees away from a zone, for a total of 35 degrees, then the second crystal will need one segment of 15 degrees and another of 40 degrees (90-35-15) to get best completeness.

It is also possible to automatically find the best rotation(s) for a smaller total rotation. Once the program has come up with the STRATEGY => prompt (*i.e.* after it has found the best solution for a single 55 degree rotation in the above case) one can then type:

```
STRATEGY => PART 1                ! Include all data from first crystal
STRATEGY => AUTO ROTATE 40 SEGMENTS 2 ! Use 2 segments (each 20 degrees for
                                     ! second crystal
```

This means include ALL data that has already been collected (from -20 to +19 in the above example) and then determine the best phi values giving a total rotation of 40 degrees (in two 20 degree segments) from the second crystal.

4.2.3 Optimising anomalous data collection

To optimise the number of anomalous pairs rather than the completeness of the unique data simply include the subkeyword ANOMALOUS:

```
STRATEGY ROTATE 60 SEGMENTS 2 ANOMALOUS
```

This will not necessarily be the same phi range(s) as that which maximise the overall completeness.

4.2.4 A complete list of the STRATEGY subkeywords

STRATEGY subkeywords

subkeywords: AUTO ROTATE SEGMENTS SIZES START END PARTS SPEEDUP ANOMALOUS

AUTO Determine the starting phi angle and the phi rotation required to give a complete dataset (if possible from a single crystal setting), and give statistics on completeness and multiplicity. Do NOT use **START** or **END** with the **AUTO** keyword. This is the default mode of running strategy.

ROTATE <phirot> Only for use with the **AUTO** option. Restrict the total rotation to "phirot" degrees.

SEGMENTS <nseg> Only for use with the **AUTO** option. Allow "nseg" discontinuous segments of data to give a total rotation of **PHIROT** degrees. Unless specified explicitly with the **SIZES** keyword (see below) the segments will have approximately equal widths in phi.

SIZES <size1,size2,size3...> The sizes for the "nseg" segments. If **SIZES** are given, then the "phirot" value given on the **ROTATE** keyword is ignored, and the total rotation is the sum of the **SIZES**. Default: Use approximately equal sizes with total "phirot"

START <phistart> **END** <phiend> As an alternative to **AUTO** mode, specify the start and end phi values to be used in generating the reflection list. Up to 10 different sets of **START** and **END** can be given on successive **STRATEGY** keywords. *e.g.*

```
STRATEGY START 0 END 30
STRATEGY START 35 END 60
STRATEGY START 70 END 90
```

PARTS <nparts> If some data have already been collected (from the same or other crystals), set "nparts" to the total number of segments of data already collected plus one (which is the current crystal or segment whose phi range is to be determined). This need only be given on the first **STRATEGY** keyword.

SPEEDUP <n> Speed up the calculation by a factor "n".

ANOMALOUS Optimise anomalous pairs rather than completeness of data.

4.3 Determining the oscillation angle for each image (TESTGEN option)

The completeness analysis assumes that NO reflections are spatially overlapped. Providing that spots within a lune (*i.e.* in the same plane in reciprocal space) are not overlapping, spatial overlaps can usually be reduced to an acceptable level by an appropriate choice of oscillation angle for each image.

The **TESTGEN** option will calculate the maximum allowed oscillation angles as a function of the phi value for a given maximum acceptable percentage of overlapped reflections (which can be zero).

The determination of whether or not a reflection is spatially overlapped depends crucially on the mosaic spread, beam divergence parameters and the minimum allowed spot separation. The mosaic spread and the minimum spot separation can be reset at the **STRATEGY** prompt to test how critical these values are, using keywords **MOSAIC** and **SEPARATION** respectively.

In versions of MOSFLM before v6.2.2, the oscillation angle had to be more than half the sum of the mosaic spread and beam divergence for post-refinement to work. This is no longer the case *except* when using the **POSTREF NOMULTI** option

Note that the **TESTGEN** keyword can be given at the **STRATEGY** prompt or at the normal MOSFLM prompt (without running the **STRATEGY** option).

Keywords:

TESTGEN subkeywords

subkeywords: **START END STEP OVERLAP MINOSC MAXOSC ANGLE**

START <phstart> Define the starting phi. This keyword *MUST* be given.

END <phend> Define the ending phi. This keyword *MUST* be given.

STEP <phstep> The optimum rotation angle will be calculated every "phstep" degrees between "phstart" and "phend". Default: 5 degrees

OVERLAP <x> The maximum rotation angle giving less than x% overlapped reflections will be calculated. Note that x is in PERCENT. Default 0%

MINOSC <rotmin>

MAXOSC <rotmax> Only rotation angles between "rotmin" and "rotmax" will be considered. Default: rotmin 0.2, rotmax 5.0

ANGLE <oscang> If the **ANGLE** keyword is given, then the overlap for a fixed oscillation angle "oscang" is calculated between phi=phstart and phi=phend. No attempt to find the "best" oscillation angle is made.

example:

```
TESTGEN START 0 END 90 OVERLAP 3 MINOSC 0.5
```

Exiting the **STRATEGY** option

Use keyword **EXIT** to end the strategy option.

An Example command file when not using the X-window menu

In this case, the type of detector (MAR,SMALLMAR,RAXIS etc) has to be specified, and the crystal to detector distance and wavelength as these cannot be read from the image header.

```
STRATEGY AUTO
DISTANCE 80
DETECTOR SMALLMAR
MATRIX lyso_1.mat
SYMM 19
BEAM 90 90
DIVERGENCE 0.35 0.3
MOSAIC 0.2 ! or MOSAIC ESTIMATE)
```

```
SEPARATION 1.5 1.5
POLARISATION MONOCHROMATOR
WAVELENGTH 1.5418
RUN
```

[Back to Contents Page](#)

5: Determining Accurate Cell parameters

- 5.1 [Using Post-refinement to refine the cell](#)
 - 5.1.1 [Doing post-refinement interactively](#)
 - 5.1.2 [Doing post refinement in background](#)
 - 5.1.3 [What the program actually does](#)
 - 5.1.4 [Using several segments or different crystals](#)
 - 5.1.5 [Tips on post-refinement](#)
 - 5.1.5.1 [Processing images showing strong diffuse scatter](#)
-

The unit cell parameters are refined as part of the autoindexing, but in general not all the parameters will be well defined (in particular, the cell parameter along the X-ray beam direction is ill-determined). Improved values can be obtained by using two or more images widely separated in phi for the autoindexing. However, accurate cell parameters are best determined by post-refinement, for which it is necessary to have a number (at least two) of abutting oscillation images. To obtain accurate cell parameters for orthorhombic or lower symmetry spacegroups, it is essential to have data from two orientations widely separated in phi, but for trigonal or higher symmetry only one "block" of data is normally required.

A pragmatic procedure is as follows:

If more than about 15 degrees of data are available from a single crystal, or several crystals in approximately the same orientation (within 20 degrees) use the "Refine cell" menu option (or the POSTREF SEGMENT option if running in background) to get an accurate cell and then do NOT refine it during integration.

If less than 15 degrees is available, use the refined cell from the autoindexing in processing and try post-refinement using an angular wedge of data, but if this is unstable (large sd's or shifts from cycle to cycle) then fix the cell parameters, as the values from the autoindexing, while they may be in error, will be sufficiently accurate to process a "local region" in reciprocal space, *i.e.* up to 10-15 degrees from the starting phi value.

5.1 Using Post-refinement to refine the cell

Post-refinement uses the distribution of the intensity of partially recorded reflections over the images on which the partial is recorded (the previous limit of two images no longer applies) to refine cell parameters, orientation and mosaic spread. It has the distinct advantage that the derived cell parameters are entirely independent of all detector parameters (crystal to detector distance and detector orientation) and distortions (ROFF and TOFF) which, if inaccurate, can lead to significant errors in the cell parameters derived from autoindexing.

**** IMPORTANT ****

The default post-refinement can now use partially recorded reflections which extend over several images; MOSFLM now determines the number of images over which half the current list of partials extend. It is still

possible to restrict MOSFLM to using partial reflections which are spread over only two images by using the **POSTREF NOMULTI** option. If refining the cell interactively, use the "**Keyword input**" menu item to give these keywords.

5.1.1 Doing post-refinement interactively

Having obtained the crystal orientation by autoindexing, choose the "**Refine cell**" menu option. You can then select the number of "segments" of data to use in the refinement, the first image and the number of images to be used in each segment. Note that there must be at least two images in each segment, but there is generally little to be gained from using a total of more than 8-10 images in ALL segments (unless there are only a few partials on each image).

Note that when using data from two segments widely separated in phi, it is possible that the crystal orientation will have changed sufficiently that the orientation matrix for the first segment of data does not accurately predict the first image of the second segment. This can be quickly checked by reading in this image ("**Read image**" menu option) and then predicting the pattern ("**Predict**"). If the prediction is poor, there are two things that can be done. Either find spots on this second image and use them (together with the spots from the first image of the first segment) to repeat the autoindexing. This may give a matrix that predicts both images successfully. This should work unless the crystal orientation has genuinely changed between the two images (or the rotation axis is not normal to the X-ray beam). If this does not work, you should derive a new orientation matrix for an image from the second segment image. Remember to change the name of the file that the orientation matrix will be written to. REMEMBER to delete all the spots used to autoindex the first image if you have not already done so, or use the "**Select images**" menu option to choose only spots from the second image. Then use the "**Autoindex**" option to get an orientation matrix for this image. define a separate orientation matrix for each segment of images.

Because the post-refinement uses partially recorded reflections, it is important to have a realistic estimate of the mosaic spread BEFORE starting post-refinement. In particular, if no value has been supplied (*i.e.* the mosaic spread is zero) the program will issue a warning message because it is unlikely that the post-refinement will work. Use the "**Estimate mosaicity**" menu item to obtain an initial estimate of the mosaic spread. The postrefinement will give a refined estimate of the mosaic spread, but this is not very reliable for mosaic spreads greater than about 0.7 degrees.

5.1.2 Doing post refinement in background

To use this option, the keyword :

POSTREFINEMENT SEGMENT <number of segments>

should be used, followed by **PROCESS** keywords defining the images to be included in each segment, with each **PROCESS** (see 5.3.1) keyword followed by a **RUN** keyword.

Example:

```
NEWMAT postref_3seg.mat      ! Defines the filename for the new matrix
POSTREF SEGMENT 3
PROCESS 1 3 [ ANGLE 1.0 START 0.0 ]
RUN
PROCESS 43 45 [ ANGLE 1.0 START 42.0 ]
RUN
```

```
MATRIX test_88.mat
PROCESS 86 88 [ ANGLE 1.0 START 85.0 ]
RUN
```

Would use 3 segments of data (with phi values 0-3,42-45,85-88.) Note that a new **MATRIX** keyword has been given for the last segment, which could be necessary if the crystal has slipped during data collection. See [section 5.1.1](#) for the best procedure to use when deriving an orientation matrix for the second or subsequent segments.

Note that the procedure uses only partially recorded reflections, and so in this case it would use partials that span images 1 and 2, 2 and 3, and 1,2 and 3 for the first segment etc. For this reason the **PROCESS** keyword *MUST* specify at LEAST 2 images,

e.g. `PROCESS 1 1 ANGLE 1.0 START 0.0`

would provide NO data for post-refinement.

5.1.3 What the program actually does

During postrefinement, the images are not fully integrated (only the intensities of partially recorded reflections are measured, and by summation integration rather than profile fitting) so there is no output generate file or MTZ file. The crystal orientation will be refined for every image independently, but the cell parameters will only be refined once the final segment of data has been processed

(Note that the very last image (88 in the example above) is apparently (from the logfile) not measured at all...this is *NOT* an error, since the intensities of the partials at the start of image 88 are obtained while processing image 87.)

If the cell parameters change by more than 2.5 standard deviations from the input values, all images will be remeasured using the updated cell and another round of cell parameter post-refinement will be carried out. This will happen up to a maximum of 4 repeats. It is quite common that two or even three complete rounds of integration are required for convergence. For this reason it is not a good idea to include too many images in the refinement. A target of between 500 and 2000 reflections in the refinement is perfectly adequate.

It is recommended that the final cell parameters are then used to integrate all the images in the dataset, fixing the cell parameters in the post-refinement:

```
POSTREF FIX ALL
```

5.1.4 Using several segments or different crystals

Note that if the crystal has been slipping during data collection, it is possible to provide different **MATRIX** keywords for each segment of data, and supply a new orientation (*e.g.* derived by autoindexing the first image of the segment). When doing this, the orientation matrices for all segments (including the first) **SHOULD BE OBTAINED FROM THE SAME INTERACTIVE RUN OF MOSFLM**. This ensures that the matrices for the second and subsequent segments are all referred relative to the orientation matrix for the first segment. It is also a good idea to **FIX** the cell parameters when autoindexing the images from the second and subsequent segments, as only one set of cell parameters is allowed when refining the cell by post-refinement. It is also possible to provide new crystal identifiers for each segment (eg if the crystal has been translated and the images given a different identifier). It is also possible to use data from different crystals, but in this case there

is the restriction that the orientation of the crystals must be the same (to within 20 degrees) and the relative phi values must be correct. Providing the different crystals are all indexed in the same run of MOSFLM, the relative phi values are taken care of automatically.

A possible complete example is then:

```
TITLE  Refine cell with 3 segments
DIVERGENCE 0.35 0.2
SYMMETRY 96
[DISTANCE 124.1]
[WAVELENGTH 1.542}
DIRECTORY /scr0/andrew/
BEAM 89.33 90.10
GAIN 1.2
NEWMAT postref_3seg.mat
POSTREF SEGMENT 3
IDENT oval1
MATRIX oval1.mat
PROCESS 1 3 [ANGLE 1.0 start 0.0]
RUN
IDENT oval2
MATRIX oval43.mat
PROCESS 43 45 [ANGLE 1.0 START 42.0]
RUN
IDENT oval3
MATRIX oval86.mat
PROCESS 86 88 [ANGLE 1.0 START 85.0]
RUN
```

When doing post-refinement, the crystal orientation around the X-ray beam direction (the X axis) is not defined (the refinement is based solely on the observed degree of partiality and not on the positions of the spots) and this parameter is therefore not refined, but missetting angles around Y and Z axes are refined (see [Appendix III](#) for a definition of coordinate frames). The refinement of the detector parameter CCOMEGA allows for crystal slippage around the X-ray beam direction.

If only a narrow angular wedge of data is available for a low symmetry spacegroup (orthorhombic or lower) it is possible to FIX cell parameters that are not well defined (those closest to the direction of the X-ray beam)

e.g. POSTREF FIX A

5.1.5 Tips on post-refinement

In the great majority of cases the post-refinement will provide accurate cell parameters without any user intervention (providing the mosaic spread estimate is realistic). There are, however, some special cases where additional input is required to get the best results.

5.1.5.1 Processing images showing strong diffuse scatter.

It is not uncommon to observe diffuse scatter on the images, particularly for data collected at a synchrotron source. Sometimes this takes the appearance of a "halo" around the Bragg spot, because the intensity of some types of diffuse scatter peak at the positions of the Bragg reflections. This can cause difficulties in post-refinement, because it has the same effect as a crystal with a very large mosaic spread. Under these circumstances, it is best to refine the cell parameters using spots that are close to half-recorded, as the refinement is then less sensitive to the model for the "rocking curve". The minimum and maximum fraction recorded can be specified as shown below:

```
POSTREF FRMIN 0.4 FRMAX 0.6
```

will only use reflections that are between 0.4 and 0.6 recorded. (Default is 0.1 to 0.9).

[Back to Contents Page](#)

6: Collecting data and processing the images

- 6.1 [Overview](#)
 - 6.2 [Special MOSFLM features](#)
 - 6.2.1 [Accumulating profiles over several images](#)
 - 6.2.2 [Addition of partials \(ADDPART\)](#)
 - 6.2.3 [Post-refinement of orientation and cell parameters](#)
 - 6.2.4 [Optimisation of measurement box parameters](#)
 - 6.3 [Running a processing job](#)
 - 6.3.1 [Running MOSFLM interactively](#)
 - 6.3.2 [Processing the first block of data\) \(Non-interactively\)](#)
 - 6.3.3 [Finally, Processing the dataset](#)
-

6.1 Overview

Before starting the serious data collection, integration of one or more images should be carried out to determine:

- a) Is the crystal single?
- b) Is the exposure time correct?
- c) Is the crystal to detector distance correct (*i.e.* the whole of the detector is being used)?
- d) Can the images be processed...are the spots separated and is the number of spatial overlaps small?

6.2 Special MOSFLM features

There are 4 features of MOSFLM which are unusual and require explanation. These are:

- 1 Accumulation of standard profiles over several images.
- 2 Addition of partially recorded reflections over adjacent images.
- 3 Post-refinement of cell parameters and crystal orientation.
- 4 Optimisation of the measurement box parameters.

6.2.1 Accumulation of profiles

In order to form well defined standard profiles (which are then used to evaluate the profile fitted intensities) fully recorded (or partially recorded) reflections over several images are added together. This improves the signal to noise and results in a better determined profile. The number of images used to form the profiles (usually between 5 and 10) is determined automatically by the program (in a way that avoids having just a few images in the final block). It can also be set manually by the **BLOCK** subkeyword on the **PROCESS** keyword line.

The positional refinement for all images in a block is carried out prior to forming the standard profiles and integrating the images. Thus each image is processed in two passes, the first pass for the positional refinement and writing all the "measurement boxes" for the spots to the SPOTOD file, and the second for actually evaluating the reflection intensities.

6.2.2 Addition of partials (ADDPART option)

The program has the option to add together the measurement boxes of the two halves of partially recorded reflections on adjacent images, thus giving the equivalent fully recorded reflection which can then be used to form standard profiles or for positional refinement of the detector parameters.

To make use of this option, the keyword:

ADDPARTials

should be given. (The default is now NOT to add partials).

Note that this procedure, which involves adding pixel values on two adjacent images, involves two assumptions:

Assumption 1

That the images have the same effective exposure time (*i.e.* total incident flux). If the rate of rotation of the spindle axis is determined by the ionisation chamber reading (as it may be on the MAR detector) then this assumption should be met. If not, then there may be an error introduced by this procedure especially if the incident beam is rapidly decaying (eg on an unstable synchrotron source). If post-refinement is being used (and by default it is used) then the program will print a warning message (to the summary file and the end of the logfile) if the exposure varies by more than 5% from one image to the next (as judged by the X-ray background).

Assumption 2

That the detector origin, orientation etc is identical for successive images, and that the images are exactly abutting (*i.e.* no overlap in rotation angle). These conditions will normally be met by the Mar, R-axis and Mac Science scanners, but mechanical wear can lead to the scanner not locking into the correct "home" position after a scan (it does one more or one too few rotations). This will show up as a variation in the **ROFF** distortion parameters in units of one pixel (0.15mm on a Mar IP). The program keeps track of variations in **ROFF**, **TOFF** and **CCOMEGA** and will give a warning message if undue variation is detected.

****** IMPORTANT ******

If either of these assumptions is not met (this will be indicated by warning messages) then the **ADDPART** option should not be used.

With **ADDPART**, what are actually partially recorded reflections over 2 images are reclassified as fully recorded when stored in the MTZ file and they will therefore be used in scaling (**SCALA**). However, summed partials do carry a special flag, so that they are still classified as partials in the statistical analysis in **SCALA**. Thus information on partial bias, for example, is still available.

Because of the ability of **SCALA** to scale data when there are no fully recorded reflections, the use of this option less important than it once was. Because its use depends on the assumptions listed above, which may not always be met, the *DEFAULT* is now *NOT* to add partials.

6.2.3 Post-refinement of cell parameters and crystal orientation

By default the program will refine both cell parameters and crystal orientation using post-refinement during integration of the images. However, it is in fact preferable to determine accurate cell parameters prior to integration using the **Refine cell** menu option for interactive work or the **POSTREF SEGMENT** option in a background job. The resulting cell parameters are then input using a **CELL** or **MATRIX** keyword and the cell is *NOT* refined during integration (by using keywords **POSTREF FIX ALL**). This will refine the crystal orientation (and mosaic spread) but not cell parameters.

If cell parameters are refined in a processing job, the way in which the refinement is carried out depends on the crystal spacegroup. For crystals of trigonal or higher symmetry data from each pair of images in turn is used in the refinement. (This is equivalent to the **POSTREF SINGLE** mode.) Thus cell parameters, crystal orientation and mosaic spread are refined after every image using intensities on that image and the next one in the series. (For off-line scanners, reflections on the current image and the preceding image are used).

For lower symmetry this is not recommended, because not all the cell parameters will be well defined using data from only one pair of images. Thus for orthorhombic and lower symmetries data is accumulated from a number of images and only then will cell parameter refinement be carried out (the crystal orientation is still refined after every image as this is well defined). By default, the number of images required for the cell parameter refinement (**NADD**) is set to correspond to a rotation of 10 degrees. However, this can be changed using the **WIDTH** subkeyword. Thus:

POSTREF WIDTH 15

specifies that 15 degrees of data must be accumulated for post-refinement of cell parameters. The actual **WIDTH** of data required for a satisfactory refinement will depend on the resolution (the higher the resolution, the fewer images are required) and the strength of the data (the weaker the data, the more images are required). Some experimentation may be required to find a **WIDTH** that gives a stable refinement. If the refinement appears unstable (*i.e.* large shifts in cell parameters) the **WIDTH** should be increased. If this is not possible (*e.g.* only a limited number of images have been obtained from the crystal before radiation damage set in) then the refinement of some or all cell parameters should be turned off. Thus

POSTREF FIX ALL

will fix all cell parameters.

POSTREF FIX C

will fix the "c" cell parameter etc. Normally one would fix the cell parameter that is closest to being parallel to the X-ray beam as this will be the least well defined. Alternatively, look at the standard deviations of the cell parameters to see which one(s) are least well defined. Normally the cell parameters obtained from autoindexing are quite adequate to measure 10-20 degrees of data from the image on which the autoindexing was run.

Once the appropriate number of images (**NADD**) have been processed, and the cell parameters have been refined for the first time, if there is a large shift in any cell parameter the program will start processing from the first image again, using the updated cell parameters. The maximum shift allowed is determined by the subkeyword **SHIFTFAC**; thus

POSTREF SHIFTFAC 5

sets the maximum shift to 5 standard deviations; if larger than this the images will be reprocessed. The default value is 2.5.

From this point on, the cell parameters will be re-refined after every image, using data from the previous **NADD** images. For example, with 1 degree oscillation images and a width of 10 degrees, the first cell refinement will be carried out after processing image 10, using data from images 1 to 10. After processing image 11, cell parameters will be refined using data from images 2-11 etc. etc.

The missetting angles should *ALWAYS* be refined by post-refinement, but it may be necessary in some cases to suppress or limit refinement of cell parameters if the refinement is not stable.

The crystal mosaic spread is also refined by default, but the refined value *IS NOT USED BY DEFAULT*. This is because if the refinement is unstable, this can have rather drastic effects on the processing. If the refinement is stable, and there is evidence for a change in mosaic spread during the run (this often results from radiation damage), the refined values should be used by including the subkeyword **USEBEAM**:

POSTREF USEBEAM

This is the default from MOSFLM version 6.2.3.

If you wish to refine the horizontal and vertical beam divergence independently (good data is required to do this) use **BEAM 2** :

POSTREF BEAM 2

Again, you need to include **USEBEAM** to actually make use of the refined values.

6.2.4 Optimisation of measurement box parameters

By default the program will automatically determine the best measurement box parameters. It will first determine the spot size from spots in the centre of the image (parameters for this search are set by keyword **SPOT**). This information is used to set initial sizes for the overall dimensions (**NXS,NYS** in figure below) and the corner and rims parameters (**NC,NRX,NRY**). Following detector parameter refinement using spots from the centre of the first image, the program will then optimise the rim and corner parameters **NRX,NRY** and **NC**.

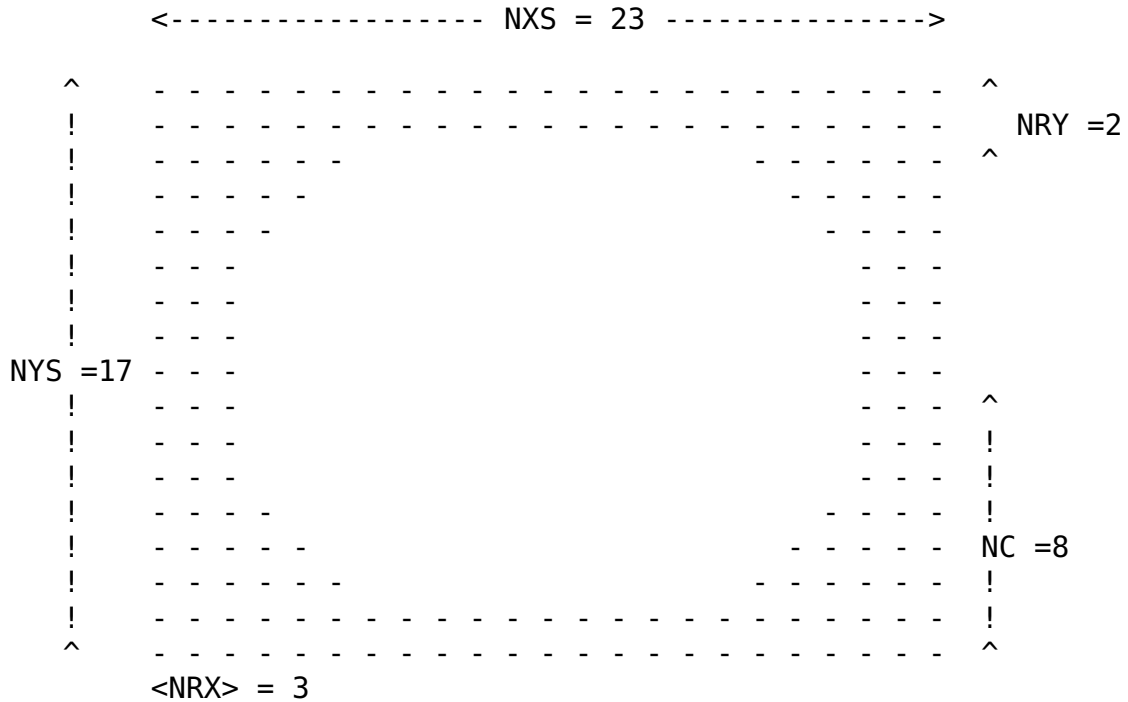


Figure 1. The measurement box used in MOSFLM. NXS and NYS (odd integers) define the overall size of the measurement box in pixels. NRX and NRY define the widths of the background rim and NC defines the corner cutoff. In the figure a "-" denotes a background pixel, all other pixels belong to the peak. There is no "safety rim" between peak and background.

The algorithm employed is as follows. The parameters NRX, NRY, NC are varied in turn and the value giving the highest ratio of the integrated intensity I to the standard deviation in the intensity $\sigma(I)$ is found. This is the notional optimum value for that parameter. The total intensity for this value is checked, and if it is less than the maximum intensity (for any value of the parameter) by more than a factor 0.01 (**TOLERANCE**), then that parameter is decreased by up to **IBOUND** pixels.

For example, the max $I/\sigma(i)$ might be found for a X rim value of 4, but the intensity might be only 97% of the maximum intensity found for any value of NRX (*e.g.* 1). NRX will therefore be decreased, one pixel at a time, and for each value the integrated intensity tested against the maximum value. If for NRX=2, the intensity is within 0.01 (*i.e.* 1%) of the maximum value, then 2 will be taken as the optimal value for NRX.

Thus it can be seen that the higher the **TOLERANCE** parameter, the **SMALLER** the optimised peak area will be. It is difficult to define a "correct" value for **TOLERANCE**, because this will depend on the degree of diffuse scatter associated with the Bragg peaks and how well adjacent spots are resolved. Values between 0.01 and 0.04 are typical, it should **NOT** be necessary to use values above 0.04. Note that two values may be supplied to the **TOLERANCE** keyword. In this case the first value is used for profiles in the centre of the image (closest to the direct beam) and the second value for the outermost profiles. An interpolated value is used for other profiles. The default value for the innermost profile is 0.01. For very close spots this can be increased to 0.02 (rarely larger).

For the first round of this iteration, when optimising NRX and NRY, NC is set to zero. NRX and NRY are varied from 1 to a maximum value which would give a peak dimension of 5 pixels. NC is varied from the smaller of NRX and NRY up to the smaller of (NX-2) and (NY-2).

Two rounds of optimisation are performed, the second round using the results of the first.

The optimisation is first carried out on the average spot profile for the central region of the detector. The optimisation of the overall dimensions (NXS,NYS) is only carried out at this stage. The background rim parameters are optimised for *ALL* the standard profiles.

The background rim parameters are reoptimised for each new **BLOCK** of images.

If the optimisation of the standard profiles causes problems (because of unusual spot shapes with long tails or other features) it can be suppressed using keywords:

PROFILE NOOPTIMISE

In this case the measurement box parameters will still be optimised for the average spot profile using spots from the centre of the detector (this profile is NOT used for integration). This box will then be expanded automatically to allow for the increase in spot size due to obliquity of incidence on the detector, but the measurement box parameters will NOT be optimised for the standard profiles (one for each area of the detector) that are used for integration. To suppress the optimisation of the measurement box parameters altogether (so that the program will use the parameters supplied on the **RASTER** keyword), give the keywords:

PROFILE NOOPTIMISE ATALL FIXBOX

To just suppress the optimisation of the overall size of the measurement box (parameters NXS,NYS) include keywords:

PROFILE FIXBOX

6.3 Running a processing job

Normally at this stage one would go straight on to process the first **BLOCK** of images (See "6.3.2 Processing the first block of data" below).

6.3.1 Running MOSFLM interactively

The following keywords might be used for an interactive run. Before starting **MOSFLM**, it is convenient to edit them into a file (*eg comm*) to save typing them several times. Then at the **MOSFLM** prompt simply type:

```
@comm
```

and it will read the commands from the file

```
TITLE Processing test data
XNAME plate_32_4A
DNAME native
PNAME oval
! Crystal parameters
MATRIX oval_2_3.mat
SYMMETRY 96
MOSAIC 0.1
```

```

! image parameters
IDENT oval
PROCESS 1 TO 1 [ ANGLE 1.0 START 0.0 ]
DIRECTORY /scr0/andrew/
EXTENSION image

! beam parameters
DIVERGENCE 0.35 0.2
POLARISATION MIRRORS

! detector parameters
BEAM 89.33 90.10
BACKSTOP CENTRE 88.5 91 RADIUS 14
GAIN 1.2

! The following are read from the image header if not supplied on
! keywords for most modern detectors, . The phi values (on
! PROCESS keyword) will also be read from the header if not supplied.

DISTANCE 124.1
WAVELENGTH 1.542

```

The following are optional, if not given, the program will set suitable defaults.

```

!HKLOUT oval.mtz
!SEPARATION 1.3 1.3
!RASTER 17 17 9 3 3
!GENFILE oval_1to1.gen
!RESOLUTION 3.0

PLOT
RUN

```

*** NOTE WELL ***

If the data has been collected at a synchrotron source, the polarisation of the beam and the horizontal and vertical divergences (horizontal here means in the plane of the X-ray beam and the rotation axis) should be given. Values default to those for the SRS at Daresbury, UK.

The **TITLE** is written to the mtz and generate files.

XNAME, **DNAME** and **PNAME** are data harvesting labels which (from version 6.2.3) are used for carrying forward information in the MTZ file about the *crystal* (from crystallisation plate 32 well 4A), *dataset* (native rather than derivative) and *project* (a protein called "oval").

MATRIX is the filename for the orientation matrix derived from a previous autoindexing or cell parameter post-refinement run. If you want to override the cell parameters or missetting angles in this matrix use the **CELL** and **MISSETTS** keywords respectively.

SYMMETRY gives the space group of the crystal, either as the name or the number in International Tables. Note that axial systematic absences (*e.g.* 0k0 with k odd in spacegroup P2₁) are measured by MOSFLM, so

that the symmetry can be checked. Lattice absences however (due to face or body centring) will NOT be measured.

The **IDENTifier** (oval) is used as a template for the image file names, which have the form:

oval_003.image

for image number 3 for example. There is an absolute limit of 40 characters for **IDENT**. The extension (image in this case) is set using the **EXTENSION** keyword.

The template keyword is an alternative to the **IDENT** keyword. In this case:

TEMPLATE oval_###.image

would work identically.

PROCESS gives the images to be generated, in this case from image 1 to image 1 (*i.e.* only one image is going to be examined), with a rotation angle of 1.0 degrees, starting at phi=0 (relative to the phi values given in autoindexing). Note that for Mar Research, ADSC, R-AxisIV and Mac Science scanners, the phi values need not be given, as they will be taken from the image header.

DIRECTORY gives the directory name where the images are stored (up to 10 different directories can be given on one or more **DIRECTORY** keywords).

EXTENSION defines the extension of the image filenames (default is "image" for Mar Research scanners, "osc" for R-axis, "ipf" for Mac Science, "cor" for ESRF CCD, and "image" for unrecognised detectors).

DIVERGENCE... if two values are given, they are the horizontal and vertical beam divergences (which will differ if a monochromator is used). Only a single number need be given for isotropic divergence. "Horizontal" in this context actually means in the plane containing the rotation axis and the X-ray beam, which is horizontal in the case of the Enraf Nonius oscillation camera and the Mar Research IP scanners, but vertical for standard R-Axis machines.

POLARISATION specifies the polarisation of the X-ray beam, and can be given as **PINHOLE** or **MIRRORS** (both specify an unpolarised beam), **MONOCHROMATOR** (for a **GRAPHITE** monochromator) or **SYNCHROTRON** followed by the degree of polarisation (0.95 for SRS).

BEAM defines the direct beam position, in mm, relative to the position of the first pixel in the image. This can be determined by taking a wax image (or plasticine) and measuring the centre of the circles using the X-windows display.

Special note for R-Axis II scanners: The definition of the detector coordinates in the R-axis software is different to that adopted in MOSFLM. Thus if the direct beam coordinates have been obtained from the R-axis software, then the X and Y coordinates must be interchanged.

BACKSTOP Defines the centre and radius of the backstop shadow. Reflections lying within this circle will not be integrated. The position and size of the shadow are best determined using the X-windows display.

GAIN defines the gain (adc units per X-ray photon) of the detector. This should be constant for a given detector (and a fixed wavelength). It is *CRUCIAL* to have a reasonable estimate of the gain as many aspects of the program use counting statistics derived standard deviations to determine acceptable spots for refinement,

profiles etc. The correct value of the gain should give a **BGRATIO** of unity (the **BGRATIO** is printed as part of the MOSFLM output, as a function of intensity). The actual value of **BGRATIO** obtained can be used to get an improved estimate of the gain using the relationship:

$$\text{true gain} = \text{estimated gain} * (\text{bgratio})^{**2}$$

The gain is typically between 1 and 2, but can be as high as 5 for R-Axis II detectors. See [section 8.1](#) for a way of estimating the **GAIN** of your detector.

IMPORTANT. The **GAIN** should not be interpreted too literally. The way it is derived in section 8.1 is only strictly correct if all pixels are independent, which in practice they are not. For this reason, this parameter should NOT be used to assess the relative performance of different detectors.

DISTANCE is the crystal to detector distance in mm.

WAVELENGTH is the radiation wavelength (defaults to 1.5418).

HKLOUT gives the name of the output mtz file. This contains the Lp corrected intensities and standard deviations, with the reflection indices reduced to the asymmetric unit. This file can be used (after sorting) as input to the CCP4 programs **SCALA**. **HKLOUT** can also be given on the command line, but if both are given that specified by the keyword takes precedence. If not given, the filename is made up from the image identifier and the first image number, so in this case would be "oval_001.mtz".

The subkeyword **MULTIPLE** can be used to force the program to write a new mtz file for each block of images.

SEPARATION (in mm in detector directions X and Y, *i.e.* horizontal and vertical for the Mar IP scanner) gives the minimum allowed separation of two spots before they are flagged as spatially overlapped (spots are treated as being ellipsoidal in shape, the numbers given being the full axial lengths in the X and Y directions). No attempt is made to integrate spatially overlapped reflections. If not given, the program will work out suitable values based on the spot size in the centre of the image. It will also determine if the "CLOSE" option needs to be used (see **SEPARATION** in the helpfile for more details). For spots that are not, in fact, completely resolved, the values determined by the program may be too conservative and lead to a very large number of spots being rejected as overlapped. In such cases, the **SEPARATION** should be defined explicitly.

RASTER gives the parameters of the measurement box (see Fig. 1 above). As explained in [section 6.2.4](#) above, these values will be determined automatically if not supplied.

GENFILE specifies the generate filename. If not given, it will default to the MTZ filename, but with the suffix ".gen"

RESOLUTION: If not given, the resolution is set by the physical size of the detector. Both high and low resolution limits can be given. A "dynamic" high resolution limit, which depends on the mean $I/\sigma(I)$, can be set using the **CUTOFF** keyword. A specific resolution range can also be excluded (eg to eliminate an ice ring) with the **EXCLUDE** keyword.

PLOT invokes the X-window interactive graphics option

RUN will start the processing

The image will be displayed with the predicted pattern overlaid. Fully recorded reflections are displayed as blue boxes, partials as yellow boxes, spatial overlaps as red boxes and reflections rejected as being too wide in ϕ (default 5 degrees, can be reset with keyword **MAXWIDTH**) as green boxes. Clicking (left mouse button) on the centre of a box will result in the reflection indices, ϕ value and ϕ width being displayed in the "Output" window. If examining the image after integration, the profile fitted intensity and standard deviation will also be given. See below for more details on examining the image after integration.

The image should be examined carefully to ensure that the predicted pattern actually matches what is on the image. If it does not, then any of the relevant parameters (cell dimensions, missetting angles, beam divergence, etc etc) can be adjusted and the pattern re-predicted (use the "Predict" menu item). A brief description of the various menu options is given below:

Predict If the cell dimensions or any other parameter in the "Display Parameters" window is changed, selecting this menu item will re-calculate the predicted reflections and display them.

Clear prediction This will delete the predicted pattern. It can be restored by choosing the "Predict" menu item.

Adjust If there is an error in the beam coordinates or camera constants the calculated pattern will be displaced relative to the observed pattern. The "Adjust" option allows this to be corrected. The mouse is used to input the calculated and the observed positions of two spots. From this the program calculates the shift, rotation and scale factor required to superimpose these two spots. The values of the shifts required are given and the user given the choice of accepting the transformation or not.

Auto-refine ***** *THIS OPTION IS NO LONGER SUPPORTED* *****

This will refine the crystal missetting angles using the **AUTOMATCH** option (see help library for more details), which essentially adjusts the missetting angles to try to optimise the fit of the predicted to the observed pattern. This can converge from initial errors of 1-2 degrees, but the final parameters are NOT as accurate as those obtained from post-refinement. With the use of either the **REFIX** or **DPS** indexing option, this should not normally be necessary. Note that it will use the default parameters for the refinement (*e.g.* only using data to 6Å). You may wish to modify these using the appropriate keywords...see help library documentation. After refinement it will go on to measure the image.

***** *THIS OPTION IS NO LONGER SUPPORTED* *****

Integrate This measures the image in the normal way but keeps the display; this is considerably slower than running the integration as a background job and is not normally recommended. You are given the option of displaying the image again after positional refinement has been carried out, or after the integration has been carried out. (See below)

Find hkl If this menu item is activated, the user is prompted for the hkl indices of the spot they wish to find (the predictions must be displayed). A blue cross is drawn on the position of the spot. A warning is given if the spot does not lie in the displayed part of the image. This can be useful in identifying bad spots.

Pick This will display the actual pixel values in a box around the cursor position (the size of the box can be set in the "Display Parameters" window).

Measure Cell Allows measurement of cell parameters, but the distance and wavelength must be set correctly.

Circles Puts up resolution circles (or ellipses for a detector swung out on a two-theta arm).

Beam/backstop Will determine the centre of a circle defined by clicking with the mouse at a series of points lying on the circle. After selecting "Fit circles" select a series of points with the mouse and then select "Fit points". The rms fit, radius and centre of the circle will be given. The user has the option to update the direct beam coordinates to the circle centre. Used to determine the direct beam position from wax rings, or to define the (circular) backstop size and position.

SAVE/EXIT This gives the user the option of saving the current processing parameters to a file, then either closing down the display or continuing to process *via* the GUI. If the image was being displayed with the IMAGE keyword, the mosflm prompt will return. If all keywords for processing are given, then MOSFLM will proceed to measure the image(s).

Examining the image after integration

There is the option to update the display after integration (At the bottom of the "Processing Parameters" window, this is one of a number of On/Off or Yes/No toggles). If this option is chosen, once the profiles have been determined and the image integrated, the image plus the predicted pattern will again be displayed again. (Note that there is an overhead associated with this, because the image will have to be read into memory again (unless only a single image is being processed)).

If there are any "bad spots" (*i.e.* poorly measured reflections) on the image a window will be displayed which gives the user the option to examine and/or edit the badspots. If this option is selected, the image will be displayed with a new menu option "Bad spots". Bad spots can then be reclassified as acceptable and other (accepted) spots can be re-classified as rejected.

In addition to the predicted reflections, vectors will be drawn indicating the difference between the predicted and observed spot positions for *FULLY RECORDED* reflections (these vectors are in red....it may be necessary to use the "Clear Prediction" menu option to see them clearly. The vectors can be scaled using the "Vector scale" menu item in the Processing parameters window. Also a minimum intensity threshold for the display of these vectors can be set using the "Threshold" menu item.

The vectors will of course be longer for weak spots than strong ones, but for all reflections the direction of these vectors should be random. If this is not the case, it suggests errors in crystal cell parameters or orientation, or misclassification of fully recorded/partially recorded reflections, or the existence of spatial distortion which is not being correctly modelled.

"Badspots" will be indicated as red crosses, rejected reflections as blue crosses. Rejected reflections will normally be those containing a zero pixel value (because the measurement box extends outside the scanned area of the image plate); these are NOT classified as "badspots". They may also arise if a very large number of background pixels have been rejected.

Overloaded reflections will be indicated as green crosses. Note that you *MUST* include keywords **PROFILE OVERLOAD** in order to estimate the intensities of overloaded reflections by profile fitting.

By clicking on a reflection, the LP corrected profile fitted intensity and standard deviation will be given in the output window.

6.3.2 Processing the first block of data (Non-interactively)

It is usually advisable to process a block of (say) 10 degrees of data prior to processing the complete dataset (as this is quite time consuming) just to check that the processing is satisfactory.

The commands for MOSFLM might now be:

```
TITLE Processing test data

! Crystal parameters
MATRIX oval_2_3.mat
SYMMETRY 96
MOSAIC 0.1

! image parameters IDENT oval
PROCESS 1 TO 10 [ ANGLE 1.0 START 0.0 ]
DIRECTORY /scr0/andrew/
EXTENSION image

! beam parameters
DIVERGENCE 0.35 0.2
POLARISATION MIRRORS

! detector parameters
BEAM 89.33 90.10
BACKSTOP CENTRE 88.5 91 RADIUS 14
GAIN 1.2
DISTORTION ROFF 0.3 TOFF 0.1

! The following are read from the image header if not supplied on
! keywords for most modern detectors, . The phi values (on
! PROCESS keyword) will also be read from the header if not supplied.

DISTANCE 124.1
WAVELENGTH 1.542

!The following are optional, if not given, the program will set
!suitable defaults.
!HKLOUT oval.mtz
!SEPARATION 1.3 1.3
!RASTER 17 17 9 3 3
!GENFILE oval_1to1.gen
!RESOLUTION 3.0

PLOT
RUN
```

See above ([6.3.1 Running MOSFLM interactively](#)) for a description of each of the keywords. The only difference to the commands described above is that the **PROCESS** keyword has been set up to process the first 10 images. The **ADD** subkeyword on the **PROCESS** line specifies that the batch number on the output mtz file should be (1000+image number) (*i.e.* 1001-1010 in this example).

DISTORTION specifies the **ROFF** and **TOFF** values for this scanner (See [Appendix III](#)). It is not normally necessary to specify these values unless they are large (greater than 0.3).

The program will then form the standard profiles by summing reflections over the first block of images 1 to 5 and print the resulting profiles. The number of images in a block is set by the program, but may be set explicitly by the **PROFILE BLOCK** keywords).

6.3.2.1 Formation of the Standard Profiles

The standard profiles are determined in a number of areas across the detector. By default the detector is divided into 9 regions for data to a resolution lower than 2.5Å and 25 regions of which only 22 lie within the active area of a circular detector for resolution higher than 2.5Å. Alternatively the user can define the set of lines parallel to the detector X and Y axes which define the standard areas. This is done with keyword **PROFILE XLINES... YLINES...**

EXAMPLE:

```
PROFILE XLINES 0 45 90 135 180 YLINES 0 45 90 135 180
```

will divide a detector measuring 180x180mm into 16 areas. See MOSFLM Help file for more details.

A separate standard profile is evaluated for each of these areas.

The program prints out some statistics on the standard profiles, followed by statistics on profiles that it has averaged (if any) and followed by a representation of each of the standard profiles using a single character (0 - 9, then A - Z) to represent the value at each pixel (A "[" denotes a negative value). In these representations, a minus sign denotes the background region, and a * denotes rejected pixels. Background pixels which are overlapped by the peak regions of neighbouring spots are automatically rejected by the program. It will also warn you if the peak regions of neighbouring spots overlap.

6.3.2.2 Inadequate profiles

Each standard profile has to satisfy two criteria before it is considered acceptable. There must be at least ten contributing reflections, and the rms variation in the background plane (after rejecting outliers) must be less than 10 (after scaling the profile to a maximum value of 255). If a profile fails to pass either test, then it is averaged using the profiles from neighbouring areas on the detector. Profile averaging should be avoided if at all possible. The averaging inevitably produces a profile that is less broad than the original profile because it is dominated by the stronger, lower resolution data. Look at the printed profiles before and after averaging to confirm this.

Accumulating the profiles over a **BLOCK** of say 10 images (see below) should help provide a sufficient number of reflections, but is unwise to accumulate over too wide a phi range because this will average out any genuine variation in the profiles with phi (*e.g.* due to a change in effective diffracting volume). Both rejection criteria can be changed using subkeywords (**NREF** for number of reflections, **RMSBG** for rms variation in background, on the **PROFILE** keyword line) and it is usually better to avoid averaging by changing these criteria if necessary:

```
PROFILE RMSBG 20 NREF 5
```

6.3.2.3 Spots running into each other

The program will give a warning if it detects that the peak areas of adjacent spots overlap. There are two possible ways around this:

- 1) Increase the **SEPARATION** parameters. Making the separation significantly smaller than the actual spot size in the centre of the image can lead to serious problems and is *NOT* recommended.
- 2) The actual spot size that the program works with when testing for peak overlaps (after rejecting those that are too close as determined by the **SEPARATION** parameters) is determined by the "profile optimisation"...that is when the program works out the best values of the measurement box parameters NC, NRX, NRY, which is done independently for each of the standard profiles. If there is significant diffuse scatter on the image, the "optimised" raster parameters may well produce a peak area that is actually slightly broader than the true Bragg peak and includes part of the "diffuse" peak. This can be checked by examination of the standard profiles...if the peak area contains many pixels with values of 0 or 1 then it suggests the peak is too broad. This effect can be overcome by increasing the **TOLERANCE** parameter (See the help library for more details on how the optimum parameters are derived and the effect of the **TOLERANCE** parameter). The default value for this parameter is 1% *i.e.* 0.01. Increasing it (try steps of 0.005, but it should not be necessary to go above 0.04) will result in a reduction of the optimised peak size. It is up to you to decide what the optimum value is on the basis of the appearance of individual spots in the image.

6.3.2.4 Output Statistics

The other statistics produced are for general information, and are described in the MOSFLM help library under "Output". Probably the most useful is the breakdown of $I/\sigma(I)$ as a function of resolution. This will give an immediate idea of the quality of the data...particularly at the high resolution end. For guidance, a mean $I/\sigma(I)$ of 3.0 will give an R-merge of between 20% and 30% in SCALA. If there are symmetry related fully recorded (or summed partial) reflections on a single image, statistics are also provided on the agreement between their intensities.

Check the following in this job:

- 1) Check the standard profiles look OK (*i.e.* the peak is within the peak region).
- 2) Check the weighted residual is about 1.0.
- 3) CHECK FOR WARNING MESSAGES. These are given at the end of the logfile and in the summary file. They will point out possible problems and suggest a way around them.

6.3.3 Finally, processing the dataset

The whole philosophy of MOSFLM is to allow the entire dataset (or all images obtained from a single crystal) to be processed in a single job. To make this possible, the crystal orientation can be refined continuously for every image, to take account of possible crystal slippage, and the cell parameters can also be refined if the initial estimates are not accurate. An accuracy of 1 part in 1000 or better is required for optimal processing of high resolution data.

There are a large number of adjustable parameters within MOSFLM, but considerable effort has gone into making the program select an appropriate value for these parameters. The program defaults should therefore ALWAYS be used unless there is a very specific reason for changing parameters (*e.g.* if it is suggested in the

warning messages in the summary or logfile)

See [section 9.3](#) for a complete example command file.

[Back to Contents Page](#)

7: Interpreting the output

- 7.1 [The log files](#)
 - 7.2 [The summary file](#)
 - 7.3 [Checking the quality of the data](#)
-

7.1 The log files

The MOSFLM log files can be very long, and to simplify assessing the performance of the processing, the program writes a summary file which contains most of the important information.

However the initial part of the logfile which gives information on the parameters used in the processing (along with the keywords used to change these parameters) should *ALWAYS* be read.

The standard profiles should also *ALWAYS* be checked to ensure that the background mask optimisation has worked correctly (particularly if there is a high degree of diffuse scatter or the spots are very close).

Warning messages are written to the end of the logfile and the summary file if the program detects possible problems. These messages should be acted on where appropriate.

7.2 The summary file

A graphical representation of the information in the summary file can be obtained by running the CCP4 program "loggraph" (followed by the name of the summary file).

The information contained in the summary file is listed below:

7.2.1 Statistics on Processing

IMAGE

The image number (as given on the PROCESS keyword)

CCX,CCY,CCOM

The camera constants (in mm and degrees). The magnitude of CCX and CCY reflects the accuracy of the supplied direct beam coordinates. However their actual magnitude is really immaterial (providing they do not exceed your expected error in the direct beam coordinates) but they should remain *CONSTANT* throughout the data collection (assuming the beam is not moving!). If using the post-refinement option to refine crystal missetting angles, then CCOMEGA may show a gradual drift, since any errors in orientation around the direct beam are taken up by CCOMEGA. From version 6.2.0 this is no longer the case: any change in CCOMEGA is now transferred to an equivalent change in the crystal missetting angles (PHIX, PHIY, PHIZ) and CCOMEGA is reset to zero for each image.

Thus, in the summary file, CCOMEGA will now always be close to zero. Because the orientation around the X-ray beam (determined from spot positions) is generally less well defined than the misetting angles refined by post-refinement, the new refined missets (PHIX, PHIY) may appear to be slightly noisier than when using previous versions of the program. This is artefactual, the actual precision of the crystal orientation is the same.

Note that if the refinement of CCOMEGA is fixed (REFINEMENT FIX CCOMEGA) then this component of the crystal orientation will not be refined. This is NOT recommended.

DIST

The refined crystal to detector distance.

YSCALE

The relative scale factor in the detector Y direction. This should be very close to unity (to within 1 part in 1000 or better). Deviations from unity indicate errors in cell parameters. Note that for the R-AXIS II detector the difference in the pixel sizes in the X and Y directions is taken up by YSCALE, which should have a value close to 1.032.

TILT,TWIST

Deviations from normal incidence on the detector. **TILT** is a rotation about a horizontal axis, **TWIST** about a vertical axis, and the values given are expressed in hundredths of a degree. These values should be close to zero (less than 20) and should be *CONSTANT* (to within experimental error).

ROFF,TOFF

The radial and tangential offsets for the Mar Research and Mac Science scanners only (see [Appendix III](#)) in mm. These should be *CONSTANT* to within experimental error. Watch out particularly for variations in **ROFF** by one pixel. This indicates that the limit switch on the scanner is not functioning correctly (this should only occur very rarely if at all). A Warning is given if this happens.

RESID

The rms positional residual (in mm) after refinement of the detector parameters. For strong images this should be between 0.02 and 0.04. For weak images or large spots (due to high mosaic spread) it can be significantly higher. If partials have to be included in the refinement it will also be higher. A dataset with rms residuals of 0.2 to 0.3 can still give a final Rmerge of under 10% ! Remember a residual of 0.15 is still only one pixel. If the residual is greater than 0.04 for a strong image, there is almost certainly an error in the cell parameters, and they should be refined using the POSTREFINEMENT options. Even for weaker images, the positional residual for the initial refinement (using only central spots) should be small.

WRESID

The weighted residual. This should be close to unity (independent of the strength of the image). Larger values suggest errors in cell parameters or crystal orientation.

FULL,PART,OVRL,NEG

The number of fully recorded, partially recorded and overloaded reflections measured on the image (this includes reflections with part of the measurement box outside the scanned region...these are not actually integrated unless the **PROFILE** **EDGE** keywords are included, in which case those with at least half of the pixels inside the scanned area will be measured by profile fitting). Overloaded reflections are also not integrated (by default) but if the **PROFILE** **OVERLOADS** keywords are included, then their intensity will be estimated by profile fitting the non-overloaded pixels. **NNEG** is the number of reflections with negative (summation integration) intensity. The proportion of the dataset that falls into this category obviously depends on the strength of the images.

BAD

The number of badspots. Reflections will be classified as badspots if they fail any of five criteria: **BGRATIO** greater than 3, **PKRATIO** greater than 3.5 (Fully recorded reflections only, and you have the choice of either rejecting the reflection completely (this is the default) or setting the profile fitted intensity and sd. to the summation integration values (**REJECTION PKRATIO 3.5 ACCEPT**)...see help library), intensity negative and greater (absolute value) than 5 sigma, background gradient too large (background gradient/average background exceeds 0.03), too few background pixels remaining after rejecting outliers (less than 10). All these criteria (except negative intensity) can be changed using the **REJECTION** keyword. There should be less than 5-10 badspots on any one image. If there are more it suggests problems. The badspots are listed in the logfile with the information necessary to see why they have been flagged. It may also be helpful to have the pixel values of flagged spots. These can be obtained by using the keywords: **REJECTION PLOT**. This should enable the source of the problem to be identified.

I/SIGI (two columns)

This gives the average $I/\sigma(I)$ for the whole dataset (first column) and the outermost resolution bin (second column).

Rsym

This gives the R-factor (on intensities) for symmetry related fully recorded (or summed partial) reflections on the same image.

Nsym

The number of reflections (not the number of observations) included in **Rsym**.

SDRAT

The ratio of the observed agreement between symmetry related reflection intensities to their estimated standard deviations. This should have a value of 1.4 if there are two measurements of each reflection, or very close to unity if four or more. This can be more useful than the **Rsym** value, as it should not depend on the intensity of the measurements while **Rsym** will always be higher for weak spots.

7.2.2 Results of the post-refinement

These should be self-explanatory. Check that any change in missetting angles is gradual. Changes greater than about one fifth of the sum of the mosaic spread and beam divergence (*i.e.* typically 0.05 degrees for a good crystal) per image will give rise to errors in the intensities of partially recorded reflections if multiple oscillations have been used in the data collection. There is no way of correcting for this. Changes in cell parameters (if refined) should really not occur (unless the cell genuinely changes with radiation damage). If they do, consider increasing the number of images used in postrefinement (**ADD/WIDTH**). Ideally there should be a data to parameter ratio of 10:1 in the post refinement. If there is not, consider reducing **SDFAC** which sets the $I/\sigma(I)$ criterion for selecting reflections. The refined angular residual should be about one tenth of the summed mosaic spread and beam divergence, although this will depend on the strength of the reflections included in refinement. If the beam parameters are refined, check that they are stable, particularly if they are being used in the reflection list generation (**USEBEAM**). If they are not stable, take an average value and use that in **MOSFLM** and do *NOT* use the refined values in generating the reflection list within **MOSFLM** (*i.e.* include the **POSTREF USEBEAM OFF** combination).

7.3 Checking the quality of the data

Although there are indicators of data quality in MOSFLM (in particular the $I/\sigma(I)$ and R_{sym} values as a function of resolution), the only satisfactory way of assessing data quality is to look at the results of merging measurements of symmetry related reflections using SCALA. Remember that the R-factor alone is not a good indicator, it will always be high for weak data. What is probably more important is the standard deviation analysis at the end of SCALA. If this suggests that the observed agreement is that expected based on the standard deviations (*i.e.* SIGM is 1.0) then you cannot hope to do any better. Inevitably there are errors which are not accounted for in the estimated standard deviation. Thus it is quite normal to have to boost the standard deviations by 20-30% (*i.e.* an SDFAC of 1.2-1.3) to achieve a SIGM value of 1.0. In addition, the agreement is generally worse for the strongest data, so an SDADD of between 0.02 and 0.03 is quite common. If these parameters have to be made significantly greater than this to achieve a SIGM value of 1.0 across the intensity range, then this indicates problems with the processing which should be investigated. One possibility is the presence of a few large outliers, which can destroy the SIGM analysis...look at the monitored reflections for evidence of this. In cases where the crystal has a high mosaic spread, pay particular attention to the partial bias analysis. If this is more than 1-3%, then the mosaic spread has probably been incorrectly defined. In difficult cases it may well be necessary to process a dataset several times (*e.g.* with different mosaic spreads, or different numbers of images included in post refinement or in forming the standard profiles) in order to achieve the best final dataset, but this should be simply a case of using more cpu time, and not require a lot of intervention.

Finally, always consider the possibility that the spacegroup is incorrect!

[Back to Contents Page](#)

8: General tips

- 8.1 [Estimating the GAIN of a detector](#)
 - 8.2 [Processing images with no \(or very few\) fully recorded reflections](#)
 - 8.3 [Processing images when the spots are not fully resolved](#)
 - 8.4 [Processing data from other detectors, or standard detectors with different rotation axis orientation.](#)
-

8.1 Estimating the gain of the detector.

The detector GAIN is the factor that converts counts in the digitised image into the equivalent number of absorbed X-ray photons. It is used to estimate standard deviations, reject outliers etc.

Thus:

$$(\text{value in digitised image}) = \text{GAIN} * (\text{Equivalent number of photons})$$

The simplest way to get an estimate of the gain is to display an image using the IMAGE keyword. Select an area which is free of any diffraction spots, but has a reasonable number of background counts (*i.e.* at least 100 per pixel, say). Drag out a small rectangle in this "spot free" area with the mouse. Then within the "Output" window of the display you will find the entries:

Average	335.5
Rms	19.1

Number

345

(The values will be those for your image of course). Try to get an area with at least 100 pixels ("Number" is the number of pixels in the area you have dragged out), but don't make it too large because what you are looking for here is a "flood field", *i.e.* an area within which you have got a uniform background.

The **GAIN** can then be estimated as:

$$\text{GAIN} = (\text{rms} * \text{rms}) / \text{Average}$$

For the example numbers above this would be 1.09

Try this for a few areas and choose the **LOWEST** value you get (any features in the background, such as diffuse scatter, a weak spot etc, will **INCREASE** the rms, but nothing will **DECREASE** it).

(NOTE: This procedure assumes that the counts in each pixel are independent. For some detectors this will **NOT** be the case, particularly if a small pixel size (100 microns or less) is used with a standard image plate. In these circumstances this will give an **UNDERESTIMATE** of the true **GAIN**.)

This should give a reasonable starting value. If you then process some images, the program calculates a parameter **BGRATIO** which is the ratio of the observed rms variation in the background around spots to what is expected from counting (Poisson) statistics based on the supplied value of the gain. This ratio should be 1.0 if the **GAIN** is correct (and providing the spots are all contained within the peak areas, *i.e.* there is no "diffuse scatter halo" surrounding the spots).

If **BGRATIO** differs from 1.0 by more than 10%, the program gives a "Warning" message. The correct gain can then be estimated as:

$$\text{TRUE GAIN} = (\text{INPUT GAIN}) * \text{BGRATIO} * \text{BGRATIO}$$

IMPORTANT. The **GAIN** should not be interpreted too literally. The way it is derived is only strictly correct if all pixels are independent, which in practice they are not. For this reason, this parameter should **NOT** be used to assess the relative performance of different detectors.

8.2 Processing images with no (or very few) fully recorded reflections

If less than one third of the total number of reflections predicted is fully recorded (this value can be changed with keywords **REFINE FULLFRAC**) then partially recorded reflections are automatically included in the refinement of the detector parameters.

When adding partials (**ADDPART** option, this is **NOT** the default and must be requested by giving the keyword **ADDPART**) then only partials at the end of the oscillation range will be selected and the other "half" of the partial will automatically be added in from the next image, so the centroids of these reflections will be determined as accurately as those of fully recorded reflections (provided the detector origin is indeed fixed).

If this is not the case (*i.e.* by default), partials from both the beginning and end of the oscillation will be selected but only if their degree of partiality is greater than 0.8 (this limit can be changed by giving an appropriate value after the subkeyword **PARTIALS**

e.g. **REFINEMENT INCLUDE PARTIALS 0.5**

will include any partial more than 50% recorded). If insufficient reflections are found for refinement, the limit on the degree of partiality will be relaxed until sufficient reflections are found.

Similarly, if less than one third of the total number of reflections predicted is fully recorded, partials will also be included in forming the standard profiles. Providing the profiles are being accumulated over a number of images, this will not introduce a significant error as both parts of most reflections will be included, so the fully recorded profile will be generated. The inclusion of partials in the standard profiles can be prevented by keywords: **PROFILE FULL**.

8.3 Processing images when the spots are not fully resolved

When adjacent spots are very close the **SEPARATION CLOSE** option for integration should be used (see the help library for details of what this option involves). The program will automatically invoke this option if the **SEPARATION** keyword is **NOT** given and the closest possible spot separation (for any orientation of the crystal) is such that adjacent spots are not separated by more than 2 pixels.

REMEMBER that when using the **CLOSE** option, a scratch file (**COORDS**) is used and this file *MUST* be unique (*i.e.* you cannot run two jobs from the same directory unless **COORDS** is assigned to a unique filename).

Spots which are not completely resolved on the detector or high levels of diffuse scatter may result in a larger peak region in the measurement box than is desirable, due to the algorithm used in optimising the background raster parameters (see [section 6.2.4](#)). If the peak region includes too much of the broad diffuse peak under the Bragg reflection or if it includes the "tail" of a neighbouring spot then it can be made tighter by adjusting the **TOLERANCE**. Try increasing the "inner" tolerance first,

e.g. **PROFILE TOLERANCE 0.02 0.03**

It may also be necessary to specify the minimum spot separation explicitly:

e.g. **SEPARATION 0.95 0.95 CLOSE**

The spot separation can be made as small as the spot size in the centre of the detector (even though the spots are always larger at the outside of the detector due to oblique incidence).

In severe cases, it may be worth trying to reject pixels around the edge of the peak if they do not fit the profiles well (because there is a neighbouring very strong peak). See keyword **PROFILES** subkeywords **WDLIM1 WDLIM2** in the help library for more information.

8.4 Processing data from other detectors

If a commercial detector is being used on a synchrotron beamline, it is not uncommon for the direction of rotation to be different to that used on the commercial instrument. This will not affect autoindexing from a single image, but the resulting orientation will not correctly predict the next image. To correct for this, use the keywords:

DETECTOR REVERSEPHI

If the orientation of the rotation axis has changed by 90 degrees (eg horizontal rather than vertical) this can be allowed for by redefining the **OMEGA** angle (see [Appendix III](#)), using keywords:

DETECTOR OMEGA

For example, the value for OMEGA is 90 degrees for a Mar IP scanner, but if it was being used with a vertical rotation axis of a goniostat, you would need to give the keywords:

```
DETECTOR MAR
DETECTOR OMEGA 0
```

Note that for RAXIS, DIP and LIPS IP scanners there is a specific subkeyword to specify the orientation of the rotation axis.

Default values of OMEGA are 90 for Mar, R-Axis IP scanners, 180 for DIP2000 or 2020 or 2030 (assumed to have a horizontal rotation axis), 270 for the ESRF CCD detector (with image intensifier). If you want to know what the default value of OMEGA is for a given type of detector, add the keyword:

DEBUG CONTROL

Then in the output file immediately after reflecting the DETECTOR keyword the following line of debug will appear:

```
Machine type: RAXI Model type: RAXISIV INVERTX F SPIRAL F ORTHOG T CIRCULAR
OMEGA 180.0
```

It should be possible to process data from other detectors but, depending on how the images are written, it may be necessary to write new code to read in the image. See keyword DETECTORS in the help library for further details.

[Back to Contents Page](#)

9: example command files

- 9.1 [Autoindexing an initial image \(interactively\)](#)
 - 9.2 [Determining an accurate cell](#)
 - 9.3 [Integrating a series of images](#)
-

There follow a number of example command files. In general, it is best to let MOSFLM choose appropriate default values for processing, and to only specify additional parameters if you get warning messages suggesting changes (these warning messages are given in the Summary file and at the end of the logfile).

It is usually *NOT* a good idea to start with someone else's command file, as they will probably have set some parameters which are not appropriate for your data.

9.1 Autoindexing an initial image

This will normally be done interactively. The commands can be put in a file (*e.g.* "runit") and executed by typing @runit at the MOSFLM prompt.

```
=> TITLE My lysozyme data           ! This title is transferred to the
                                     ! MTZ file
```

```
=> IMAGE lyso_001.image [PHI 0 TO 1]      ! Filename of first image. For Mar,
                                           ! ADSC, Mac Science and R-AxisIV images
                                           ! the phi values will be taken from
                                           ! the image header if not given here.
```

```
=> BEAM 150.0 149.0                      ! Direct beam coordinates
```

If not a Mar Research IP scanner (unnecessary for most commercially produced detectors from version 6.2.3):

```
=> DETECTOR RAXISIV                      ! or RAXIS (for RAXIS II) or DIP2000 et
```

If not processing Mar Research, R-axis or Mac Science images:

```
=> WAVE 0.91                             ! For Mar, ADSC, R-AxisIV and Mac Scienc
=> DISTANCE 300                           ! information is taken from the header
                                           ! but can be overwritten using the
                                           ! keywords.
```

```
=> SYMM p43212                           ! If known, give cell and symmetry
=> CELL 79 79 38                          ! otherwise omit completely.
```

Not essential for first stages, but needed for integration

```
=> DIVERGENCE 0.1 0.03                   ! If isotropic, the beam divergence
                                           ! can be included in the mosaic spread.
=> SYNCHROTRON POLARIZATION 0.9          ! Defaults to 0.95 (SRS, Daresbury UK)
=> GAIN 1.7                               ! See section 8.1 for a way to
                                           ! estimate the gain if not known.
=> GO
```

Further options are then usually invoked via the menu of the X-window interface.

9.2 Determining an accurate cell

This example assumes that an orientation matrix has been obtained for the first image, and that accurate cell parameters are to be determined using two (one can use one or more) segments of data. Note that this can be done interactively via the X-windows menu, the example given below is for a background job.

```
ipmosflm spotod /scr0/andrew/fladpx3.spotod \
<< eof-ipmosflm
TITLE Postrefining the cell with two segments
```

```
! Source parameters
```

```
WAVE 0.91                                ! For most modern detectors,
                                           ! this information is taken from the
                                           ! header but can be overwritten using
                                           ! the keywords.
```

```
SYNCHROTRON POLARISATION 0.86           ! Default polarisation is 0.95 (SRS)
DIVERGENCE 0.10 0.02                    ! Horizontal and vertical divergence
DISPER 0.00020                           ! Wavelength dispersion
```

```

! Detector parameters
DETECTOR RAXISIV                                ! If not a Mar Research IP scanner:
                                                ! can be RAXIS (for RAXIS II), DIP2000 etc

BEAM 150.0 149.0                                ! Direct beam coordinates
GAIN 1.2                                          ! Detector gain
DISTANCE 300                                    ! For most modern detectors,
                                                ! this information is taken from the
                                                ! header but can be overwritten using
                                                ! the keywords.
BACKSTOP CENTRE 148 151 RADIUS 12              ! Beamstop shadow

! Crystal parameters
SYMMETRY P212121
MATRIX image_001.mat                            ! orientation matrix for first image
MOSAIC 0.22                                     ! Mosaic spread, should be a reasonable estimate

! Image parameters
IDENT fladpx3                                  ! Sets template for image filenames
DIRECTORY /scr1/andrew/images/                 ! Directory where images are stored

! Processing parameters
POSTREF SEGMENT 2                              ! postrefine using two segments
PROCESS 1 to 4 [START 0 ANGLE 1]               ! Images to be used in first segment.
                                                ! For most modern detectors,
                                                ! the phi values are taken from the
                                                ! header but can be overwritten using the
                                                ! keywords.
NEWMATRIX fladpx3_postref.mat                 ! Filename for output orientation matrix
                                                ! This will contain the refined cell
                                                ! and the refined missetting angles for
                                                ! the first image (1 in this case).

G0
PROCESS 86 to 89 [START 85 ANGLE 1.0]         ! Images to be used in second segment
! MATRIX image_086.mat                        ! If necessary (i.e. crystal has slipped) can
                                                ! specify an orientation matrix for the first
                                                ! image of the second run.

G0
eof-ipmosflm

```

Possible additional keywords:

1) If there is significant diffuse scatter in the image, or if the mosaic spread is very large (greater than 0.5 degrees) it is usually best to limit the post-refinement to using reflections that are nearly half-recorded, using the FRMIN, FRMAX keywords. This will make the refinement less dependent on the model of the rocking curve:

e.g. POSTREF FRMIN 0.4 FRMAX 0.6

2) If there is an ice ring or spots on the image, all spots within a specified resolution limit can be rejected.

e.g. RESOLUTION EXCLUDE 3.66 3.72

9.3 Integrating a series of images

This example assumes that an accurate cell has already been obtained (using a POSTREF SEGMENT run) so no further refinement of cell parameters is required. Note that integration can be done interactively via the X-windows menu, the example given below is for a background job.

Note that from version 6.2.3, it is unnecessary to delete the "SPOTOD" file as it will be automatically deleted when the program exits.

```
ipmosflm spotod /scr0/andrew/fladpx3_1to90.spotod \
summary fladpx3_1to90.sum \
coords /scr0/andrew/fladpx3_1to90.coords \
<< eof-ipmosflm
TITLE Postrefining the cell with two segments ! Source parameters
WAVE 0.91 ! For most modern detectors, this !
! information is taken from the header !
! but can be overwritten using the !
! keywords.
SYNCHROTRON POLARISATION 0.86 ! Default polarisation is 0.95 (S !
DIVERGENCE 0.10 0.02 ! Horizontal and vertical divergence !
DISPER 0.00020 ! Wavelength dispersion
! Detector parameters
DETECTOR RAXISIV ! If not a Mar Research IP scanner !
! can be RAXIS (for RAXIS II), DI !
BEAM 150.0 149.0 ! Direct beam coordinates
GAIN 1.2 ! Detector gain
DISTANCE 300 ! For most modern detectors, this !
! information is taken from the header !
! but can be overwritten using the !
! keywords.
BACKSTOP CENTRE 148 151 RADIUS 12 ! Beamstop shadow
! Crystal parameters
SYMMETRY P212121
MATRIX fladpx3_postref.mat ! orientation matrix previous job
MOSAIC 0.22 ! Mosaic spread, should be a reasonable value
! Image parameters
IDENT fladpx3 ! Sets template for image filenames
DIRECTORY /scr1/andrew/images/ ! Directory where images are stored
! Processing parameters
POSTREF FIX ALL ! do not refine cell, only crystal
PROCESS 1 to 90 [START 0 ANGLE 1] ADD 1000 ! Images to be integrated.
! For most modern detectors,
! the phi values are taken from the header
! but can be overwritten using the
! keywords. The batch numbers in
```

! MTZ file will be the image number
! (ADD keyword).

HKLOUT fladpx3_1to90.mtz

! Name of output MTZ file.

GO

END

eof-ipmosflm

Delete temporary files

/bin/rm /scr0/andrew/fladpx3_1to90.spotod

/bin/rm /scr0/andrew/fladpx3_1to90.coords

/bin/rm fladpx3_1to90.gen

Possible additional keywords:

PROFILE TOLERANCE 0.02 0.03

PROFILE XLINES 0 75 150 225 300 YLINES 0 75 150 225 300

SEPARATION 0.75 0.75 CLOSE

AUTOINDEX

CELL KEEP 283.09 107.60 139.65 90.000 90.000 90.000

REJECTION PKRATIO 4.0

RESOLUTION EXCLUDE 3.66 3.72

RESOLUTION ANISOTROPIC 3.5 2.4 2.3

TWOTHETA 15

BEAM SWUNG_OUT 99.96 124.5

LIMITS RSCAN 147.0

These are described below.

1) Optimising the Standard Profiles

Depending on the strength of the images, the degree of diffuse scatter, the spot separation on the images, the crystal mosaicity etc, it may be necessary to adjust the PROFILE TOLERANCE parameters to get well defined standard profiles. The appearance of the standard profiles should always be checked in the logfile, to ensure that adjacent spots are not included in the "peak" region, that long diffuse tails are not being included in the peak, and that not too many profiles are being averaged (see below).

e.g. PROFILE TOLERANCE 0.02 0.03

The first value is used for profiles near the centre of the image, the second value for profiles at the outside, and an interpolated value for profiles in between. Defaults are 0.01 0.01 for a lab source (wavelength 1.542) and 0.01 0.03 for a synchrotron.

It should not normally be necessary to use values above 0.04.

For weak images, you may find that some profiles are being averaged. This is to be avoided if possible. Consider if you are trying to integrate beyond the true resolution limit of the crystals. If not, try to avoid the averaging by one of the following:

- i) If profiles are being averaged because there are very few reflections, but the reflections are reasonably strong so that the profiles look OK, reduce the minimum number of reflections required (default 10):

e.g. PROFILE NREF 5

ii) If the profile is being averaged because the rms variation in the background (after scaling the peak to 255) is too large, but in fact this is because there is significant diffuse scatter which should not be included in the Bragg peak, then increase the allowed rms variation (default 10.0):

e.g. PROFILE RMSBG 20.0

iii) Try setting up fewer standard profiles, by defining the regions on the detector where the profiles are to be set up using the XLINES, YLINES keywords:

e.g. PROFILE XLINES 0 75 150 225 300 YLINES 0 75 150 225 300

will give 16 areas over which a profile will be determined.

2) Specifying the minimum spot separation

The program will automatically determine the minimum allowed spot separation based on the size of spots in the centre of the first image to be processed. Spots closer than this will not be integrated. However, if the spots are very close together, the minimum spot separation determined by the program may be too conservative and result in many spots being rejected. To avoid this, set the minimum spot separation explicitly. Note that in such cases the CLOSE option for spot integration should be used (see the Help library under SEPARATION for further details of this option). The program also decides if the CLOSE option need to be invoked, again based on the very first image to be processed. It is a good idea to ensure that if the CLOSE option is used for one segment of data (eg the 90 images processed in the job above) then it is also used for all other data from this crystal, by specifying its use explicitly.

e.g. SEPARATION 0.75 0.75 CLOSE

or if just wanting to enforce the use of the CLOSE option:

SEPARATION CLOSE

3) Refining cell parameters during integration

In this case, replace the POSTREF FIX ALL keywords with

POSTREF WIDTH 10

where WIDTH specifies the width (in phi) of data to be used in the refinement. For trigonal or higher symmetry a few degrees of data is usually sufficient, for lower symmetries at least 10 degrees is usually necessary. THIS IS NOT RECOMMENDED: It is usually preferable to determine the cell initially using the POSTREF SEGMENT option.

4) Determining the orientation as part of the job

It is possible to invoke the autoindexing as the first part of a processing job. By default the indexing will be done on the first image to be processed, but the images to be used can be specified explicitly (See section 3.2). In these circumstances the unit cell derived from the autoindexing will be used in the integration, unless the KEEP subkeyword is given with the CELL. Usually an accurate cell (from post-refinement) will be available

and should be used in the integration:

```
AUTOINDEX [ IMAGE 1 2 3 ]
CELL KEEP 283.09 107.60 139.65 90.000 90.000 90.000
```

5) Specifying the size of the integration box

The overall dimensions of the integration or measurement box are determined automatically based on strong spots from the first image to be processed. If you want to make the box smaller or larger than this, specify its size explicitly and tell the program not to change it:

```
RASTER 25 25 17 3 3
PROFILE FIXBOX
```

The corner and rim parameters (17 3 3 in above case) will still be optimised, but the overall dimensions will not (use PROFILE NOOPTIMISE (ATALL) to suppress optimisation).

6) Excessive numbers of "Bad spots"

Excessive numbers of "Bad spots" is usually a sign that the processing is not as good as it should be and there are errors in the cell parameters, crystal orientation, mosaic spread or in the detector itself. However, if the cause of the bad spots cannot be corrected, it may be preferable to change the rejection parameters to avoid too many reflections being rejected. This is particularly true if a large number of strong reflections are being rejected because of a poor profile fit, because this could have a significant effect on Patterson based methods used in Molecular Replacement, definition of solvent masks etc.

e.g. REJECTION PKRATIO 5.0

This should ONLY be done as a last resort, it is much better to find the cause of the rejections. See REJECTION keyword in Help library for more details.

7) Ice rings

If there is an ice ring or spots on the image, all spots within a specified resolution limit can be rejected.

e.g. RESOLUTION EXCLUDE 3.66 3.72

8) Anisotropic diffraction

If the crystal diffraction is very anisotropic, an anisotropic resolution limit can be applied.

e.g. RESOLUTION ANISOTROPIC 3.5 2.4 2.3

It should be noted that refinement programs which use maximum likelihood targets may object that there are many missing data if this option is used. In practice, it seems best to use it for postrefinement of parameters but to include all reflections to the isotropic resolution limit for integration (see (14) below).

9) Inclusion of partially recorded reflections in refinement and profiles

The program will decide automatically if partials should be included in the refinement of the detector parameters and formation of the standard profiles. If however you want to explicitly force the program to do so use the following keywords: **REFINEMENT INCLUDE PARTIALS** for detector refinement **PROFILE PARTIALS** for profile formation. To explicitly **EXCLUDE** partials from profile formation use: **PROFILE FULLS**

10) Processing data from an offset detector (in two-theta)

To process data collected from an offset detector, the swing angle needs to be specified, and the direct beam coordinates should be those corresponding to a two-theta value of zero, unless the **SWUNG_OUT** subkeywords are given. The areas for the standard profiles *MUST* be specified explicitly using the **PROFILE** keyword.

```
TWOTHETA 15 BEAM SWUNG_OUT 99.96 124.5 ! Direct beam position on swung out
                                           ! detector
                                           !
PROFILE XLINES 0 50 100 150 200 YLINES 0 50 100 150 200
```

11) Excluding a "shadow" around the rim of the image

The **LIMIT RSCAN** keywords can be used to limit the "active" area of a circular detector, or **XSCAN** and **YSCAN** for an rectangular detector. All these limits are measured from the physical centre of the detector.

e.g. **LIMITS RSCAN 146.5**

12) Excluding more general "shadows"

The **NULLPIX** keyword can be used effectively to deal with more general shadows, for example if a solid backstop support is used. All pixels with a value less than or equal to that specified by the **NULLPIX** keyword will be treated as lying outside the active area of the detector, and any spots containing these pixels will be rejected.

In addition, rectangular areas of the detector can be excluded from spot finding and integration by use of the **LIMIT EXCLUDE** keywords. See the helpfile for details.

13) Resetting the saturation point of the detector

If there is evidence that the detector is saturating before the default cutoff value, it can be reset with the **OVERLOAD CUTOFF** keywords:

e.g. **OVERLOAD CUTOFF 75000**

14) Limiting the resolution

Normally the resolution limit is set by the physical dimensions on the detector. This can be overridden by the **RESOLUTION** keyword:

e.g. **RESOLUTION 3.5**

Both inner and outer resolution limits can be set:

e.g. RESOLUTION 20.0 3.5

The resolution can also be determined by the quality of the data, so that the resolution limit is determined by the mean value of $I/\sigma(I)$ within a resolution shell.

e.g. RESOLUTION CUTOFF 5.0

[Back to Contents Page](#)

Appendix I - Setting the measurement box parameters manually

IMPORTANT

The program now automatically optimises the measurement box parameters so normally it is not necessary to optimise them manually. For difficult cases the optimisation can be turned off (PROFILE NOOPTIMISE) and the manual approach used.

Following 2 rounds of refinement, the program will display the average spot profile for the spots used in the refinement. At this stage the user can adjust the dimensions of the measurement box, and the average spot profile will be recalculated. This can be done as many times as necessary. Note that a value of zero must be entered as 99. When choosing the optimal box parameters the following points should be kept in mind.

- i) The accuracy of the weakest reflections will be determined by the accuracy of determining the local background. There should therefore be at least as many pixels in the background as in the peak (and preferably more, although beyond a factor of 2 there is little additional gain). The program has a sophisticated algorithm for rejecting pixels from the background if neighbouring spots intrude into part of the background region, so this should not necessarily enforce the use of a small measurement box. If possible, however, enlarge the box in the direction of greatest spot separation.
- ii) Systematic errors in the integrated intensity will result if the spot extends out of the peak region of the measurement box. This should be avoided if at all possible.
- iii) The inclusion of what are actually background pixels within the peak region of the box will degrade the summation integration estimate of the intensity (because they introduce only noise, no signal). However they will have almost no effect on the profile fitted intensity, so there is very little penalty involved. It is much more important to get all the spot within the peak region than to keep background out of the peak region.
- iv) The processing time is approximately directly proportional to the number of pixels in the measurement box, so if this is a consideration, don't make the box too large (but remember point (i) above).
- v) The positional residual will increase as the peak size is increased, but this really doesn't matter.

When you are happy with the measurement box parameters, just give a carriage return to the query on changing the measurement box parameters, and the program will go on to do the second round of positional refinement using reflections from the entire detector (rather than just the centre). The residual vectors are plotted as before. The program asks if you wish to repeat this refinement. Even if you say No, it may repeat it anyway if the residual is too high. It will then go on to integrate all the spots on the image. As it does so, it will display the residual vectors between observed and calculated positions for all fully recorded reflections

above some threshold intensity. Finally it gives some statistics on the measured intensities, writes the intensities back to the generate file, and returns the prompt:

==>

[Back to Contents Page](#)

Appendix II - Overview of MOSFLM

1) Introduction

The MOSFLM program is designed to facilitate processing of rotation data collected on either image plate, CCD or film. The suite originates from the MOSCO system developed in Cambridge by Nyborg and Wonacott for use on a PDP 11/10 ([Nyborg & Wonacott, 1977](#)) but it has been extensively developed since that early version, primarily at Imperial College by A.J. Wonacott, P. Brick and A.G.W. Leslie, and more recently at LMB. In particular the much greater memory and cpu resources available in current machines have been exploited (the first version ran on a machine with 28 Kbytes of memory, the current version uses 5 Mbytes of memory just to store digitised images). All necessary processing steps are now performed by a single program which incorporates routines for indexing, refinement, integration and display of results.

The basic procedure for data processing is independent of the type of detector (film, image plate or CCD) although there are a number of useful features which are only available for image plate data (particularly automatic updating of cell parameters and crystal orientation).

The MOSFLM program

a) Overview.

MOSFLM performs the actual integration of the reflection intensities. It generates the reflection list, reads the digitised image, integrates the spots and writes the intensities and standard deviations into the generate file and mtz file. The image plate version has the additional capability of refining the crystal orientation and cell parameters during data processing using the intensities of partially recorded reflections in the same manner as the POSTCHK program. The program can be run interactively making use of the graphical output options (X-windows only at the moment), which is most useful when first characterising a new crystal or when dealing with pathological cases.

Routine processing is generally done as a background job.

The normal operation of the program can be broken down into several steps as outlined below. A summarised flow diagram is given in Figure 2.

- ii) Generation of a reflection list for this image using the latest refined values for crystal orientation, cell parameters and beam parameters. Crystal orientation can be refined using a pattern matching procedure.)
- iii) Location of diffraction spots in the central region of the image and refinement of the detector parameters using the observed positions of these spots. Determination of an average spot profile optimisation of measurement box parameters.

iv) Location of diffraction spots in the outer regions of the detector and further refinement of detector parameters.

v) The pixel values corresponding to the measurement boxes for all reflections are extracted from the digitised image and written to a scratch file. If addition of partially recorded reflections is being done (this is *NOT* the default) then only partials at the end of the oscillation range are chosen, and the pixel values for the other "half" of the partial on the NEXT image are added to those of the current image before writing the pixel values to the scratch file. Thus the partial addition is done within **MOSFLM**, rather than at the **SCALA** stage if the **ADDPART** option is used. This is only valid if the effective exposure time of both images is the same (*e.g.* the data is collected in the "dose" mode rather than simply by time), and the origin (*i.e.* direct beam position) and orientation of every image is identical. This is *NOT* recommended, partly as the trend is towards collecting fine phi-sliced datasets which have reflections spread over several images.

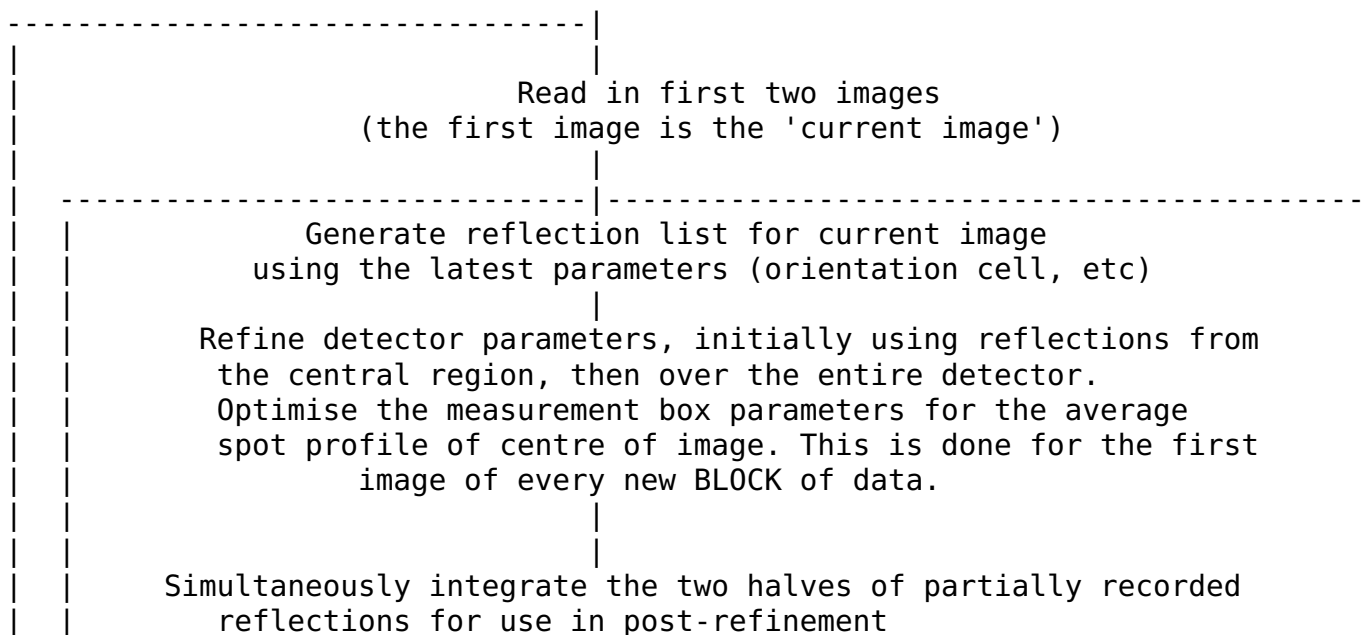
vi) If post-refinement of the crystal orientation, cell parameters or beam parameters is to be carried out, then the intensities of partially recorded reflections occurring at the end of the oscillation range of the current image and the start of the next image are evaluated as the measurement boxes (for the current image only) are written to the scratch file in step (v) (these intensities are evaluated by summation integration rather than profile fitting). Providing that data over the required angular range is available, post-refinement of the crystal parameters is then carried out.

vii) Steps (i) to (vi) are repeated for all the images to be used in forming a set of standard profiles for evaluation of the reflection intensities. Thus the scratch file will finally contain the measurement boxes of a number of different images.

viii) The standard profiles are evaluated for several different regions of the detector using the reflection data accumulated in the scratch file. These profiles are then used to evaluate the reflection intensities for each image, and the profile fitted and summation integration intensities and standard deviations are written back to the generate file.

ix) Steps (i) to (viii) are repeated until all images have been processed.

A more detailed description of each of these steps follows.



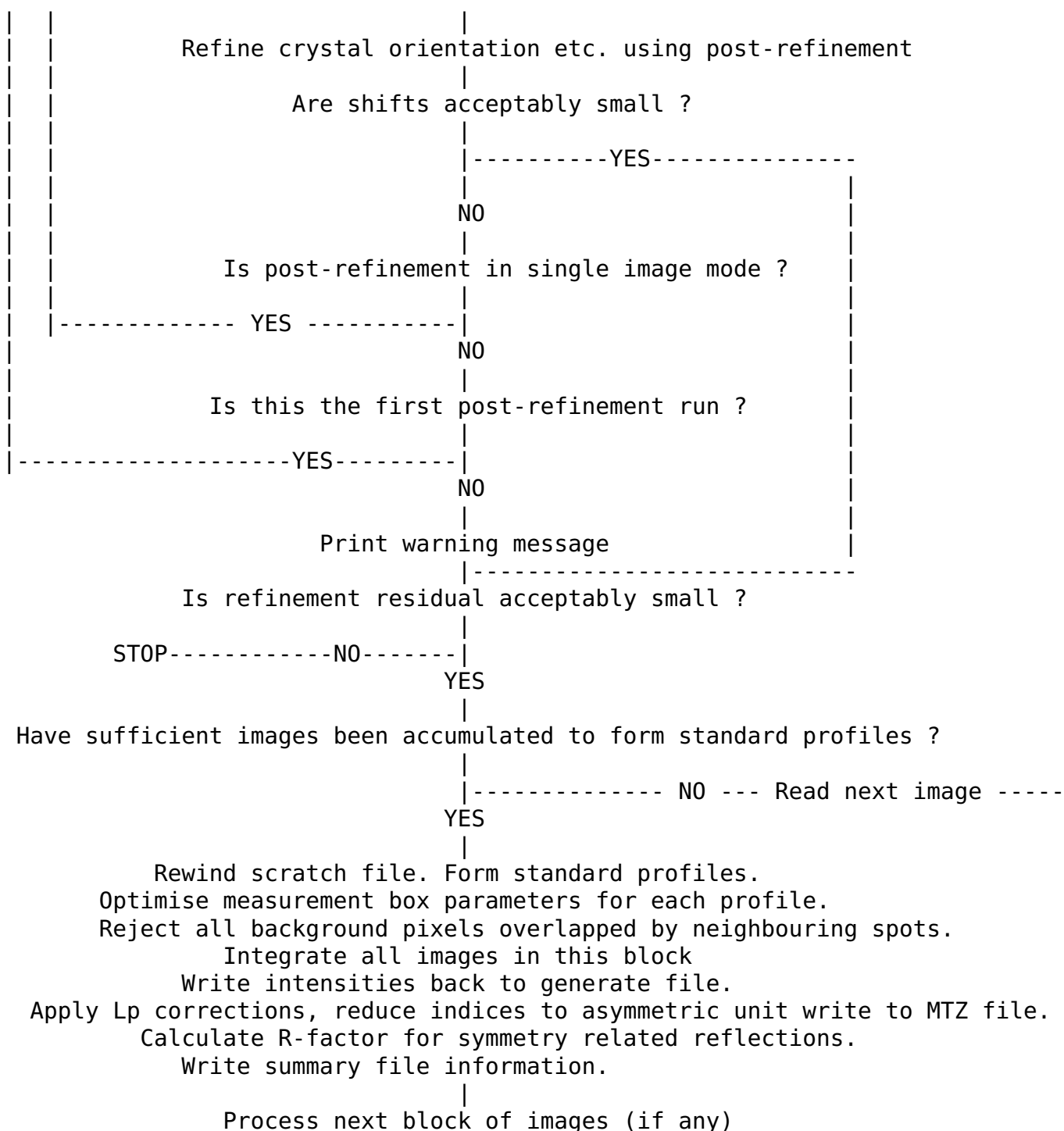


Figure 2. A Flow diagram of the operation of MOSFLM when using the post- refinement option.

c) Generation of the reflection list.

Generation of the reflection list is performed using the Reeke algorithm.

d) Initial refinement of detector parameters.

The first step in the refinement of the detector parameters is the location of up to 60 suitable reflections. Generally, only fully recorded reflections are selected, as the position of the centroid of a partially recorded reflection will depend on its degree of partiality. For crystals with a high mosaic spread, there may be too few fully recorded reflections to allow a satisfactory refinement, so partially recorded reflections may have to be used. Similarly, overloaded reflections will have a poorly determined centre of gravity, although for very intense images it may be necessary to include some overloaded reflections.

The refined parameters are:

- i) The crystal to detector distance (XTOFD).
- ii) The position of the centre of the diffraction pattern (XCEN,YCEN).
- iii) A relative scale factor applied to the Y coordinates (YSCALE).
- iv) Small rotations of the detector about a horizontal axis (TILT) and a vertical axis (TWIST)
- v) Rotation of the detector about the X-ray beam direction (OMEGA)
- vi) Radial (ROFF) and tangential offsets (TOFF) (Mar scanners only, see below)
- vii) ONLY if explicitly requested (REFINEMENT FREE keywords), the amplitude of the radially dependent radial and tangential offsets RDROFF, RDTOFF in mm (Mar scanners only).

Further details of the refinement procedure are given in [Appendix III](#).

Refinement is carried out for a fixed number of cycles, and is followed by a display of the average spot profile of the reflections used in refinement (but not partials) if requested. If the program is being run interactively and automatic measurement box optimisation is NOT being performed, the user has the option to update the parameters defining the measurement box at this stage.

If insufficient reflections are found for refinement (*i.e.* less than 20), then the program will automatically attempt to find additional reflections in a number of different ways. If the majority of reflections are too weak, then the threshold will be reduced (but only down to one sigma). If the majority of reflections are overloaded, it will allow the use of overloaded reflections. If there are fewer than 20 non-overloaded reflections in the central region of the detector, then the size of this region will be enlarged by 10mm (in X and Y). If the final residual is greater than a preset limit, processing will be abandoned.

e) Location of reflections in outer region of the detector and further refinement.

Following successful refinement using reflections from the inner region of the detector, a list of all fully recorded reflections outside this central region is prepared (partials will be included if requested). After sorting this list on the detector X coordinate, it is divided into 8 bins with an equal number of reflections in each bin. Within each bin, reflections with an $I/\sigma(I)$ value greater than a preset limit are chosen until a maximum of 5 reflections have been found. If less than 30 reflections are found in total, the $I/\sigma(I)$ cutoff is automatically reduced and the search repeated (to a minimum $I/\sigma(I)$ cutoff value of 2). These reflections are combined with 20 reflections selected from the central region for final refinement of the detector parameters. The final rms positional residual (given in mm) will depend on the spot size, the strength of the image and the accuracy of the cell parameters, but typical values are 0.02-0.03mm for a reasonably strong image and up to 0.07-0.08mm for relatively weak images. Larger values suggest an error in cell

parameters, which can usually be corrected using the post-refinement option. If partial reflections are included in the refinement the positional residual will be significantly larger (*e.g.* 0.06-0.07mm for a strong image, 0.10-0.15mm for a weak image).

g) Extracting the measurement boxes from the digitised image.

A final list of all reflections to be integrated is prepared and sorted on the detector X coordinate. Using a circular buffer and working through the digitised image stripe by stripe, the pixel values corresponding to the measurement boxes of these reflections are accumulated and written to a scratch file.

The size of the peak area of the measurement box is expanded automatically (by 2 pixels at a time to maintain an odd number of pixels) in order to allow for the increase in spot size due to obliquity of incidence of the diffracted beam on the detector. This is a function of the spot size (determined by collimation and mosaic spread of the crystal), the effective detector thickness and the Bragg angle.

If post-refinement is to be carried out (image plate data only), then for each reflection that is partially recorded at the end of the oscillation range of the current image, the identical pixels on the next image in the series are also accumulated. This requires both images to be stored in memory for the old-style postrefinement; **POSTREF MULTI**, the current default, stores the necessary information in arrays. In either case, this is only practical if the detector has a fixed origin and orientation (*e.g.* the Mar Research IP scanner) so that the predicted position of the partially recorded reflection is identical in the digitised image on both images. It is unclear at present whether images scanned using an off-line image plate scanner will give sufficiently reproducible results for the method to be applicable. The measurement boxes for the two halves of the partial are then integrated (using summation integration) to give the intensity of the two components, which in turn gives an observed degree of partiality:

$$P(\text{obs}) = (I_{\text{curr}} / (I_{\text{curr}} + I_{\text{next}}))$$

where I_{curr} is the intensity on the current image and I_{next} is the intensity on the next image. In order to minimise the errors due to the finite pixel size, the peak area of each measurement box is interpolated onto a grid centred on the calculated position of the reflection, using linear interpolation. If this is not done then the standard profiles, which are formed by averaging all fully recorded reflections within a limited area of the detector, will be artificially broadened because the true reflection position can be up to half a pixel away from the centre of the measurement box (in any direction). The interpolation procedure significantly improves the fit of the profiles, particularly for image plate data, although in this case the linear interpolation procedure is not ideal in view of the very large dynamic range, which can lead to local gradients of 20,000 counts per pixel for strong reflections. The summation integration intensity is unaffected by the interpolation.

h) Post-refinement of crystal orientation, cell parameters and beam parameters.

The degree of partiality of a partially recorded reflection is a function of the crystal orientation, cell parameters and beam parameters (mosaic spread and horizontal and vertical divergence of the X-ray beam). These parameters can therefore be refined using the observations $P(\text{obs})$ described in section (g) if there is a model for the rocking curve, that is, the relationship between the distance of a reciprocal lattice point from the Ewald sphere at the end of the rotation and the resulting intensity of the reflection on that image expressed as a fraction of the total intensity. This rocking curve is therefore used to convert the observed partiality $P(\text{obs})$ to the position of the reciprocal lattice point at the end point of the rotation. It is this conversion which is a function of the mosaic spread and beam divergences. The position of the reciprocal lattice point is then a function of the crystal orientation and cell parameters. It should be noted that this is independent of the

reflection coordinates on the detector or the crystal to detector distance.

When data from a single pair of images is used to refine crystal orientation, there is essentially no information on the orientation of the crystal around the X-ray beam, and only the missetting angles around the rotation axis (PSIZ) and the direction normal to the rotation axis and X-ray beam (PSIY) are refined. Following refinement these missetting angles are converted back to the standard missetting angles PHIX, PHIY, PHIZ which describe the crystal orientation at $\phi=0$. In general, if the crystal symmetry is lower than trigonal, not all the cell parameters will be well defined using data from one pair of images (the spacing along the X-ray beam direction is poorly defined) and in such cases it is advisable to use an angular wedge of data (typically 5-10 degrees in ϕ) accumulated over a number of successive images. Even then, some cell parameters may not be well defined (large standard deviations and shifts) and in such cases individual cell parameters can be fixed if desired. However it is preferable to use the POSTREF SEGMENT option to get an accurate cell in advance of the processing run in these cases.

When an angular wedge of data is to be used, cell parameter refinement is delayed until sufficient images have been processed (the crystal orientation is still refined after every image). After the first refinement, if there are large shifts in any of the refined parameters the processing will be restarted at the first image using the updated parameter values. This can be repeated a number of times, allowing a reasonably large radius of convergence from inaccurate starting parameters. Post-refinement is then carried out after every image, by deleting the data from the image with the smallest ϕ value and adding in the data from the latest image. When a single pair of images is used for the refinement, then a large shift in missetting angles or cell parameters will result in the reprocessing of that image.

The post-refinement procedure is, not surprisingly, most successful with reasonably strong, high resolution, data. For weak data at low resolution, the correlation between different parameters in the refinement can lead to instability which manifests itself as unrealistically large variations in cell parameters and missetting angles. For example, in the case of monoclinic data collected by rotating about the unique axis, there is a strong correlation between the monoclinic beta angle, the missetting angle around the rotation axis, ϕ_{HIZ} , and the relative values of the a and c cell parameters. In these cases data from a number of "blocks" of images at different ϕ values should be used to obtain accurate cell parameters prior to processing, and the cell parameters should be kept fixed during the processing (POSTREF SEGMENT option)

i) Formation of the standard profiles.

Normally a number of different standard profiles will be determined for different areas on the detector. These areas are by definition rectangular, and are defined by a set of lines parallel to the detector X and Y directions.

By default for high resolution data (greater than 2.5Å) the program will generate 6 lines in each direction, giving $5 \times 5 = 25$ standard profile areas of which 4 will be outside the circular resolution limit, leaving 21 active profiles. For resolutions below 2.5Å it will generate 4 lines giving $3 \times 3 = 9$ standard profiles. However the user may define his own areas on the detector by supplying the coordinates (in mm) of the defining lines (XLINES, YLINES).

The overall measurement box size for each of the standard profiles will be determined by expanding the peak region of the measurement box determined for the central region of the detector to allow for the increase in spot size due to obliquity (this depends on the Bragg angle and the nominal detector thickness). However the optimal background parameters (NRX, NRY, NC) will be determined separately for each of the standard profiles.

The standard profiles themselves are determined by summing all fully recorded reflections above a certain threshold intensity. When measuring data from crystals with a very large mosaic spread, there may be very few fully recorded reflections on each image. In these circumstances there is the option to add together the two components of a partial at the stage when the measurement boxes are written to the scratch file (see [Fig. 2](#)). In this way reflections which are actually partially recorded at the end of the rotation range of the current image will be treated as if they were fully recorded on the current image, and in particular these reflections are used in forming the standard profiles. This procedure depends on successive images having the same exposure time and incident flux. This threshold is defined in terms of the average, background subtracted, peak pixel value (over 9 central pixels) exceeding the rms variation in background by a specified factor, usually 2. A best least-squares background plane is fitted to the summed reflections (see below for details of background outlier rejection). The resulting pixel values, after scaling the central pixel value to 10,000, is taken as the standard profile. Two criteria are applied to determine whether the resulting profile is satisfactory. First, there is a minimum acceptable number of reflections that contribute to the profile (10) and secondly the rms variation in the background plane (after scaling) must not exceed a specified limit. If a particular profile fails either test, then it is improved by adding in the summed reflections from adjacent regions on the detector. This will normally produce an acceptable profile, but if it does not, then no profile fitting will be attempted for that region of the detector. Clearly the accuracy of a given profile will depend on the number (and intensity) of the reflections contributing to that profile. The program therefore allows the standard profiles to be accumulated over a number of successive images, rather than forming them on an image by image basis. Typically, the profiles will be accumulated over 5 to 10 images. This significantly improves the signal to noise in the profiles, but it would be unwise to average over too many images as the profiles themselves may vary slightly with rotation angle as a result of changes in the diffracting volume of the crystal or radiation damage, which can lead to an increase in mosaic spread and hence spot size.

j) Reflection integration.

Following formation of the standard profiles, the reflections on all images contributing to the profiles are integrated. The first step in this procedure is fitting the best background plane to the background pixels in the measurement mask. In order to deal with outliers (due to cosmic rays, spots from a satellite crystal etc) in the background region of the reflection, the initial background plane is determined using a fraction (between 0.5 and 1.0) of the total number of background points, selecting those pixels with the lowest values. The constant component of the background plane is then adjusted to correct for the systematic bias introduced by selecting the lowest pixel values (assuming a Gaussian distribution). This plane is then used to reject outliers, which are defined as pixels with values which deviate from the plane by more than a fixed number (usually 3) standard deviations, where the standard deviation is based on Poissonian counting statistics. The same procedure is applied to determining the centre of gravity of spots used in positional refinement (sections **d** and **e**). The profile fitted intensities and standard deviations are evaluated using weighted profile fitting and methods based on those originally described by [Rossmann \(1979\)](#). For every reflection, a new profile is evaluated by linear interpolation of the standard profiles of the regions surrounding that reflection. The interpolation is based on the distance of the reflection from each standard profile, where the coordinates of the profiles are calculated as the intensity weighted mean of all the reflections contributing to that profile. For most reflections, four different profiles will contribute to the variable profile, but for reflections near the periphery of the detector there may be only three or two contributing profiles, while no sensible interpolation can be done for the outermost reflections. This procedure provides a more accurate modelling of the way in which the profile varies across the face of the detector.

The summation integration intensity and standard deviation is also evaluated for every reflection, and both profile fitted and summation integration intensities and standard deviations are written back to the generate and mtz files.

Individual reflections are flagged as 'badspots' (which are rejected by MOSFLM) if they fail any of the following tests:

- 1) The rms fit of the background plane must not exceed a factor (3) times the variation expected on the basis of counting statistics.
- 2) The intensity should not be negative with an absolute value greater than 5 standard deviations.
- 3) The fit of the profile should not be worse than a factor (3) times the expected fit based on counting statistics. If a reflection fails this test, only the profile fitted intensity (not the summation integration intensity) will be rejected.
- 4) The background plane gradient must not exceed a preset value (by default, spots with a ratio of the either background plane gradient (a or b) to the average background (*i.e.* a/c or b/c) greater than 0.03 will be rejected.
- 5) The reflection must not contain more than a specified number of saturated pixels (*i.e.* be overloaded). The intensities of these reflections can be estimated by profile fitting if requested, using only that part of the peak which is not saturated in fitting the standard profile.

Typically only a handful of reflections (between 0 and 5) will be rejected on any image. Larger numbers are indicative of problems in processing, *e.g.* poor profile fit due to inaccurate cell parameters.

2) Orientation refinement by pattern matching.

This option permits the refinement of the crystal orientation by automatically optimising the fit between the predicted and observed patterns, by maximising the predicted intensity (*i.e.* finding the orientation for which the sum of the intensities of the predicted reflections is a maximum). The basic algorithm is identical to the 'convolution' procedure described by [Rossmann \(1979\)](#). There are minor differences, the most significant being:

- a) The total predicted intensity is calculated while stepping the missetting angle (either PSIIY or PSIZ) through a fixed number of angular steps (typically 9) centred on the current value of that angle. The best value for that parameter is then chosen as the midpoint of the two steps at which the intensity has dropped by 15 standard deviations from its value at the centre of the range (*i.e.* the current value). This is in contrast to Rossmann's procedure which simply chooses the value at which the intensity is at a maximum. This makes the refinement less susceptible to errors introduced by statistical fluctuations in the total summed intensity. The step size is then multiplied by a damping factor (normally 0.25) and the grid search repeated.
- b) When generating the reflection list for different values of PSIZ (say) with a view to optimising the summed intensity, the beam divergence in the horizontal direction is increased by an amount equal to the largest step applied to PSIZ, so that the search is insensitive to errors in PSIIY. Similarly, when PSIIY is being refined, the vertical divergence is increased by the same amount.

With these modifications, the algorithm will converge from an initial orientation up to 2-3 degrees in all three missetting angles for problems with cell parameters between 50 and 100Å. However, the advent of autoindexing algorithms has made this option essentially obsolete.

[Back to Contents Page](#)

Appendix III - Coordinate systems used by MOSFLM

- a) Detector Coordinates
 - b) Camera Constants
 - c) Coordinate Transformations
 - References
-

a) Detector Coordinates

The coordinates systems used by MOSFLM are illustrated in Fig. 3. The crystal orientation is defined relative to the camera coordinate frame (X,Y,Z), with X along the X-ray beam (away from the source), Z along the rotation axis, and Y forming a right-handed set. A positive phi rotation is anticlockwise viewed from +Z looking towards the origin. The orientation matrices (A and U) are defined relative to this axial system.

Coordinates of reflections are calculated in the detector frame (Xd, Yd), which has its origin at the point of intersection of the direct beam and the detector, with Yd parallel to the camera Z axis and Xd parallel to the camera Y axis.

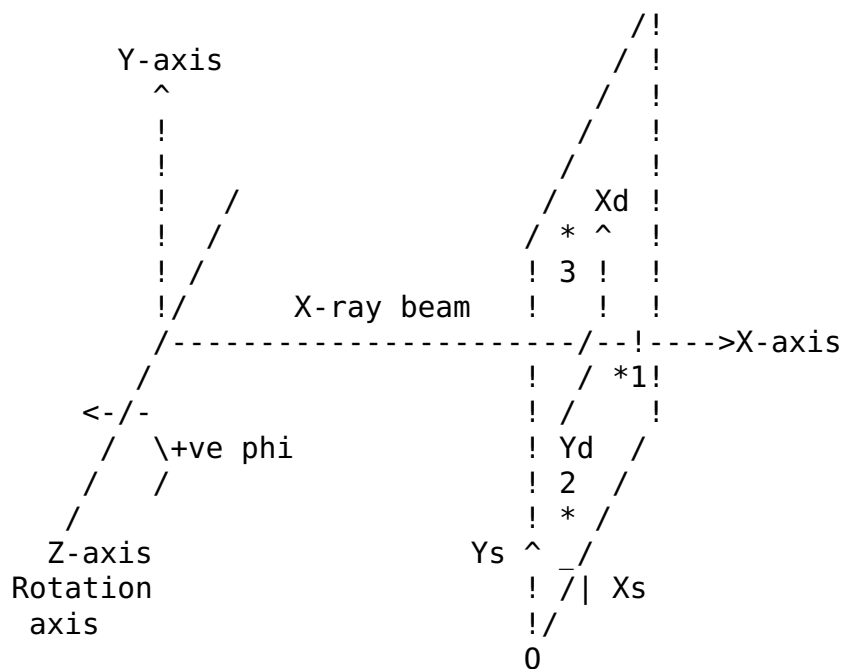


Figure. 3

For film data, there is the option to scan the film in an orientation corresponding to a 90 degree clockwise rotation from that shown in figure 3, in which case Xs is (ideally) parallel to Xd and Ys is (ideally) parallel to Yd.

The coordinates of the direct beam position in the scanner coordinate frame (Xs, Ys) are denoted XCEN, YCEN. If the angle between Xd and Xs is denoted by OMEGA, then for image plate data the ideal value of OMEGA is 90 degrees. For film data scanned in the normal orientation (as in Fig. 3), OMEGA may deviate

slightly from 90 degrees as its precise value will depend on how the film is placed on the scanner. For film scanned in the alternative orientation, OMEGA will be close to zero.

Finally, the pixels of the digitised image have coordinates (X_s , Y_s). For image plate data and films which have been scanned on a microdensitometer in the normal orientation (that is, with X_d around the circumference of the scanner drum, Y_d along the axis of the drum), the origin of this coordinate frame is the lower left corner of the image viewed from behind the detector looking towards the X-ray source (Fig. 3). Thus the scanner axis X_s is (ideally) antiparallel to the detector axis Y_d , while scanner axis Y_s is (ideally) parallel to detector axis X_d .

The coordinate systems used by MOSFLM.

X, Y, Z are an orthogonal frame centred at the point of intersection of the X-ray beam and the rotation axis. Positive ϕ rotation is anticlockwise as viewed down the Z axis towards the origin (as indicated). The ideal detector coordinate frame (X_d, Y_d) has its origin at the point of intersection of the detector plane and the X-ray beam. Y_d is parallel to the rotation axis and X_d orthogonal to it. The scanner coordinate frame (X_s, Y_s) has its origin at the lower left corner of the image (camera-mans view, *i.e.* looking towards the source). If necessary (as it is for images from the Mar scanner) the image will be inverted left to right when read by the program so that internal to MOSFLM the first pixel in the digitised image is in this corner. However this is entirely hidden from the user, and in the display, for example, the first pixel in the image (with pixel coordinates 1,1) will be in the lower right corner of the displayed image.

Y_s is then the most rapidly changing direction in the digitised image, X_s the more slowly changing direction.

b) Camera Constants

i) Image plate or CCD data

The coordinates of the direct beam position $XCEN$, $YCEN$ in the scanner coordinate frame must be supplied by the user. Any deviation in the refined position of the centre of the diffraction pattern from these coordinates are denoted by the camera constants CCX , CCY . It should be noted that CCX , CCY are defined in the detector frame (X_d, Y_d), NOT in the scanner coordinate frame. Thus a positive CCX represents a displacement along $+Y_s$, while a positive CCY is a displacement along $-X_s$. Any deviation of the angle OMEGA from its expected value of 90 degrees is referred to by the camera constant $CCOMEGA$. The camera constants allow for errors in the user defined position of the direct beam (CCX and CCY) and in the alignment of the scan direction of the image plate relative to the camera (and detector) axes ($CCOMEGA$).

ii) Film data

For film data the camera constants CCX , CCY denote the deviation of the refined position of the centre of the diffraction pattern from the midpoint of fiducial marks 1 and 3 (Fig. 3), as appropriate for an Enraf Nonius oscillation camera. They are expressed in the detector frame (X_d, Y_d), so that they are independent of the orientation of the film on the scanner. The line joining fiducial marks 2 and 3 is assumed to be parallel to the detector axis X_d , and any deviation is denoted by the camera constant $CCOMEGA$. Note that for film data, the value of OMEGA will not necessarily be 90 degrees (normal orientation) or 0 degrees (rotated orientation) because it will depend on exactly how the film is placed on the scanner. The orientation of the film on the scanner does NOT affect the relative orientation of the line joining fiducials 2 and 3 and the detector axis X_d at the time of collecting the data. The camera constants allow for errors in the positioning of the cassette on

the carousel and misalignment of the camera itself.

c) Coordinate Transformations

The basic transformation between the calculated detector coordinates of the reflections (X_d , Y_d) and the scanner coordinates in the digitised image (X_s , Y_s) is given by:

$$X_s = XCEN + XTOFRA * (X_d \cos(\text{OMEGA}) - Y_d \sin(\text{OMEGA}))$$

$$Y_s = YCEN + XTOFRA * YSCALE * (X_d \sin(\text{OMEGA}) + Y_d \cos(\text{OMEGA}))$$

where:

$XCEN$, $YCEN$ are the coordinates of the centre of the diffraction pattern (the direct beam position) in the scanner coordinate frame.

$XTOFRA$ is a refined scale factor which accounts for variations in crystal to detector distance. Its nominal value is 1.0, as the actual crystal to detector distance is implicit in the detector coordinates.

$YSCALE$ is a further refined scale factor defining the relative scales along the Y_s and X_s axes. For image plate data $YSCALE$ should be unity, but for film data the pixel size around the circumference of the drum is increased slightly due to the finite thickness of the X-ray film (by about 0.2%), so $YSCALE$ should have a value close to 0.998.

$OMEGA$ is the (refined) angle between X_d and X_s .

For image plate data, an additional four parameters have been introduced to allow for errors in the alignment of the camera and distortions introduced by mechanical misalignment of the scanner. $TILT$ and $TWIST$ allow for the detector plane being non-normal to the incident X-ray beam, and correspond to a rotation about an axis parallel to the camera Z axis and camera Y axis respectively (*i.e.* horizontal and vertical axes in the conventional arrangement). The effect of these rotations on spot positions can be adequately modelled (for alignment errors of a few degrees or less) by a variation in crystal to detector distance that is dependent on X_d ($TILT$) or Y_d ($TWIST$). Thus $XTOFRA$ is replaced by:

$$XTOFRA' = XTOFRA + TILT * X_d + TWIST * Y_d$$

When expressed in this way, $TILT$ and $TWIST$ when divided by the crystal to detector distance are the actual rotations (in radians). The refined values printed by the program are expressed in hundredths of a degree.

For the Mar Research image plate scanner, it is assumed that the first pixel of the spiral scan (at the outer edge of the image plate) is at a known radius from the centre of the image plate. Any error in the presumed radius of this first pixel will result in a (fixed) radial offset in the positions of all reflections. This is modelled by a refineable radial offset ($ROFF$). A second assumption is that the radial scan is indeed exactly radial, *i.e.* would pass through the centre of rotation of the image plate. If this is not the case, to a good approximation it will result in a (fixed) tangential offset in the position of all reflections ($TOFF$). (Note that this assumption would break down for very small radii, but as the inner radial limit of the scan is 10mm this is not a problem in practice).

It has been found that ROFF and TOFF show a radial dependence that is significant in some 30cm Mar scanners.

This has been modelled as a sine-wave dependence, giving:

$$ROFFTOT = ROFF + RROFF * \sin(\pi * R / RSCAN)$$

$$TOFFTOT = TOFF + RTOFF * \sin(\pi * R / RSCAN)$$

Where RSCAN is the scanner radius and R is defined below.

Defining:

$$X_m = X_d \cos(\Omega) - Y_d \sin(\Omega)$$

$$Y_m = X_d \sin(\Omega) + Y_d \cos(\Omega)$$

$$\Psi = \arctan(Y_m / X_m)$$

then the transformation, including the distortion components, is given by:

$$X_s = X_{CEN} + (X_{TOFRA} + TILT * X_d + TWIST * Y_d) * X_m + ROFFTOT * \cos(\Psi) - TOFFTOT * \sin(\Psi)$$

$$Y_s = Y_{CEN} + (X_{TOFRA} + TILT * X_d + TWIST * Y_d) * YSCALE * Y_m + ROFFTOT * \sin(\Psi) + TOFFTOT * \cos(\Psi)$$

These equations are used in the least squares refinement of the detector and distortion parameters, using the measured centres of gravity of the selected refinement spots (in the scanner coordinate frame) as the observed positions.

For film data, three distortion parameters are refined. The first two, TILT and TWIST are identical to those used for image plates. The third parameter is to allow for bulging of the X-ray film in the cassette. This is modelled as a radially dependent variation in the crystal to detector distance, corresponding to a conical distortion which is radially symmetric.

Using X_m and Y_m as defined above, and defining in addition:

$$R = \sqrt{X_d^2 + Y_d^2}$$

the transformation for film data becomes:

$$X_s = X_{CEN} + (X_{TOFRA} + TILT * X_d + TWIST * Y_d + BULGE * R) * X_m$$

$$Y_s = Y_{CEN} + (X_{TOFRA} + TILT * X_d + TWIST * Y_d + BULGE * R) * YSCALE * Y_m$$

Refinement proceeds in an identical manner to image plate data, with the exception that as the BULGE parameter is poorly defined when using only data from the central area of the image, it is held fixed at its input value during the initial refinement.

The refinement of crystal to detector distance (as X_{TOFRA}) and the YSCALE parameter allow for errors in crystal cell parameters to be at least partially compensated when predicting reflection positions. This can be

important in cases where no accurate values are initially available, as the post-refinement option for refining cell parameters is of course only available once the images have been measured.

References

A. Jones, K. Bartels & P. Schwager (1977) in 'The Rotation Method in Crystallography', U.W. Arndt & A.J. Wonacott, eds, North Holland Publishing Co.

W. Kabsch (1988) J. Appl. Cryst., 21, 67-71.

W. Kabsch (1993) J. Appl. Cryst., 26, 795-800.

I. Steller, R. Bolotovskiy & M.G. Rossmann (1998) J. Appl. Cryst. 30, 1036-1040

A.G.W. Leslie (1990) in 'Crystallographic Computing', Oxford University Press.

J. Nyborg & A.J. Wonacott (1977) in 'The Rotation Method in Crystallography', U.W. Arndt & A.J. Wonacott, eds, North Holland Publishing Co.

M.G. Rossmann, A.G.W. Leslie, S. Abdel-Meguid & T. Tsukihara (1979), J. Appl. Cryst., 12, 570-581.

M.G. Rossmann (1979), J. Appl. Cryst. 12, 225-238.

F.K. Winkler, C.E. Schutt & S.C. Harrison (1979), Acta Cryst. A35, 901-911.

A.J. Wonacott (1977) in 'The Rotation Method in Crystallography', U.W. Arndt & A.J. Wonacott, eds, North Holland Publishing Co.

Harry Powell Last modified: Fri Oct 17 15:30:52 BST 2008