



# THE *HKL* MANUAL

A Description of the Programs

*Denzo*

*XDisplayF*

*Scalepack*

An Oscillation Data Processing Suite for Macromolecular Crystallography

by Daniel Gewirth

written with the cooperation of the program authors  
Zbyszek Otwinowski and Wlodek Minor  
revised and updated by Wladyslaw Majewski

Edition 6 – September 2003

# Table of Contents

<b>Table of Contents</b>	<b>1</b>
<b>Preface</b>	<b>3</b>
<b>Introduction</b>	<b>5</b>
<b>Data Collection Process</b>	<b>7</b>
Experimental Procedure	7
Detector	8
Summary	10
<b><i>The Denzo Manual</i></b>	<b>11</b>
<b>Denzo for Dummies</b>	<b>13</b>
<b>A Denzo Session</b>	<b>16</b>
<b>Displaying the Image Using XdisplayF</b>	<b>17</b>
Visualisation of the diffraction space	22
Starting Denzo	22
<b>Getting that First Orientation: Indexing</b>	<b>23</b>
Indexing	23
<b>Refinement: Using the Fit Command</b>	<b>32</b>
Mosaicity and Spot Shape	35
How it works: Integration of the Diffraction Maxima	37
Ice Rings	39
Refinement Statistics	40
Finishing Up	41
Understanding the DENZO Log File	42
Understanding the DENZO .x file	44
<b>Input Formats and Conventions</b>	<b>46</b>
<b>Denzo Orientation Conventions</b>	<b>47</b>
Detector Conventions	47
Camera Conventions	48
Determining the values of FILM ROTATION and Y SCALE	49
<b>Alphabetical Denzo Keyword Descriptions</b>	<b>50</b>
<b>Denzo Appendixes</b>	<b>75</b>
Keywords by Category	75
Default Parameters	76
Recommended Hardware	77
Manual Indexing of Diffraction Images	78
Film Cassette and Off-line Image Plate Users	81
Detector Specifications	82

<b>The SCALEPACK Manual</b>	<b>87</b>
<b>Program Description</b>	<b>88</b>
Statistics and Scalepack	89
<b>Implementation</b>	<b>93</b>
Limits on variable dimensions	93
Platforms	93
<b>Scenarios</b>	<b>94</b>
List of Scenarios	94
Scenario 1: Scale Data from One Crystal Processed with Denzo	95
Scenario 2: Scale Data from One Crystal not Processed with Denzo	97
Scenario 3: Heavy Atom Search: Scale native plus a few frames of a potential derivative	98
Scenario 4: Scale Multiple Native Data Sets Together. Data from Denzo	100
Scenario 5: Reindexing	102
Scenario 6: Comparing two data sets from Scalepack	111
Scenario 7: Compare I+ and I- : detection of anomalous signal	112
Scenario 8: Fitting polarization	114
Scenario 9: High resolution + low resolution data sets from one crystal	115
Scenario 10: Thin or few frames (SCALE RESTRAIN and B RESTRAIN)	116
Scenario 11: Space Group Identification	117
<b>Alphabetical Scalepack Keyword Descriptions</b>	<b>121</b>
<b><i>Q &amp; A</i></b>	<b>141</b>
2 $\Theta$	142
Absorption Correction Factors	142
Auto editing	142
Autoindexing	143
Box definition	145
CCP4 Scaling	145
Computer problems and questions	146
Consequences of Profile fitting errors in Denzo	148
DENZO output	149
Display	151
Example of Bad Detector Pathology	151
Ignore	153
Mosaicity	153
Negative Intensities	154
Other	154
Overload	156
Refinement	156
Rhombohedral	158
Twinned Crystals	158
<b>Index</b>	<b>159</b>

# *Preface*

The **HKL** suite is a package of programs intended for the analysis of X-ray diffraction data collected from single crystals, and consists of three programs: *XdisplayF* for visualization of the diffraction pattern, *Denzo* for data reduction and integration, and *Scalepack* for merging  
~~Denzo~~ scaling of the data

audience for whom this manual was written, and at whom the barbs are directed, is the author himself."

## HKL user support

Questions about the programs described here should be directed to Zbyszek Otwinowski or Wladek Minor at the addresses below. You may also use `hkl@hkl-xray.com` or `news://bionet.xtallography` as a medium for user's ideas exchange):

mailing list:	<code>majordomo@hkl.hkl-xray.com</code>
Information and user support:	<code>hkl@hkl-xray.com</code>
<b>HKL WWW homepage:</b>	<b><a href="http://www.hkl-xray.com">http://www.hkl-xray.com</a></b>
Zbyszek Otwinowski:	<code>zbyszek@mix.swmed.edu</code>
Wladek Minor:	<code>wladek@iwonka.med.virginia.edu</code>

## How to reference *Denzo* and *HKL*

If you want to **reference** the HKL package for publication, use the following citation:

Z. Otwinowski and W. Minor, " Processing of X-ray Diffraction Data Collected in Oscillation Mode ", Methods in Enzymology, Volume **276**: Macromolecular Crystallography, part A, p.307-326, 1997,C.W. Carter, Jr. & R.M. Sweet, Eds., Academic Press



The authors encourage referring to the use of the programs when describing structural work. This is particularly significant when statistics from data reduction are being quoted, for example, resolution limit, signal-to-noise ratio, or  $R_{\text{merge}}$ . The package, when properly used, gives a realistic estimate of measurement uncertainties. By default, the  $R_{\text{merge}}$  statistic in does not reject observations. In contrast, most other programs define the  $R_{\text{merge}}$  statistic with a large number of observations deleted from  $R_{\text{merge}}$ . This difference in the definition of  $R_{\text{merge}}$  (a more detailed discussion is in the program description) between programs frequently is very large. Quoting  $R_{\text{merge}}$  statistics without referring to the program / definition used does not properly describe the crystallographic experiment.

## History

Until 1988 and were developed at the University of Chicago, from 1989 through 1994 at Yale University, and currently at the University of Texas, Southwestern Medical Center at Dallas. was developed at Purdue University until 1995, and currently, at the University of Virginia, in Charlottesville.

## Acknowledgments

We would like to thank Janet Smith, Anne Bloomer, Alan Friedman, John Tesmer, Iwona Minor, Bob Sweet for comments and contribution to programs and documentation, Michael Sintchak for the first UNIX port of programs *Denzo* and *Scalepack*, Paul Sigler for his encouragement in this project, and Michael Rossmann for providing the initial stimulus and ideas for the first implementation of the programs. We also would like to acknowledge Janet Smith for the initial version of the *Scalepack* manual.



Every January in Daresbury Laboratory the CCP4 Study Weekend is held. Proceedings, distributed by The Librarian, Daresbury Laboratory, Daresbury, Warrington WA4 4AD, UK, contain comprehensive coverage of the X-ray diffraction data analysis, especially the volume "Data collection and Processing" (January 29-30, 1993) compiled by L. Sawyer, N. Isaacs, and S. Bailey.

The HKL package can estimate Bragg intensities from single-crystal diffraction data that are recorded on a position-sensitive X-ray (also potentially neutron-diffraction or electron-diffraction) detector, for example, film, IP scanners, or charge-coupled device (CCD) area detectors. The programs allow for data collection by oscillation, Weissenberg, and precession methods. The detector can be either flat or cylindrical. The detector readout can be either rectilinear or spiral, although spiral coordinates must be converted to rectilinear before processing. The package allows for random changes in the position and the sensitivity of the detector between consecutive exposures. The package features profile fitting, weighted refinement, eigenvalue filtering, and a universal definition of detector geometry.

is a program which takes raw X-ray diffraction data, for example, from a scanned piece of film or imaging plate, and reduces it to a file containing the h, k, l indexes and background and L-P corrected intensity of the spots on the image, along with an estimate of the error. In this way the useful information is extracted from the image. Depending on the number of spots on the image, a typical 7 to 18 MB imaging plate raw data file (or, in case of the new RAXIS-IV scanner and 50 micron scan, the size of the image is 72 MB!!) will yield a reduced file of approximately 0.1 to 2.0 MB.

### Advantages of using *Denzo*

to process raw X-ray diffraction images:

1. The use of a **flexible, weighted, profile fitting** algorithm for measuring the intensity of the spots provides highly accurate data processing.
2. Crystal and detector parameters can be refined in any order and combination.
3. The initial **indexing procedure is automatic** and uses only a single still or oscillation image.
4. **Partially recorded reflections are used in the positional refinement** of the crystal orientations, for high accuracy and computational stability. In addition, , when combined with the companion program , accurately tracks and sums up partially recorded reflections over an essentially unlimited number of frames. This provides the user with the option of collecting thin oscillation frames or extracting usable data from crystals of high mosaicity.

To " " a data set involves two separate steps. The first is to determine the precise crystal and detector orientation parameters for a single image in the collected series. The intrinsic crystal parameters include lattice type and unit cell dimensions. The crystal orientation parameters describe the orientation of the reciprocal lattice with respect to the spindle, beam, and vertical axes of the camera. The detector and X-ray parameters include the wavelength, the crystal to detector distance, the precise coordinates of the direct beam, the detector missetting angles, and the internal scanner alignment parameters, such as the non-orthogonality of the scanner head motions or the scale of the y pixels to the x pixels of the scanned image.

The second step in ' ing' a data set is to process all of the images in the series using the parameters determined from indexing one frame in the series. The user has full control over which parameters are refined in every cycle. The strategy in refinement, therefore, is to start at low resolution, adjusting just a few parameters, and then to extend the refinement to a higher resolution, where more parameters can be fitted.

Once a data set is processed with , the output is ready to be scaled using the companion program, . Together these programs make up a useful package for the reduction and merging of X-ray diffraction data.

# Data Collection Process

A recent review of basic problems of practical crystallography is given in Jan Drenth's "Principles of X-ray Protein Crystallography" Springer-Verlag 1994 (ISBN 0-387-94091-X)

Every element of the data collection process must function close to its optimum in order for one to solve a macromolecular structure. The sheer amount of collected data makes computer programs an inevitable intermediary between the researcher and the experiment. The package provides several levels of insight into the data at each stage of the measurement and data analysis process:

- , which provides statistics for the full data set;
- , which provides numerical analysis of one oscillation image; and
- , which presents data visually, up to the single pixel level.

Different problems manifest themselves most clearly at different levels of analysis. The traditional method of judging the success of the experiment by the final statistics (e.g. from ) is not sufficient, since it does not show if the experiment was done optimally. The biggest problem with final statistics is that they do not differentiate well among the sources of problems, and often come too late to fix them. Therefore, the experimenter must be aware of how the detector, the X-ray beam, the crystal, and the procedure all contribute to the final data quality and how each of them can make the experiment a failure.

## Experimental Procedure

In the traditional approach, one collects data first and then starts analyzing the results. This strategy has a risk that there may be a gross inefficiency in the setup of the experiment, for example: the data set may be incomplete, the reflections may overlap, the zones may overlap, a large percentage of the reflections may be overloaded, etc. At that stage the only solution is to repeat the experiment, which may be difficult with unique crystals or with experiments that require a synchrotron source.

Data collection is best performed as a highly interactive process. Immediate data processing, which the authors encourage, provides fast feedback during data collection. Most macromolecular crystallographic projects go through iterative stages of improving crystal quality and data collection strategy. Typically, most of the data collection time and effort is spent before the optimal point is reached. Then, if data collection is going well, there is pressure to use the expensive detector and X-ray beam resource efficiently.

The three basic questions are:

1. whether to collect?
2. what to collect?
3. how to collect?

### 1. Is the data worth collecting?

Quick scaling of a partial data set collected in the first minutes may eliminate the need to collect a full set of non-derivative data. Observing many diffraction spots in the image encourages one to collect a full data set; however a high number of spots may be due to high mosaicity, making such a data set unprocessable. One image is enough to index it, estimate mosaicity, and note the severity of the problem caused by overlapping between the reflections. If the Bragg peaks are not resolved, there is no point in collecting such data, despite the number of spots in the image.



## 2. What range of data to collect?

Typically, one wants to collect up to the resolution limit, defined by the ratio of average intensity to  $\sigma$  (noise) being about 2. The safest way to establish this, is by processing a test image, rather than by guessing. One has to note that some space groups have inherent ambiguities in indexing, which only scaling of the initial image to the previously collected data can resolve. Otherwise, one risks recollecting already measured indexes rather than filling in the missing data.

## 3. How to collect data?

The detector should be placed as far away from the sample as possible to cover the desired resolution limit. Long unit cells, large mosaicity, or large oscillation range all affect spot separation and potential overlaps. Some overlaps are immediately visible - the ones arising from a long unit cell axis in the plane of the detector. At high resolution, due to the weakness of the spots, the overlaps may be less obvious. The simulation of a diffraction pattern, based on indexing of the first image and proposed data collection protocol, is the right tool to define an adequate oscillation range and correct detector placement. There is no particular need to collect fully recorded reflections, so the optimal oscillation range is typically narrow, or even equal to a fraction of the crystal mosaicity.

## Detector

Problems are best diagnosed by collecting data with benchmark, high quality crystals, e.g. tetragonal lysozyme. **There is no particular advantage of lysozyme** crystals, with the possible exception of how easy it is to grow them, and a larger unit cell crystal would be preferable, e.g. **tetragonal chymotrypsin**.

One should expect very high quality data from test crystals. The resulting anomalous-difference Fourier map should identify all the sulfurs in the protein. The detector parameters refinement should produce a **very small spread** (tens of microns, hundredths of a degree) from one image to another. Such a test may require the mounting of a test crystal in a way that avoids slippage and minimizes absorption.  $R_{\text{merge}}$  statistics in the range 2-3%, based on high redundancy (4 fold or higher) and high resolution (2Å or better) should be expected. Only very few (less than 0.1%) outliers should be found during merging.

Results worse than the above, indicate a problem with the test crystal or with the experimental setup. Preferably, the test crystal should be kept at 100K to minimize radiation damage. Problems with the test crystal may mask detector problems. For instance, slippage of the test crystal makes it very difficult to notice a spindle motor backlash or malfunctioning of the X-ray shutter.

## Benchmarks and Calibration

Many macromolecular crystallography labs have not developed stringent benchmarks of acceptable performance. The value of such a test depends on how it is analyzed. Frequently, lack of rigor leads to the acceptance of many outliers in the test data. The **outliers rejection** in the merging of symmetry-related data is a valid statistical procedure, but it should be applied with great caution. The definition of an outlier is "a large, but sporadic fluctuation in the data", for example, due to a cosmic radiation hit. A small number - less than 0.01% - of outliers is something to be expected, even in a well functioning system. However, the practice of many labs has been to accept a much larger number of outliers, even as high as 10%. Many serious problems may be masked by such a liberal outlier rejection. It should be emphasized that outlier rejection always improves the consistency (including consistency indexes, e.g.  $R_{\text{merge}}$ ), but not necessarily the correctness of the merged data.

If the detector marginally passes a standardized test, it may fail during data collection. For example, if the test data are collected in the large oscillation-angle mode, a shutter opening delay or spindle motor backlash may affect fewer partials than if the data are collected in the narrow frame mode. **It is dangerous to accept results from the test with a significant number of reflections flagged as outliers**, even if the  $R_{\text{merge}}$  statistics seem to be good. This is almost a sure sign of a **problem**, and unless the problem is well understood, it may not be sporadic when one collects data to solve a crystal structure. One way to attempt to understand the nature of outliers is to locate them in the detector space in order to identify the problem. The clustering of outliers in one area of the detector may indicate a damaged surface; if most outliers are partials, it may indicate a problem with spindle backlash or shutter control. The zoom mode may be used to display the area around the outliers in order to identify a source of problem, for example the existence of a satellite crystal or single pixel spikes due to electronic failure. Sometimes a histogram of the pixel intensities may suddenly stop below the maximum valid pixel value, indicating a saturation of the data acquisition hardware/software.

### Detector geometry

A correct understanding of the detector geometry is essential for accurate positional refinement. Unfortunately, **most detectors deviate from perfect flat or cylindrical geometry**. These deviations are detector specific. The primary sources of error include:

- **misalignment** of the detector position sensors (Mar, R-axis),
- **non-planarity** of the film or IP during exposure or scanning,
- **inaccuracy of the wire placement** and distortions of the positional readout in multi-wire proportional counters (MWPC),
- **optical distortion** (which can also be due to a magnetic field acting upon the image intensifier) in the TV or CCD based detectors.

➡ **If the detector distortion can be parametrized, then these parameters should be added to the refinement.** For example, in the case of the spiral scanners there are two parameters describing the end position of the scanning head. In the perfectly adjusted scanner these parameters would be zero. In practice, however, they may deviate from zero by as much as 1 mm. Such misalignment parameters can correlate very strongly with other detector and crystal parameters. If the program does not have the ability to describe detector distortions, then the other parameters, such as the unit cell and crystal-to-detector distance, will be systematically wrong.

With films and IPs handled manually in cassettes, like at many synchrotrons, the biggest problem lies in keeping the detector flat during exposure and subsequent scanning. In the manual systems, it is much harder to model the possible departures from ideal flat or cylindrical geometry, and makes only limited attempts to correct such distortions. **Non-ideal film or IP geometry is one of the main factors behind the variable quality of data collected with the manual systems.**

### X-ray beam

See Arndt U.W. "X-ray Collimation and Generation" in "Data Collection and Processing", Daresbury 93

The main properties of the X-ray beam that need to be checked are: stability, focus, angular spread, and wavelength in the case of MAD experiments. Large **fluctuations in beam intensity** produce a variable background intensity and variable scale factors during scaling. The quality of the **beam focus** is immediately visible in the spot profile of low resolution reflections. The **angular spread of the beam** contributes to the reflection width, and it may introduce overlaps between reflections for crystals with very long unit cells. The beam properties (except stability) are best analyzed by the inspection of images. The beam parameters are less significant for crystals with large mosaicity.

## Summary

Macromolecular crystallography is an iterative process. Rarely do the first crystals provide all the necessary data to solve the biological problem being studied. Each step benefits from the experience learned in previous steps. To monitor progress, the package provides two tools:

- a) **Statistics** - both weighted ( $\chi^2$ ) and unweighted ( $R_{\text{merge}}$ ) are provided. The Bayesian reasoning and multi component error model facilitates obtaining proper error estimates.
- b) **Visualization** of the process plays two roles: it helps the operator to confirm that the process of data reduction, including the resulting statistics, is correct, and it allows one to evaluate problems for which there are no good statistical criteria. Visualization also provides confidence that the point of diminishing returns in data collection and reduction has been reached. At that point the effort should be directed to solving the structure.

The methods presented here have been applied to solve a large variety of problems, from inorganic molecules with 5Å unit cells to a rotavirus crystallized in the form with cell dimensions: 700x1000x1400Å (Temple B., Harrison S. C., private communication). The overall quality of the method has been tested by many researchers in successful applications of the program to MAD structure determinations.

# THE *Denzo* MANUAL

A Description of the Program

An Oscillation Data Processing Program for Macromolecular Crystallography



# Denzo for Dummies

If you have a good crystal and just want to index your frames without knowing anything, follow this strategy. This is not the only way to process images with `denzo`, nor it is necessarily the best, fastest, or most efficient, but it covers the basics. Sorry, but if your problem is more complicated or you have data from a non-automatic scanner, you may have to read the rest of this manual. Basically, you will be using as many defaults as possible. Hopefully someone has installed the programs on your computer and set up the proper aliases.

<b>Step 1</b> see	Load your frames onto a local disk (for fast access). Put all your files in the same directory. Make sure you have the four processing command files called <code>site.dat</code> , <code>myexperiment.dat</code> , <code>auto.dat</code> , and <code>refine.dat</code> . <code>Site.dat</code> should have been set up for your X-ray and detector system.
<b>Step 2</b>	Sit down at your fast computer workstation (e.g. Alpha, Indigo, etc.). Open two windows (I call them <i>ImageWindow</i> and <i>DenzoWindow</i> ) and in each window set your directory to the directory where you are going to store the results from your processing.
<b>Step 3</b>	Edit the file <code>myexperiment.dat</code> to update the experimental parameters: distance, oscillation range and start, space group or lattice, sector range, mosaicity, and file names. If you don't know your space group or lattice, enter P1 there. Edit file <code>refine.dat</code> to include the refinement resolution ranges.
<b>Step 4</b> see appendix: _____	<p><b>ImageWindow:</b> <code>xdisp &lt;format&gt; xtal###.osc 1</code> _____ launches the display program</p> <p><b>DenzoWindow:</b> <code>denzo</code> launches Denzo; get a blank line "prompt."</p>
<b>Step 5</b>	<p><b>ImageWindow:</b> Click on . See tiny circles.</p> <p><b>DenzoWindow:</b> <code>@site.dat</code> <code>@myexperiment.dat</code> <code>@auto.dat</code> begins autoindexing</p>
<b>Step 6</b> see	<p><b>ImageWindow:</b> Make sure green circles (corresponding to whole reflections) cover real reflections. Partial reflections are displayed as yellow circles. If circles don't cover real spots, sorry, gotta read more manual.</p> <p><b>DenzoWindow:</b> Examine the distortion table. Note the highest symmetry lattice with low distortion. Hopefully this corresponds to the lattice type you entered in the file <code>myexperiment.dat</code>. If not, exit both windows, edit <code>myexperiment.dat</code> appropriately, and repeat from _____ in the <i>DenzoWindow</i>.</p>
<b>Step 7</b>	<b>ImageWindow:</b> Hit  . Select a region of your pattern to examine. Hit the <b>Int. Box</b> button. Now you are going to make sure your integration box parameters are correctly set. This exploits the advantage of

: interactive analysis.

Examine the diffraction pattern to determine the resolution limits. The resolution at the position of the cursor is displayed in the green box (on the right side of the window). Let's say it's 2.2 Å.

**DenzoWindow:** RESOLUTION LIMITS 100 2.2  
FIT ALL  
GO GO GO GO GO

**ImageWindow:** Examine the diffraction pattern. Are all the real reflections covered by predicted reflections ('preds')? If not, you will have to increase the mosaicity. If there are too many preds, you will have to decrease the mosaicity. Let's say you started with a mosaicity of 0.4 degrees and you need to decrease it a bit. Here's how to change it:


**DenzoWindow:** MOSAICITY 0.3 GO

**ImageWindow:** Check again. Let's assume you've got it right now. Note it. If you have to go to unreasonable values of mosaicity (>> 1 degree), see later in this manual.

Look at the integration boxes in the zoom window. Zoom in enough to see the spots. Are the preds big enough to cover the spot? Are they too big? Are the reflections so close together that the boxes overlap adjacent spots (lots of red circles)? If you have overlaps that a small adjustment of the spot or box size is not going to fix, then you will have to adjust your data collection, perhaps by changing the oscillation range or the crystal-to-film distance.

Let's say that your reflections are smaller than the inner white jagged outline, which is the spot area, and that a slight adjustment in the box size will prevent most of your overlaps. In the `site.dat` file the spot size was defined as a 0.75 x 0.75 mm ellipse and the box was 3.5 mm on a side. Change that now:

**DenzoWindow:** SPOT ELLIPTICAL 0.5 0.5 0.0  
BOX 2.5 2.5 GO GO GO

**ImageWindow:** Hit . The display will show the profile fitting radius as a white circle and put preds only on those spots with  $I/\sigma$  greater than the weak level (default value  $5\sigma$ ). If you don't have enough spots for profile fitting (say > 10), you may want to increase the profile fitting radius. Let's say you need to increase it slightly:

**DenzoWindow:** PROFILE FITTING RADIUS 30 GO

Hopefully you get the idea of how to adjust your parameters.


## Step 8

see  
and

**DenzoWindow:** @refine.dat  
begins sequential refinement of frames specified in the sector n to m section of `myexperiment.dat`. Depending on how fast your computer is and how fast you can read, examine the screen output for the correct histogram shape (meaning you chose the mosaicity correctly, and for good  $\chi^2$  values and convergence). If your `site.dat` file was set up correctly, things should work.

see

**ImageWindow:** Watch refinement proceed. Little circles should line up on

		reflections. Cool! If not, gotta read more manual.
<b>Step 9</b>	<b>DenzoWindow</b>	After last frame, type <b>STOP</b> . This ends the program.
	<b>ImageWindow</b>	Hit  button to kill image. You're done! Result is a bunch of <code>xtal###.x</code> files.
<b>Step 10</b>		Scroll backward through the <b>DenzoWindow</b> to examine the histogram and the final $\chi^2$ values. If they are normal, you can proceed to . If not, see the discussion in the following chapter. If this scenario didn't work and you entered the correct values in the <code>myexperiment.dat</code> file, it is most likely due to an error in the <code>site.dat</code> file. Candidates include <b>X BEAM</b> and <b>Y BEAM</b> .

see

in

✦ Notice that the session was run interactively, not as a batch job. This is so that you can watch the progress of the refinement in the **ImageWindow**. Also, it is to encourage you to process your images as they are collected, so that you can spot problems before you have wasted lots of time. If the refinement is not working, you will be able to see very quickly that the predicted reflections (preds) do not line up with the real reflections and stop the job. If the data collection has problems, you can stop the experiment and fix your crystal, oscillation range, etc. If the computer is reasonably fast, the data are on a local disk, and the number of reflections is average, you can probably process 90 frames in less than half an hour, which is not too bad. If you still want a log file to examine later once you catch your breath, try typing something like:

```
> denzo | tee denzo.log
```

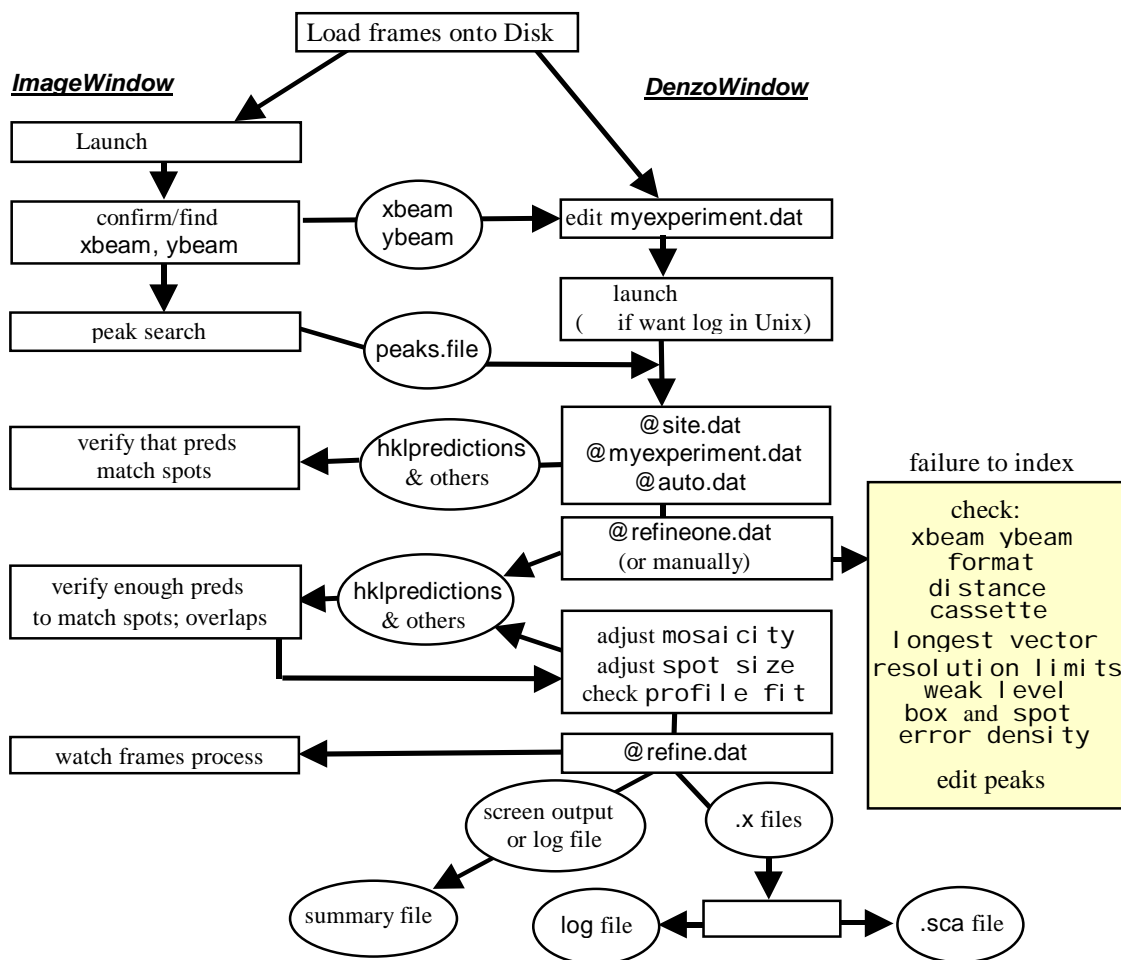
when starting .



# A Denzo Session

This chapter describes data processing in more detail than in “ for Dummies”.

## Denzo session outline



1. Display the image and start .
2. Autoindex to deduce the orientation of the crystal, and refine the parameters for the image you have displayed.
3. The entire series of frames is processed, either in batch or interactively.
4. The log file of the batch job is then examined to see the results of the processing, or the .x files are displayed over the diffraction image, or the data are processed in .

What follows is a description of each of these operations.

# Displaying the Image Using XdisplayF

In the process which leads to  $\langle I \rangle$ ,  $\langle I^2 \rangle$ ,  $F^2$  and  $\sigma$ , the final statistics are affected by many factors. The program outputs a substantial amount of information to help identify the potentially sub-optimal steps in the data collection and reduction. Interactions between factors like detector geometry, crystal symmetry, radiation damage, detector failures, non-ideal crystals, etc. are far too complex to be formalized in a method that could be implemented strictly in a computer algorithm.

`xdisp` allows the crystallographer to visualize the data in their original form as well as to observe the progress of data reduction. Displaying the raw data takes advantage of the human ability to grasp the significance of complex patterns that would otherwise be hard to analyze numerically.

Visualization of the data collection and reduction process allows efficient recognition of what is correct or wrong by taking advantage of the brain's ability to recognize complex visual patterns. Problems like detector vibration may produce very striking patterns on the diffraction image, yet the deleterious effect on the final  $R_{\text{merge}}$  statistics may not be so obvious.

The optimal experiment has to take into account the interaction between all of the following: the experimental point spread function, the unit cell length, the detector position, the angular range of data collection, the mosaicity, and the definition of the integration box. The ability to visualize the influence of different data acquisition parameters allows the experimenter to understand his particular experimental limitations and help to overcome them.

## Starting XdisplayF

The display program for `xdisp` is based on the X-window interface. `xdisp` runs concurrently with a `denzo` session either in a separate window or on a separate graphics terminal. Both `xdisp` and `denzo` store the entire image in memory, so the host computer must have approximately (depending on the exact size of the image) 64 to 512 MB of RAM in excess of that needed for the operating system. It also needs a substantial amount (e.g. 1.5 GB) of swap space. Note that if you are indexing 72 MB Raxis-IV images you will need 144+ MB of memory.

In order to run `xdisp`, open a window on a graphics terminal (**ImageWindow**) and launch the program. Launching the program involves the following:

- calling the program itself ('`xdisp`' is an alias)
- calling the format variable (the same as the modifier of the keyword **FORMAT** in `denzo`)
- telling the program the name of the image file to be displayed

Here's an example:

```
xdisp raxis xtal###.osc 1
```

In this example, the name of the frame to be displayed is `xtal001.osc` as number 1 represents the sector (frame) number. The old versions of the program used a little different convention:

```
xdisp raxis xtal001.osc [hklpredictions]
```

Communication between `xdisp` and `denzo` is achieved by using `hklpredictions` and `.test` files. This explains why the programs must be started in the same directory. The program version number is stored in these files, so you should delete them when switching to a new

version of the programs. `hklpredictions` is a name of the file, which creates. It contains the predicted reflections. The use of `hklpredictions` file is a default, therefore is put in parentheses [].

The old convention still works for some detectors. The change of convention has been forced by the variety of naming schemes used by detector vendors and developers.

The new convention has to be applied to display predictions from the output files. To display predictions from `xtal150.x` file one has to type:

```
xdisp raxis xtal###.osc 150 xtal###.x
```

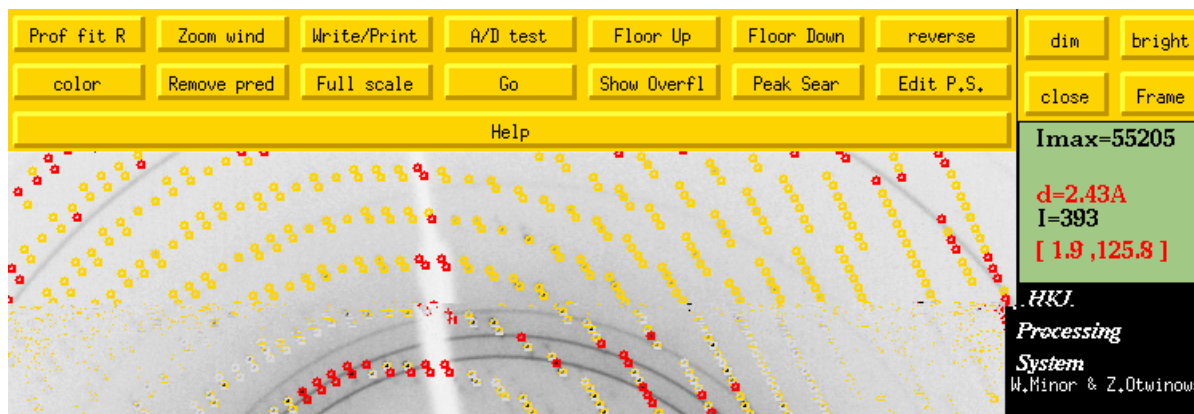
- ➡ `xtal150.x` represents the name of the file, which contains a list of reflections in output format. You need to click the **Update Pred** option in the *ImageWindow* for predictions to appear on the screen.
- ➡ You may get a list of all supported formats by typing “`xdisp`”. Formats are also listed in the Appendix: \_\_\_\_\_

## Using the XdisplayF

The frame should now be displayed on the screen.

For the most part, the features of the display program are self-explanatory. The display is changed by clicking the middle mouse button (or both buttons for a two-button mouse) once on any of the buttons above the image. It is possible to zoom in and out, to change the color and intensity of the image, and to determine the resolution and x-y coordinates (in pixels and mm) of any point on the image. To exit from the display program click once on **close**. What follows is a more detailed description of all the options.

## The Main Display Window



**Prof fit R**

Toggles (turns on/off) the profile fitting radius. You must have already indexed the image for this to be of any use, because the preds shown in the display are those which are stronger

than the **WEAK LEVEL**, and the **WEAK LEVEL** is determined by running . The profile fitting radius is changed by entering it in the *DenzoWindow*. See **PROFILE FITTING RADIUS** in the Denzo keyword documentation.

<b>color</b>	Toggles between black & white and color. Mono is often combined with reverse imaging to get black spots on a white background.
<b>Zoom wind</b>	Launches the zoom window. Any time you click the middle mouse button in either the regular or the zoom window, that point becomes the new center of the zoom display.
<b>Remove pred</b>	Toggles between updating and removing predictions. Once 'update predictions' is turned on, new predictions from each cycle will be updated automatically in the display.
<b>Write/Print</b>	This tells the program to write out a file which contains the current image in the .xwd format. You may use a local utility to convert this file into a postscript file.
<b>A/D test</b>	Toggles this function on and off. To use, click on <b>A/D test</b> , then use the right mouse button to bracket a region of the diffraction pattern. A black and red window will appear with a histogram of pixel intensities. This plot (but not the window) can be expanded or contracted by clicking on the arrow buttons. This function is useful for diagnosing hardware problems. Detectors with problems will have dips in what would otherwise be a smoothly varying curve. To exit from the A/D test, click on the <b>A/D test</b> again.
<b>Full scale</b>	Changes the right side scale bar so you can get an idea of where your image is on the full dynamic range of the detector.
<b>Floor Up</b>	Changes the lower level for the display range. Basically causes the image pixels with intensities below the "floor" to be displayed at the lowest color/darkness level, so in effect it narrows the displayed dynamic range, but increases the contrast. The left-middle-right button produces a small-medium-large change.
<b>Go</b>	Used in conjunction with the <b>Frame</b> . Starts a "movie" by sequentially incrementing or decrementing the last digit of the file name being displayed. If your file is called xtal001.osc, and <b>frame up</b> is selected, hitting <b>Go</b> will tell the display to sequentially display xtal002.osc, then xtal003.osc, etc. until you stop it or run out of frames. This is a neat way of screening through your frames and spotting any that have obvious problems. Of course, it works better on a fast computer. You can do the same thing manually with <b>Frame</b> (below).
<b>Floor Down</b>	Is the opposite of <b>Floor Up</b> . Increases the dynamic range being displayed by lowering the floor, and again, the change depends on which mouse button is pressed.
<b>Show Overfl</b>	Colors saturated image pixels red.
<b>reverse</b>	Inverts the dark/light display. Changes white to black, and black to white, and all others in between. Especially useful for black and white displays, so that the spots appear black and the background appears white.
<b>Peak Sear</b>	Starts the peak search routine and results in the creation of a file called peaks.file The number of spots <b>Peak Sear</b> picks can be controlled with menu options, which will appear on the right-hand side after the initial peaks are found. The picked peaks are displayed as small black circles over the image. <b>This set of peaks is the starting point for the autoindexing routine in Denzo.</b>

<b>dim</b>	Increases the display dynamic range, which has the effect of emphasizing the stronger reflections with a concomitant loss of the weaker ones. It could also be called <b>Top up</b> to make it analogous with the <b>Floor Down</b> button.
<b>bright</b>	Opposite of <b>dim</b> . Decreases the display dynamic range, which has the effect of emphasizing the weaker pixels. The dark pixels will just stay dark. This is analogous to the <b>Floor Up</b> , but at the other end of the spectrum.
<b>close</b>	Quit and exit the display program. This will not interfere with any jobs running in a separate window as long as you do not log off the computer.
<b>Frame</b>	Tells the display which direction to increment the last digit of the image file when displaying sequential images. See the description above for <b>Go</b> . The right mouse button causes the file number to increment, and the left mouse button cause the file number to decrement. The middle mouse button is analogous to <b>Go</b> except that one click results in only a one frame advance (or decrease).
<b>Help</b>	
<b>Help</b>	Has some help menus. Not very helpful yet. To close click <b>Help</b> again.

## Buttons in the Zoom Window

**Right mouse button.** This displays a set of cross sections through the zoom window so that you can see the 2-D profiles of the pixels.

**Mid mouse button.** Re-centers the zoom display on the selected point. Works in either the zoom window or the full image window.

**Left mouse button.** Displays the information about a particular point in the zoom window in the green box in the lower and upper left corners of the zoom window. If the frame has been indexed already by then when you click on a spot the h, k, l indexes of that spot will be displayed. Otherwise, the display only shows the position of the spot and its intensity.

<b>Zoom in</b>	Increases the magnification of the zoom window. When you get to really high magnification, the intensity of each pixel is written in that pixel. However, this is not always accurate, and will be discussed under the Green Box window section below.
<b>Zoom out</b>	Decreases the magnification of the zoom window.
<b>Int. box</b>	This is a toggle through three different displays. The first (default) display shows a circle centered on the predicted reflection. This circle is always 6 pixels in diameter and has nothing to do with the size of the spot. This is the display you see if you don't hit the <b>Int. box</b> . The first time you hit the <b>Int. box</b> , you will see the integration box (the <b>BOX</b> of ) plus the "circles" which define the spot and the guard regions, both in white. Note that you have to have indexed the frame in for this display to be useful, since the integration box and the predicted reflections are determined in , not

the spot, guard, and box sizes in mm in , their actual sizes must be an integral number of pixels, so you are not going to see perfect circles. Also, the spot must have numerical parity with the box, i.e. the background will always be symmetrical around the spot, so if the box has an odd number of pixels in a direction, then the spot will as well, even if it is supposed to be circular. The next time you hit the **Int. box**, the box is removed and the spot outline is colored red. This makes it easy to see things when spots are close together. When you hit the **Int. box** once more, the original display is restored.

**Diff Vec**

When this button is clicked, a second, black circle is displayed which is centered on the actual observed reflection. The green or yellow circles (representing whole or partial reflections) are centered on the predicted positions of the reflections. If the indexing and refinement have gone well, the two circles should be very close to concentric. Notice that most of the time when they are not, they differ by only a little bit. When they differ by a lot, you will see a little line drawn which connects the centers of the observed and predicted reflection positions. These lines are easy to spot in a low zoom window, so you can easily examine any problem reflections.

**Zoom close**

Closes the *ZoomWindow*.

**Imax=55205**

**d=2.43Å**  
**I=393**  
**[ 1.9 ,125.8 ]**

**Green Box on the right side of the window**

**Imax=55205**

The maximum pixel intensity in an image.

**d=2.43Å**

Resolution of the cursor position, in Å. Note that this is only accurate once has been launched and has read in the distance and wavelength and communicated these to the display program. This means that at least one

## Visualisation of the diffraction space

A diffraction data set forms an **image of three-dimensional reciprocal space**. 3-d image consists of a series of two-dimensional diffraction images, each of them representing a different, curved slice of reciprocal space. In order to integrate the diffraction maxima accurately, they must appear as separated (non-overlapping) spots in the individual 2-d images. Unless the data are collected by the precession method, the diffracted image contains a **distorted view** of reciprocal space. This distortion of the image is a function of the data-collection method, the diffraction geometry, and the characteristics of the detector. For the data reduction to be successful, the distortion of reciprocal space as viewed by the detector has to be accounted for correctly by the program. The distortion of the image of reciprocal space can vary even between images collected on the same detector. This is because the position of the detector, the X-ray wavelength, the oscillation range, pixel size, scanner gain, and the exposure level all, affect the detector representation of diffraction space.

One should start data collection and reduction with a careful inspection of the data in their raw (original) form. The zoom option of the program allows one to examine reflections in pixel-by-pixel detail to check that the diffraction maxima are resolved. Because the program displays the resolution (in Ångströms) corresponding to the position of the mouse-driven cursor, the diffraction limit of the crystals can be estimated even without data reduction. The display in high zoom mode provides digital pixel values, so one can check, among other things, that the exposure level is appropriate.

If problems exist with the detector or other components of the data-collection system, the display option helps to discover these before all the data are recorded. The examination of the image may reveal if there are extraneous sources of X-ray background. There are other statistics that can be provided instantly by which may indicate for example A/D converter malfunction. If there are many diffraction maxima in the image that form a characteristic pattern of diffraction from a single crystal, then the next step is deducing a crystal lattice that accounts for such a pattern. This step is called **indexing**.

## Starting Denzo

is run concurrently with the display program from another window (let's call it *DenzoWindow*) on a multi-window workstation. To run , type something like this:

```
>/usr/users/students/HKL/denzo
```

Make sure you are in the same directory from which you launched , because the two programs share information through the `hklpredictions` and `peaks.file` files (and other X-window data structures called 'atoms'). The program has no prompt, so when you get a blank line, the program is ready to accept your input, be it commands or a macro file. Since most interactive sessions start out with trying to index a particular diffraction pattern, that's what I'll discuss next.

# Getting that First Orientation: Indexing

In order to assign an index properly to each spot on a diffraction image, the orientation of the crystal in space must be known. Historically this chore was performed by adjusting the arcs of the goniometer on which the crystal was mounted until test, or 'setting' photographs showed that the crystal was aligned. Usually this meant that one of the unit cell axes was perfectly aligned with the spindle axis. It was then a relatively straightforward matter to index the spots on an oscillation photograph, knowing the unit cell parameters, the oscillation range, and crystal alignment. Modern data collection strategies, however, typically involve the collection of oscillation photographs from crystals where the orientation of the reciprocal lattice relative to the X-ray beam and the film is not known ahead of time. The chore of indexing has thus become one of deducing the orientation from the collected oscillation photographs, rather than adjusting the position of the crystal before data collection begins.

There are **five parameters** that specify the orientation of the crystal relative to the X-ray beam:

- **VERTICAL AXIS**
- **SPINDLE AXIS**, and
- **CRYSTAL** `rotz`, `rotx`, and `roty` values.

These are what we wish to deduce in our `session`. Note that our primary goal is to deduce the orientation of the crystal relative to the X-ray beam, even though the orientation of the detector relative to the beam is of equal importance. The reason we focus our attention on the crystal orientation is simply because the detector, or "site" parameters, like **DI STANCE**, **2 $\theta$** , and the position of the direct beam at the detector are (and should be!) reasonably well known before processing begins. Thus they can be refined automatically by the program.

## Indexing

The `package` offers two indexing methods: automatic and interactive. The **automatic method**, applicable in most cases, is fast and simple. The first step in the automatic method is the peak search, which chooses the spots to be used by the autoindexing subroutine. Ideally, the peaks should come from a diffraction by a single crystal. The `program` accepts peaks for autoindexing only from a single oscillation image. It is important that the oscillation range be small enough (it even can be zero, i.e. a still) so that the **lunes** (rings of spots from one reciprocal plane) are resolved.

### Lune separation ►

The requirement of lune separation is distinct from the requirement of spot separation. If the lunes overlap, spots may have more than one index consistent with a particular position on the detector. On the other hand, the oscillation range should be large enough to have a sufficient number of spots for the program to be able to establish the periodicity of the diffraction pattern. This may require at least 0.5 degree oscillation for a small unit cell protein crystal and 2-3 degree oscillation in the case of organic small molecule crystals.

The second step in the autoindexing is the **mapping of the diffraction maxima** identified by the peak search onto reciprocal space. Because the precise angles at which reflections diffract are `unknown` for oscillation data, the center of the oscillation range is used as the best estimate of the angle at which the diffraction occurs.



The **autoindexing** in `hkl2map` is based on a novel algorithm: a complete search of all possible indices of all reflections, that are found by a peak search or manual selection. When the program finds values (integer numbers) of one index (for example, **h**) for all reflections, this is equivalent to having found one real-space direction of the crystal axis (in this case, **a**). For this reason such indexing is called **real-space indexing**. Finding one real-space vector is logically equivalent to finding the periodicity of the reciprocal lattice in the direction of this vector. The search for real-space vectors is performed by a Fast Fourier Transform (FFT) and takes advantage of the fact that finding all values of one index (e.g. **h**) for all reflections is independent of finding all values of another index (e.g. **k**). The `hkl2map` implementation of this method is not dependent on prior knowledge of the crystal unit cell, however for efficiency reasons, the search is restricted to a reasonable range of unit cell lengths, obtained by default from the requirement of spot separation.

After the search for real-space vectors is completed, the program finds the three linearly independent vectors, with minimal determinant (unit cell volume), that would index all (or, more precisely, almost all) of the observed peaks. These three vectors form a basis, but are unlikely to form a standard basis for a description of the unit cell. The process of converting a basis into a standard basis is called **cell reduction**. The program follows the definitions in the International Tables for Crystallography, and finds the best cells for all of the fourteen Bravais lattices.


The transformation of the primitive cell to a higher symmetry cell may require some **distortion** of the best triclinic lattice that fits the peak-search list. Due to experimental errors the fit is never perfect for the correct crystal lattice. Sometimes the observed reflections can fit a higher symmetry lattice than one defined by space-group symmetry. Such a condition is called lattice (or metric tensor) **pseudo symmetry**. If this happens, the lattice determination and assignment of lattice symmetry may get complicated. The procedure in such case is to index the data in lowest symmetry lattice that does not introduce wrong lattice symmetry (triclinic lattice is always a safe choice), and look for the symmetry of the intensity pattern during the scaling of symmetry-related reflections. `hkl2map` calculates the distortion index for all fourteen of the Bravais lattices. It is up to the user to define the lattice and space-group symmetry, since the program at this stage of the calculation cannot distinguish lattice symmetry from pseudo symmetry.

## Autoindexing

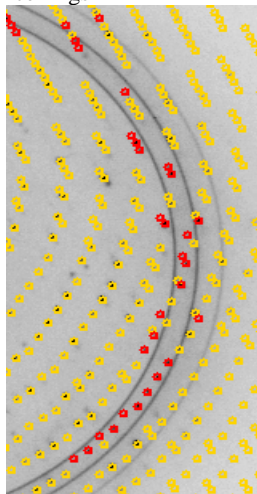
The universal autoindexing procedure is now part of `hkl2map`. This procedure is able to deduce the lattice type, crystal unit cell parameters, and crystal orientation parameters from a single oscillation image. It is also able to index more complicated lattices routinely, for example body-centered or rhombohedral lattices. It really makes life a lot simpler!

Autoindexing in `hkl2map` always finds a **standard lattice**, however the crystallographer may prefer a non-standard choice, for example similar to one in a different space group. Reindexing in `hkl2map` or manual indexing in `hkl2map` accommodates such needs.

## Peak picking

To start the autoindexing, the first step is to select a set of spots from the displayed oscillation (or still) image. A subset of these spots will be used by the autoindexing routine. To select the spots, go to the **ImageWindow** which contains the displayed image and click  (**Peak Search**). The number of spots the peak-search routine picks can be controlled with menu options, which will appear after peak search finishes. If the number of spots is too few or too many, click the **More Peaks** or **Fewer Peaks** respectively, as many times as necessary, until you have about the right number of spots. If your spots are unusually large or small for your detector, choosing **Size Peak Up** or **Size Peak Down**, will change peak search criteria (peak size and overlap) towards larger or smaller spots. You can also edit results of the peak search procedure by using **Pick (Add)** and **Pick (Remove)** options. For example, you can

Ice rings



remove a series of peaks by clicking first **Pick (Remove)**, then clicking in either zoom or main window on peaks you want to remove.

There is no “perfect” number of spots to pick for autoindexing, because it is the **quality** of the picked peaks, not the quantity, which is critical. The peaks used in autoindexing should come mostly (say > 90%) from a single lattice and should line up on real reflections in the image. If most do not, then there is something wrong and you may have to change some parameters as discussed below under

The two most common problems which interfere with the peak search are **satellite crystals** and **ice rings**. With the noisy image, the peak search program may pick spots which are clearly not part of the lattice. By clicking on **Fewer Peaks** the number of peaks selected can be reduced. If your diffraction pattern has some strong peaks, these will usually get picked before the noise.

**Ice rings** pose a different problem because the peak search program will pick tons of “spots” within the very intense ice ring. However, it is easy to overcome this by simply specifying resolution limits (say 20 to 4.1 Å) in the autoindexing macro file (`auto.dat`) which exclude the ice ring (more on this later in the section).

The picked peaks are written to a file called `peaks.file`, in whatever directory you happened to launch the display program. This file is now read by to autoindex the image.

### Running autoindexing

Back to the *DenzoWindow* now, you are ready to input the initial parameters and start the autoindexing. For this purpose you will need the following files: the oscillation image file, the `peaks.file` file, and three macro files called `site.dat`, `myexperiment.dat`, and `auto.dat`. These three macros list the detector and crystal parameters and tell the program to begin the autoindexing. Then, in your *DenzoWindow*, you would simply type `@site.dat`, `@myexperiment.dat`, and `@auto.dat`, and the program begins. Here's an example of the macro files:

#### `site.dat`

---

```

FORMAT dip2000
WAVELENGTH 1.5418
FILM ROTATION 270.00
CASSETTE roty 0.465 rotx -0.102
[ERROR density 0.60]

BOX 2.4 2.4
SPOT ELLIPTICAL 0.6 0.7
BACKGROUND elliptical 0.7 0.8
OVERLAP spot
X BEAM 99.4
Y BEAM 99.6

```

---

In this example and throughout this manual, each major keyword is in **BOLD CAPITALS**, modifiers and values are printed in lower case letters. Comments are and do not actually appear in the command file. For example, the modifiers `roty`, `rotx` modify the keyword **CRYSTAL** as well as **CASSETTE**. A precise description of each of these keywords and parameters can be found later in this manual.

myexperiment.dat

---

```
OSCILLATION range 1.0
OSCILLATION start 0.0
DISTANCE 150.0
SPACE GROUP P212121
MOSAICITY 1.0
TITLE 'my crystal'
SECTOR 1 to 90
RAW DATA FILE 'data:[mydir]xtal0##.osc'
FILM OUTPUT FILE 'data:[mydir]xtal0##.x'
X BEAM 99.4
Y BEAM 99.6
BOX 2.4 2.4
SPOT ELLIPTICAL 0.6 0.7
BACKGROUND elliptical 0.7 0.8
OVERLAP spot
ERROR positional 0.050
PROFILE FITTING RADIUS 30
```

---

auto.dat

---

```
PEAK SEARCH FILE peaks.file
FIT x beam y beam cell crystal rotx roty rotz
[RESOLUTION LIMITS 100 4.1]

WRITE PREDICTIONS
GO
```

---

When

---

Vector lengths in autoindexing from 5.7 to 152 Angstroms								
<b>primitive cubic</b>	60.08%	92.79	92.79	92.79	90.00	90.00	90.00	
		77.48	124.78	65.26	89.82	32.01	89.54	
<b>I centered cubic</b>	31.25%	130.81	130.81	130.81	90.00	90.00	90.00	
		130.97	136.43	124.78	90.00	161.72	84.76	
<b>F centered cubic</b>	37.67%	146.44	146.44	146.44	90.00	90.00	90.00	
		146.35	146.67	146.31	63.19	127.02	147.44	
<b>primitive rhombohedral</b>	9.82%	132.55	132.55	132.55	27.55	27.55	27.55	
		131.62	140.63	124.78	27.65	18.16	32.85	
		58.37	58.37	360.56	90.00	90.00	120.00	
<b>primitive hexagonal</b>	9.14%	71.24	71.24	124.78	90.00	90.00	120.00	
		76.75	65.26	124.78	90.18	90.15	147.65	
<b>primitive tetragonal</b>	11.96%	54.53	54.53	124.78	90.00	90.00	90.00	
		41.07	65.26	124.78	90.18	90.58	89.40	
<b>I centered tetragonal</b>	6.51%	71.24	71.24	252.50	90.00	90.00	90.00	
		76.75	65.26	252.50	90.08	85.24	147.65	
<b>primitive orthorhombic</b>	0.30%	41.07	65.26	124.78	90.00	90.00	90.00	
		41.07	65.26	124.78	90.18	90.58	89.40	
<b>C centered orthorhombic</b>	2.27%	41.07	252.50	65.26	90.00	90.00	90.00	
		41.07	252.50	65.26	89.92	90.60	81.22	
<b>I centered orthorhombic</b>	4.72%	41.07	77.48	269.93	90.00	90.00	90.00	
		41.07	77.48	269.93	92.77	72.99	57.38	
<b>F centered orthorhombic</b>	3.33%	41.07	136.43	252.50	90.00	90.00	90.00	
		41.07	136.43	252.50	92.71	98.78	73.08	
<b>primitive monoclinic</b>	0.18%	41.07	65.26	124.78	90.00	90.58	90.00	
		41.07	65.26	124.78	90.18	90.58	89.40	
<b>C centered monoclinic</b>	2.27%	41.07	252.50	65.26	90.00	90.60	90.00	
		41.07	252.50	65.26	89.92	90.60	81.22	
<b>primitive triclinic</b>	0.00%	41.07	65.26	124.78	89.82	89.42	89.40	
autoindex unit cell		41.07	65.26	124.77	90.00	90.00	90.00	
► crystal rotx, roty, rotz -131.740 -39.042 -4.743								
Volume of the primitive cell 334451.								
Autoindex Xbeam, Ybeam 96.42 95.58								
crystal rotx	48.260	roty	39.042	rotz	-175.257			
crystal rotx	48.260	roty	39.042	rotz	4.743			
crystal rotx	-131.740	roty	-39.042	rotz	-4.743			
crystal rotx	-131.740	roty	-39.042	rotz	-4.743			

---

Blue lines represent the 14 possible Bravais lattices in decreasing order of symmetry, with **primitive cubic** being the highest symmetry and **primitive triclinic** the lowest. After the lattice name is a percent value which represents the amount of distortion the unit cell parameters would suffer in order to fit the lattice. Next to this percentage are the 'distorted-to-fit' unit cell parameters, and below are the undistorted unit cell parameters, for comparison (in case of a **primitive rhombohedral** lattice there are actually two lines - the second with parameters in hexagonal indexing).

As you can see, for some lattices (e.g. primitive tetragonal, all the cubics, etc.) a substantial amount of distortion is required. Similarly, in all cases there will be zero distortion required to fit the data to a **primitive triclinic** cell, since all three unit-cell lengths and angles are unrestricted. **The goal here is to find the highest symmetry lattice which fits the data with minimal distortion.** In the example above, there are two candidates, aside from the trivial case of P1. These are **primitive monoclinic**, with a 0.18% distortion index, and

`primitive orthorhombic`, with a 0.30% distortion index. Since 0.30% is, in this case, an acceptable amount of distortion, especially given that the unit cell parameters were refined from a single frame, a reasonable choice for the lattice would be `primitive orthorhombic`.

Note that this choice of lattice must be confirmed, since it may well be a monoclinic lattice that has a beta angle which is very close to 90 degrees. To do this, process the data with `postrefine`, and then postrefine the unit cell parameters and scale the intensities in `scale`. There will be more discussion of space group determination in the `manual`.

The first `crystal` line (pointed to by red ►) is very important, as it gives the crystal orientation parameters.

The last group of `crystal` lines give alternative choices for these values, based on crystal symmetry. Thus, for example, if you preferred to work with small numbers, you would choose the values from the second of these lines for your crystal orientation parameters. It doesn't really matter, it's what you prefer.

Autoindex Line gives the **X BEAM** and **Y BEAM** positions. However, since the success of autoindexing is critically dependent on knowing this accurately in advance, these values should be close to the initial values in the file `site.dat`.

## When Autoindexing Fails

Occasionally the autoindexing routine does not give clear-cut results.

There are four classes of data-reduction problems:

1. Utter failure
2. Misprediction
3. Kind-of-failed
4. The predicted positions agree perfectly with the peaks, but data scale poorly

**Utter failure** means that there is no reasonable predicted lattice, the crystal unit cell parameters are unreasonable, and the predicted diffraction pattern does not correspond to the observed image. In most cases failure can be ascribed to problems in three areas:

- experimental conditions,
- the parameters of autoindexing,
- the peak search (rarely).

**experimental conditions** Problems with defining experimental conditions account for the vast majority of the problems with autoindexing single crystals. If any of the following parameters is widely off, your autoindexing is screwed up (listed in order of importance):

- the **X BEAM** and **Y BEAM** positions. They can be measured from the diffraction image using the cursor and reading off the position from the little green box in `image`. A direct beam shot or a good educated guess (say based on the intersection of Laue streaks, the symmetry of the diffraction pattern, etc.) is often necessary. At synchrotrons, when the data is collected on IPs in removable cassettes and the beam wanders all over the place on the image, it is **very** helpful to have a **beam stop** which has been **thinned** sufficiently at its center so that a tiny amount of the direct beam passes through and makes a spot on each frame whose position can be measured.
- specification of the **detector format** (believe it, it's happened),
- the crystal-to-film **DI STANCE**. The distance should be known accurately (within about 1%) and can be measured with a tape measure as a last resort
- the **CASSETTE** `rotx` and `roty` values. Misset **CASSETTE** `rotx` and `roty` values may be due to 2θ angles other than zero.

Really, knowing the detector parameters only makes sense. **If you want to deduce some unknowns** (crystal lattice and cell, and orientation), **you need some knowns too**.

- oscillation range** Very rarely autoindexing will fail because too large or too small an oscillation range was collected. In the case of **large** oscillations, the lunes may have severe overlap and the autoindexing program will have a problem with the non-uniqueness of the mapping from detector to reciprocal space. In the case of **small** oscillations or stills collected on non-mosaic crystals with a very parallel beam, too few reflections may be present to describe a lune. As a rule of thumb the following types of oscillation ranges should be used:
- **viruses:** 0.25° oscillations,
  - **proteins:** 1.0° oscillations,
  - **small molecules:** 2.5° oscillations,
- assuming well diffracting samples. There is a fairly wide margin on either side, at least a factor of 2.
- autoindexing parameters** Before autoindexing, `edit` edits the peak list produced by the peak search algorithm of the XdisplayF program. Problems with the `edit` autoindexing parameters account for a few of the remaining problems with autoindexing. These may have to be adjusted in the case of non-single crystals, very large or small unit cells, lots of noise peaks, or ice rings. These parameters are controlled by the keywords: **LONGEST VECTOR**, **RESOLUTION LIMITS**, **WEAK LEVEL**, **BOX**, and **SPOT**.
- THE LONGEST VECTOR** may need to be adjusted if the peaks are sparse due to a small unit cell or low mosaicity. In this case, it should be set to equal the longest vector of the reduced primitive real space unit cell. If you did not understand that last sentence, simply use the longest unit cell axis (which is the same in primitive space groups, anyway). Better to err on the side of larger rather than smaller values.
- RESOLUTION LIMITS** can be adjusted, if you have:
- an **ice ring** which interferes with the indexing, you can set the resolution limits to, say, 100 - 4.1 Å to eliminate these reflections.
  - a **small satellite crystal** which contributes some reflections at low resolution but not at high resolution, you can set the resolution limits to, say, 8 - 2 Å, thereby excluding the strongest contributors from the contaminating lattice.
  - a **cracked crystal** where the two lattices are only 2-3 degrees apart, you can often adjust the resolution limits as in the case of an ice ring so that you index at low resolution where the two lattices are approximately identical, and then use manual refinement to zero in on both lattices. The idea here is to index each lattice separately and to combine the data later in `merge`.
  - a **twinned crystal** where the two lattices have a non-random orientation relative to one another, the program may fail spectacularly by finding a **superlattice** which indexes both lattices simultaneously but cannot be used to integrate the data because the superlattice does not usually fit both lattices exactly. Then it's time to call an expert or grow new crystals.
- WEAK LEVEL** If you have a **number of spurious peaks** in the image distributed randomly over the image, you can adjust the value of the **WEAK LEVEL** which appears before the first **GO** statement in the `auto.dat` macro. This will affect the number of peaks accepted by the autoindexing routine. By increasing the **WEAK LEVEL**, you can exclude weak, spurious peaks, then run autoindexing, and examine the superimposition of the autoindexing preds (green circles) on the image. If most of them cover real reflections and not noise, then you are succeeding.
- BOX and SPOT** sizes may have to be adjusted if too many spots are being removed by the overlap criteria. **Autoindexing does not use** spots which `edit` determines are **overlaps**.
- peak search** Finally, you can try changing the number of peaks you select in the peak picking routine. However, since `edit` edits the list of peaks anyway, this is unlikely to yield much improvement if the suggestions above do not work. In that case, **do you have a better crystal?**

**Misprediction** results in location of reflection masks not corresponding to the positions of the Bragg peaks. It is visibly obvious and is disastrous. It may be due to:

- forcing wrong space group symmetry,
- misindexing or
- serious detector malfunction.

**misindexing** Autoindexing is based on the assumption that the diffraction spots are correctly mapped from detector coordinates to diffraction (reciprocal) space. The origin of the diffraction space is defined by the position of the direct beam on the detector. A substantial error in the beam position can shift the indexing of the diffraction pattern by an integer vector. Such misindexing can be totally self-consistent until the stage when symmetry related reflections are compared. For any assumed (starting) value of the beam position, the origin of the diffraction space during indexing will be shifted to the nearest grid point of the best primitive lattice. An initial error in the direct beam position by 0.48 times the distance between reflections will lead to correct indexing, while an error of 0.52 times the same distance will cause a misindexing of the diffraction pattern by one index. **Misindexing by one is never corrected by subsequent refinement** of the crystal and detector parameters. Misindexing often produces poor agreement between the predicted and the observed positions of the reflections, but for some crystal orientations, the agreement between the predicted and the observed positions can be equally good for both correctly indexed and misindexed cases. This property of the diffraction geometry creates a potential trap for the unwary crystallographer.

Sometimes the **crystal asymmetric unit** may have molecules related by an approximate translation by a fraction (typically **one half**) of a unit cell edge or diagonal. The resulting diffraction pattern will have odd index reflections much weaker than even index reflections. Autoindexing may find one of the two possible solutions, the choice depending on whether odd reflections are weak enough to be assumed (within experimental error) systematically absent or not. This depends on what is the fraction of odd reflections in the peak search. If there are only a few odd reflections, then most of the peak search result may be explained with a **smaller real-space unit cell**. To prevent autoindexing from finding such a smaller cell, one should enhance the fraction of odd reflections used in autoindexing by changing the peak-search criteria or by using only high-resolution reflections in autoindexing. If one still cannot index odd reflections, then one should consider ignoring them all together. In such a case, one can solve the structure in a smaller unit cell and the resulting structural error will not be very significant if odd reflections are much weaker than even ones.

**Kind-of-failed** means that a lattice other than P1 was predicted, but it had a high distortion index, say 1-1.5%. This may not be a "failure" but simply an indication that further refinement of parameters is needed because the initial detector parameters were very far off. Unfortunately, misindexing may give similar results. A quick check of the **superposition** of the preds on the reflections should tell you if the refinement is likely to succeed. Note that refinement may converge to low  $\chi^2$  values even with misindexed crystals but eventually the problem will surface and you will have to index correctly.

**Predicted positions** **agree perfectly with the peaks** and no detector or diffraction artifacts are visible, but **data scale poorly** and produce many outliers. This problem may be:

- a simple mistake in data processing, like using the **wrong file format**, or
- a **non-uniform exposure** during crystal oscillation.

The **non-uniform exposure** may be caused by spindle motor backlash, shutter malfunction (opening too early or too late), ionization chamber electronics failure (if used), decay or variation of the X-ray beam intensity (if ionization chamber is not used), variable speed of the spindle motor etc. Non-uniform exposure is best diagnosed by exclusion of other problems that may affect data quality. Graphical feedback provides confidence that the problem cannot be at the indexing/integration stage.

**Absorption** Large variations in **absorption** of X-rays by the crystal will make data scale poorly and will produce visible variation of the background, however it will not affect positional agreement. The variation in the absorption can be avoided easily by a proper mounting of the crystal. The correction for absorption is a whole field in itself.

**Errors in detector orientation** will produce distorted mapping from detector to reciprocal space. A **wrong crystal-to-detector distance** will enlarge (or decrease) the apparent reciprocal space separation between Bragg peaks. This error will not be uniform in three directions; in the first approximation, along the beam direction, the error will be proportional to the square root of the error in the distance; in the other two directions the error will have linear dependence. In most cases a significantly wrong crystal-to-detector distance (say an error of 10%) will not make the autoindexing step fail immediately; however, **the calculated unit cell will be quite wrong**. The length of unit cell along the beam direction will be 5% shorter, in the perpendicular directions 10% shorter. If the crystal is diagonally oriented (no principal axis along the beam direction) than the apparent angles between axes will violate the lattice symmetry. Incorrectly defined **angles between detector and the beam** will result in incorrect angles between crystal axes. Therefore, the interpretation of the lattice symmetry is dependent, to some extent on how precisely the detector parameters are known.

### Manual indexing

See also  
appendix.

When autoindexing fails, you can try to index oscillation images manually, but if you tried the suggestions made above, you are unlikely to do any better. Manual indexing used to be the only way to run , but autoindexing has superseded this difficult task.

### Autoindexing succeeded!

If the autoindexing looks good, then it's often nice to finish up the frame you have been working on to make sure you have the correct values for **MOSAICITY** and **SPOT** shape. This can be done interactively in the *DenzoWindow*, by calling a refinement macro, examining the superposition of the preds on the diffraction image and then testing the **MOSAICITY** and **SPOT** parameters to make sure all the reflections are indexed.

### Multiple crystals, simulation and interactive indexing

Autoindexing can fail when more than one crystal contributes to the diffraction image. Sometimes, editing of weaker reflections and resolution cuts can make one crystal dominate the peak search list enough for the autoindexing method to succeed. If crystals have a similar orientation, sometimes using only very low resolution data can be the right method. In case of twinned crystals, autoindexing sometimes finds a **superlattice** that finds integer indexes for both crystals. In such cases solves the problem of finding the best 3-d lattice that goes through all of the observed peaks. Unfortunately, for a **twinned crystal** this is a mathematically correct solution to the wrong problem. Since there is no general algorithm to index a diffraction image from **multiple crystals**, one has to rely on the human ability to sort-out complex patterns in such cases. The approximate orientation can be determined by an iterative trial-and-error process where the predicted pattern is being adjusted, keeping the diffraction image constant.

A crystal orientation can be defined relative to any principal or higher order zone perpendicular to the X-ray beam. This flexibility helps the interactive indexing when only a higher order zone is visible in the diffraction pattern. This is particularly useful in centered space groups where it is easier to orient a diagonal zone, rather than a major one.

Manipulation of the predicted diffraction patterns also can be used to simulate diffraction experiments. The simulation can help set a proper data-collection strategy in order to avoid later problems in data reduction. Using the program for **simulation of diffraction patterns also can be a tool for teaching crystallography**.



## Refinement: Using the *Fit* Command

If autoindexing has given you a pretty good match between the observed diffraction pattern and your preds, you can use `FIT` to refine the parameters to several significant figures.

The crystal and detector orientation parameters require refinement for each processed image. The refinement can be simple, for a series of images collected with an on-line detector, or more complex, if the detector orientation is only crudely known and varies from image to image, as in the case of off-line scanners. The refinement is controlled by the user and can consist of several steps. In each step the user defines the resolution limits and the order and number of parameters to be fitted. Both detector and crystal parameters can be fitted simultaneously by the fast-converging least squares method. The refinement is done separately for each image to allow for the processing of data even when the crystal (or the detector) slips considerably during data collection.

**What to fit first** Although `FIT ALL` can successfully refine all the parameters - crystal rotations, film rotations, camera distance, unit cell dimensions, etc. - simultaneously if their initial values are close enough, in practice it is a good idea to start by first refining those values you are **least sure** of, and which are **most critical**. These are usually the crystal rotations, unit cell parameters, and the beam positions. In addition, the resolution limits must be set, and generally you start refining at a medium resolution and progress to a higher resolution in one or two steps. So, the first round of refinement would start with two commands:

```
RESOLUTION LIMITS 100 2.8
FIT crystal rotx roty rotz cell x beam y beam GO GO GO.
```

This tells the program to refine the crystal rotation angles, the unit cell lengths and angles, and the direct beam position for three cycles (= 3 `GO` statements), using reflections between 100 and 2.8 Å, and applying the results of the first cycle to the second, and so on. In the *ImageWindow*, you will see the preds move after each cycle to their new positions. If things look like they are getting closer, you can do more cycles simply by typing `GO` more times.

**FIT remains on! ►** Once you specify the `FIT` operation, as in `FIT crystal rotx roty rotz`, these parameters will be refined every time you type `GO`, no matter what you have specified on the same line as `GO`. For example, let's say that after the 3 cycles of `CRYSTAL rotx roty rotz` refinement, you decide to refine the `X BEAM` and `Y BEAM` positions using the command `FIT x beam y beam GO GO GO`. Now *not only* will the beam position be refined, but simultaneously so will the `CRYSTAL rotx roty rotz` rotations. Once `FIT` is on for a parameter, it is always on unless you turn it off. Those of you with sharp eyes will have noticed that the `auto.dat` macro already included the `FIT crystal rotx roty rotz cell x beam y beam` statement, but it doesn't hurt to say them again.

**FIT & autoindexing ►** The program uses one of two target functions specified by the keyword `FIT`. In the first cycle following autoindexing, the target function uses the reflections found by the peak search. In all subsequent cycles, the target function includes all reflections which overlap predicted reflections. The first target function which follows autoindexing has a much larger radius of convergence, at a tradeoff of less precision. On all subsequent `FIT` cycles, a more

precise fit algorithm is used which completely ignores the list of reflections from the initial peak search.

### Turning FIT off

Generally, **once the refinement is progressing you don't need to turn it off.**



**Don't** set up multiple steps of fitting and fixing in an attempt to **refine a single parameter** at a time. That is neither necessary nor smart, since `GO` can fit multiple parameters simultaneously, and because single parameter refinement may not allow correlated parameters to be refined sufficiently.

When you really don't want to refine a parameter any more, for example to simulate an experiment, use the command **FIX**, as in **FIX crystal rotx roty rotz GO**. The crystal rotation parameters will now be fixed at their last refined position (unless you re-enter them manually) and will not refine on subsequent **GO** commands unless you specify **FIT** again.

### If things get worse

If the autoindexing succeeded, subsequent **refinement never should make things worse**. If it does, it is best to make sure that the autoindexing really did work after all. Go back and re-read the section on failures of autoindexing and try the suggestions given there. This is probably easier than trying to index your crystal manually.

### Finishing the Refinement

As a general guide, for automatic scanners the parameters are refined in this order:

```
CRYSTAL rotx roty rotz
X BEAM Y BEAM
UNIT CELL
CROSSFIRE x y xy
CASSETTE rotx roty
RADIAL OFFSET and ANGULAR OFFSET
Y SCALE and SKEW
DISTANCE
```

Not all parameters are refined for every crystal because some are highly correlated. In particular, the crystal-to-film distance, and the detector parameters **Y SCALE** and **SKEW** (or **RADIAL OFFSET**) are correlated with the unit cell dimensions. A particularly dangerous combination to be aware of is the simultaneous refinement of **DISTANCE**, **CELL**, and **RADIAL OFFSET**. These are so highly correlated that they can be refined only simultaneously for high quality, high resolution data. Generally, the greater the curvature of the Ewald sphere and the higher the symmetry of the crystal lattice, the better conditioned the problem. That is, correlations are weaker and parameters are more independent. Some simple rules for when you can safely refine all parameters together are given below:

Refine all parameters when resolution/ $\lambda$	$\leq 1.5$ for triclinic lattices
	$\leq 2.0$ for trigonal / tetragonal lattices

Otherwise don't refine **Y SCALE**, **SKEW** and **DISTANCE**.

If things are going well, you should see an **excellent** agreement between the diffraction spots and the preds. You can now extend your refinement to higher resolution. Here is a macro which finishes the refinement of one frame after autoindexing:

`refineone.dat`

---

```
RESOLUTION LIMITS 100 2.8
FIT crystal rotx roty rotz
GO GO GO
RESOLUTION LIMITS 100 2.2
FIT cell x beam y beam crossfire x y xy
FIT cassette rotx roty
```

---

---

**FIT** distance y scale skew  
**GO GO GO GO GO**

---

 $\lambda \leq$ 

A summary of the current values for all the parameters you have refined can be obtained by typing the command **LIST**. Now it's time to examine the refined predictions and adjust the mosaicity and spot shape.

### More about refinement - how it works

The integration of reflections requires knowledge of their indices and position. The weak reflections can be found only by predictions based on the information obtained from strong reflections. The autoindexing provides only the approximate orientation of the crystal, thus the result may be imprecise if the initial values of the detector parameters are poorly known. The least squares refinement process is used to improve the prediction. The parameters describing the measurement process either have to be known or have to be estimated from diffraction data by the refinement procedure. Depending on the particulars of the experiment, the same parameters (e.g. crystal to detector distance) are more precisely known, or are better estimated from the data. allows for the choice of fixing or refining each of the parameters separately. However, using it well requires considerable knowledge of diffraction experiments. Fortunately, **the FIT ALL option and detector specific default values seem to be reliable under most conditions.**

#### Eigenvalue filtering

Occasionally the refinement can be unstable because of a high correlation among some parameters. High correlation makes it possible for the errors in one parameter to compensate partially for the errors in other parameters. If the compensation is 100%, the parameter would be undefined, but the error compensation by other parameters would make the predicted pattern correct. In such cases **eigenvalue filtering** (described in as Singular Value Decomposition) is employed to remove the most correlated components from the refinement and make it numerically stable. Eigenvalue filtering works reliably when starting parameters are close to correct values but may fail to correct large errors in the input parameters. Once the whole data set is integrated, the global refinement (sometimes called **postrefinement**) can refine crystal parameters more precisely. The unit cell used in further calculations should come from the global refinement (in ) and not from refinement.

#### Partiality refinement

The refinement of the fraction partiality works by minimizing the function:

$$\sum \frac{1}{\delta_{(-)}^2} \left( \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} - \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} \cdot \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} \right)^2$$

where

$$\delta_{(-)}^2 = \delta^2 + \left( \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} \cdot \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} \right)^2 + \left( \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} \cdot \begin{matrix} \phantom{f} \\ \phantom{f} \end{matrix} \right)^2$$

$\delta^2$  = the sigma of the measurement of the reflection  
 = the intensity of the partially recorded reflection  
 = the fraction partiality  
 = the average intensity of the fully recorded reflections in this particular resolution shell  
 = the error in the positional measurement (see **ERROR** positional keyword)

The quantity f, as a function of the crystal parameters, is the variable in the equation. The other parts are determined by the observations and by the error model.

The detector and crystal parameters are refined by a **least squares method** that minimizes the deviation of the reflection centroids from their predicted positions. Such refinement by itself is seriously deficient when applied to a single oscillation image, since one crystal rotation parameter (rotation around spindle axis) is undefined and the others are highly correlated and/or poorly defined. To overcome this problem, another term (**partiality refinement**) is added within , in which the intensity of the partially recorded reflections is compared to the predicted partiality multiplied by an average intensity in the same resolution range.

The formula for the residual (difference between expected and predicted value) is the same as in the postrefinement; however, at this stage, the error of the predicted fully-recorded intensity is much larger, equal to the expected intensity. Nonetheless, the **concomitant positional and partiality refinement** used in `REFINE` is both stable and very accurate. The power of this method is in proper weighting (by estimated error) of two very different terms - one describing positional differences and the other describing intensity differences. The method leads to a reduced correlation between the detector and crystal parameters. An additional benefit is the uniform treatment of both detector and crystal variables in the whole refinement process.

## Mosaicity and Spot Shape

defines **mosaicity** as the smallest angle through which the crystal can rotate, about any axis or combination of axes, while a reflection is still observed. Typical values range from less than 0.1 degree to several degrees. Generally, the higher the mosaicity, the less useful the crystal. **Denzo does not refine mosaicity**, so you must input a reasonable value. The easiest way to do this is to simply increase the mosaicity (entering it as `MOSAICITY 0.5 GO`, for example) until all of the observed reflections are matched with preds. Thus, if you have:

- more observed reflections than preds ► **increase** the mosaicity.
- more preds than observed reflections ► **decrease** the mosaicity.

Useful increments are 0.1° for typical protein crystals and 0.02° for virus crystals.

Depending on how `display` you have set your display, you can get fooled into thinking that your mosaicity is higher than it really is, because you are trying to get preds to cover very weak reflections which result from diffuse or satellite scattering. To get around this, dim your display so that only the reflections representative of your Bragg scattering are displayed. Then adjust the mosaicity as above until a good match is seen.

A third, and more precise way of setting the mosaicity is to examine the histogram that produces in the log file (or outputs to the screen, if working interactively) after the `CALCULATE GO` statement is given (you can do this now, but it is the last operation you will be able to execute on this image unless you restart `denzo` since the `SECTOR` argument is incremented with the `CALCULATE` command, and thus a new frame is read; so hold off on this until you've done the best you can with matching up preds to reflections, and after you've adjusted the spot shape and size). The use of the histogram is discussed in the section `ANALYZE`, later in this chapter. To summarize briefly, you want to adjust the mosaicity so that the observed average-partiality histogram declines almost to zero in the last mosaicity zones.



**It is better to err on the side of a slightly high mosaicity** because this way you can be sure of catching all the spots. This is because `denzo` can re-label measured (partial) reflections as non-diffracting, but cannot do the opposite, i.e. make up a reflection which `denzo` did not find. Postrefinement with `REFINE` will calculate an accurate measure of mosaicity, and if the value used in `denzo` turns out to be low or way too high, the frames can be reprocessed with `denzo` using this corrected value for the mosaicity.

## Spot Shape and Size

When `denzo` actually fits a profile to each reflection and measures its intensity, it does so only to the portion of the reflection that is within the borders of the predicted spot (the 'pred'). If you choose a pred which is:

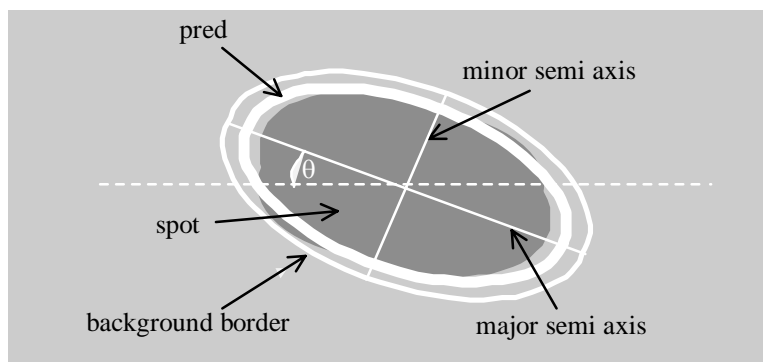
- too **small** ► intensities will be too low because part of the reflection was not measured.
- too **big** ► you run into the risk of:
  - a) rejecting reflections due to overlaps (overlaps are defined by whether or not the **predicted** spots, guard, or background regions are overlapped, not whether the actual spots are),
  - b) including portions of a neighboring reflection in the measured intensity, and

- c) including portions of the background in your intensity measurement. Therefore, it is wise to choose a pred size that matches those of your reflections as closely as possible.

Often the diffracted spots are slightly elliptical. They also vary in shape with the size of the crystal, shape of the X-ray beam, etc. allows you to change the size and shape of the pred spot to match the actual reflection better. The way to judge the correct spot size is to zoom in on an area of the image and compare the magnified shape of the actual spot with the magnified pred shape.

➡ You can't use different spot shapes for different regions of the image. If your spot shape varies over the image, you'll have to choose the best compromise. This is not so bad, though, because you pay only a small penalty for including the background in the spot when using profile-fitted intensities.

The keyword **SPOT** is modified by either **radius** or **elliptical**, followed by numbers. It is convenient to get used to using **SPOT elliptical**, because a circular spot can be described by an ellipse whose major and minor axes are equal and you will undoubtedly run into elliptical spots one day. A typical command for an elliptical spot would be **SPOT elliptical 0.6 0.5 0.0 GO**, where the first number is the length in mm of the major semi-axis, the second number is the length of the minor semi-axis, and the third number is the clockwise angle between a horizontal line and the major semi-axis. Once you have enlarged a region of your image, it is quite easy to find the appropriate spot shape. An idealized example is shown below:



**Figure 4:** DENZO spot terminology. Note that most of the time  $\theta$  is zero.

Make sure you adjust the **BACKGROUND** to match the size of the spot. In our example above, a reasonable value for the background is 0.1 mm larger than the spot in all dimensions, i.e. **BACKGROUND 0.7 0.6 0.0 GO**.

## Profile Fitting Radius

In addition to adjusting the spot shape and size, it is also a good idea to check and see if the **PROFILE FITTING RADIUS** is set properly. This is easy to do with the *ImageWindow*. In the *ImageWindow*, simply click on the **Prof fit R** button, and when you move the mouse you will see a white circle which corresponds to the **PROFILE FITTING RADIUS** set in the *ImageWindow*. The pred display changes now too. Instead of seeing all of the predicted reflections, you see only the preds that correspond to those reflections that are stronger than the weak level. Only this subset is shown because these are the only reflections used in the calculation of the average spot profile within the radius. Your radius must be large enough to include enough strong reflections for proper profile fitting, and not too large to average out significant profile variations. As a rule of thumb, you want the radius to include between 10 and 50 reflections. Useful increments of the profile fitting radius are 5% of the detector size.

## How it works: Integration of the Diffraction Maxima

To calculate the diffraction intensity, the detector **background** has to be estimated and then subtracted from the reflection profile. The standard method used to estimate the background value was to calculate an average detector signal in the neighborhood of a specific reflection. In it is assumed that the background is a **linear** function of the detector coordinates. Robust statistics (as discussed in the ) are applied to remove the contribution of pixels that deviate by more than  $3\sigma$  from the best fit to the background function. If too many background pixels are flagged as outliers the whole reflection is removed from integration.

ignores pixels in three other cases, when:

- they have been flagged as having no measurement by an auxiliary program,
- they have a special value (e.g. zero in the case of R-axis or MAR) or
- they are in the spot area (based on the predicted, rather than the measured position) of an adjacent reflection.

### Profile Fitting

The accurate prediction of spot positions is necessary to achieve a precise integration of Bragg peaks. The most important need for accurate prediction of the spot positions arises from the application of profile fitting.

Profile fitting is a two step process:

- First, the profile is predicted **based on the profiles of the other reflections** within a chosen radius. The predicted profile in is an average of profiles shifted by the predicted separation between the spots, so that they are put on top of each other. If the predicted positions are in error, then the average profile will be broadened and/or displaced from the actual profile of the reflection.
- In the second step the **information from the actual and the predicted profile is combined** by the following process:

The observed profile is a sum of the Bragg peak and background. (the estimate of ), is expressed by the formula:

$$= + \text{Const} * \quad (1)$$

where is the predicted value of the background and is the predicted profile.

Profile fitting **minimizes** function:

$$\sum_i \frac{(M_i - P_i)^2}{V_i} \quad (2)$$

where is a **variance** ( $\sigma^2$ ) of is a function of the expected signal in a pixel, which in the case of a counting detector is . The index represents all pixels in a two-dimensional profile; the same formulation of profile fitting applies to one and three-dimensional profiles. The predicted profile can be normalized arbitrarily; the most natural definition of normalization is that the sum of is equal 1. Such a choice makes the constant in expression (1) the fitted intensity , i.e. = Const

The solution to the profile fitting can be expressed by an alternative, but mathematically equivalent, approach, presented below.

Each pixel provides an estimate of the spot intensity equal with variance  $/^2$ . A profile-fitted intensity is then simply a weighted average of all observations:

$$I = \frac{\sum \frac{p_i^2 (M_i - B_i)}{V_i}}{\sum \frac{p_i^2}{V_i}} = \frac{\sum \frac{p_i (M_i - B_i)}{V_i}}{\sum \frac{p_i}{V_i}} \quad (3)$$

This approach (without an explicit solution presented in Eq.1) was first published by Diamond in 1969 for the one-dimensional case. However, in 1974 Ford proposed a simplified formula where  $V_i$  is constant. This was based on the mistaken idea that the variance of the optical density value of the exposed film is independent of the degree of X-ray exposure. Equation 3 thus became simpler:

$$I = \frac{\sum p_i (M_i - B_i)}{\sum p_i^2} \quad (4)$$

Many of the subsequent programs followed the Ford's formulation rather than Diamond's, even when applied to data collected with proportional counters or IPs. The **unweighted** formula proposed by Ford works quite well where the peak spot intensity is not much higher than the background intensity. This situation arises more often with data collected on film, which has a high intrinsic background and low saturation, or when the crystals have low scattering power due to a very large unit cell, high solvent content, or a disorder. The unweighted profile fitting improves the accuracy of the weak reflections compared to a straight summation, but at the cost of reducing the accuracy of the strong ones. This observation did lead in the past to a partial solution based on taking a weighted average between profile-fitted and summed intensities, where the weight is a function of the reflection intensity. The **weighted** formula (Eq.3) used in *Denzo* does not deteriorate the accuracy of strong, low resolution reflections. Thus, the observed problem with the unweighted formula is in the lack of weighting.

### Errors of the profile fitting

The profile fitting increases the precision (decreases the statistical error) of the measurement, but it may introduce an error due to lack of accuracy of the predicted profiles. *Denzo* applies the averaging of profiles in detector coordinates and, unlike other programs that use the profile-fitting method, **averages profiles separately for each spot**. This approach has two main advantages:

1. first, **only nearby spots are chosen for averaging**, ones which should have the most similar profiles.
2. second, ***Denzo* avoids interpolation** in the profile-prediction step; instead, it shifts the contributing profiles by vectors that make the smallest possible pixel-truncation error. These translation vectors precisely center the predicted profile on the reflection to be fitted and the error introduced by these shifts is smaller than that due to interpolation used in some other programs.

Diamond, R. 1974

Acta  
Crystallographica A25:43-  
55

Prediction of the profile shape is never exact, due to errors in the positional refinement, due to averaging of different shapes, due to truncation of pixel shifts or interpolation, etc. The resulting error of the fitted intensity was analyzed by Diamond in the case of one-dimensional Gaussian profiles and an unweighted profile-fitting formula. The important parameters are  $f$  - the root mean square (rms) width of the actual profile,  $w$  - the rms width of the predicted profile, and  $d$  - the displacement of the predicted profile from the actual profile. Let's define the relative change in the square of the reflection width:

$$\Delta^2 = \frac{f^2 - w^2}{w^2} \quad (5)$$

mxpe

(6)

The averaging of the profile used in the value of  $r^{2/3}$ , where  $r$  is the detector pixel size, to the value of  $r^{2/3}$  will increase the profile fitted intensity (or most reflections) by a constant multiplicative factor, which has little effect on crystallographic procedures. The interpolation broadens the profile by a factor dependent upon the position of the predicted reflection relative to the pixel boundaries. The interpolation will also increase  $r^{2/3}$  by a number between zero and  $r^{2/3}$ . The interpolation will in

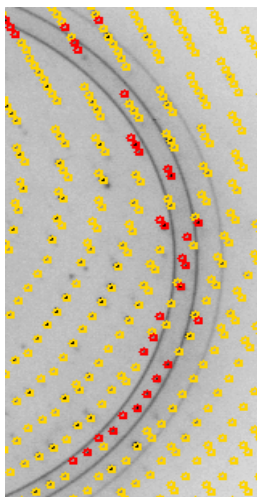


ice (non-protein lattice) peaks. You can tell if this worked by inspecting the image after autoindexing and then again after running the `@refine` macro. If the preds line up on real spots, then autoindexing worked despite the ice ring.

If the ice ring does interfere with autoindexing, here is a simple remedy: **adjust your resolution limits** to exclude the ice ring. This means you would edit the `auto.dat` macro to include the resolution limits. In the example provided above, the resolution limits were commented out. You could exclude the ice ring from autoindexing by setting the high resolution limit to about 4.1 Å. Once the initial orientation is found, all of the data can be included in the subsequent refinement.

## Ice rings in your data

There are two ways in `Denzo` to reject reflections corrupted by ice rings.



The elegant way exploits the fact that the **distinguishing feature of spots corrupted by ice rings is a very high and irregular background**. `Denzo` assesses the quality of the

background of each spot against criteria for slope and uniformity, and if a certain fraction of the pixels in the background fail this test, then the whole spot is rejected. This is controlled by the keyword `REJECT fraction`. The default value for this rejection is 0.75 (i.e. 75% of the background pixels must be acceptable). If you have an ice ring, you can increase `REJECT fraction` until most of the spots in the ice ring are no longer picked up. Useful increments are 1-2% at a time. Do this while you are looking at the display, just as you did for spot shape and mosaicity. As you increase the reject fraction, the number of good preds in the ice ring should decrease, while leaving most of the others alone.

The second way to avoid ice rings in your data is to simply **process your frames twice**, once at a resolution cutoff below the ice ring, and then again for the data beyond the ice ring. The two data sets are then merged in `Denzo` along the lines of Scenario 10 in The `Denzo` Manual. This is more of a brute-force approach and is likely to result in the loss of more good data than the approach outlined above.

## Refinement Statistics

After each `GO` command, `Denzo` not only updates the pred display, it also prints out the numerical summary of the refinement cycle:

1. The first part lists the **refined parameters**, their final values, and how much they changed in the last refinement cycle.
2. The last part lists the **number of spots** used in the refinement, the  $\chi^2$  values for the x and y positions of the preds, and the predicted decrease in the  $\chi^2$  values from the previous refinement cycle. If things refined well, the number of spots used in the refinement should be a significant fraction of the total number of spots on the image. This number reflects the number of preds that line up with observed spots. The  $\chi^2$  values represent the average ratio, squared, of the error in the fitting, divided by the expected error.

A good refinement will have  $\chi^2$  values near 1.0. In the early stages of refinement these values will be high, but towards the end they should be below 2 or so. The **predicted decrease** is a measure of the convergence of the refinement. When the predicted decrease is near zero, further `GO` commands will have no added effect. In other words, the current parameters have been refined as far as possible.

➡ **The magnitude of the  $\chi^2$  values is not a critical test for the success of *Denzo***, since they represent only the comparison of the spatial differences between the observed and predicted reflections to an error model that is typically quite strict in the prediction of small errors. Assuming default values of the error estimates,  $\chi^2$  values of 2 or even 3 are acceptable, because the position of the predicted reflection, and hence the intensity, is still very accurate. Thus, the  $\chi^2$  values may not reflect errors in the integration of the reflections.

It is a good idea to keep your error model roughly the same from crystal to crystal so that you can use the  $\chi^2$ s to pick up any other variations in your data collection. For example, if you are getting  $\chi^2$ s in the range of 2 to 5 consistently and your crystal is good, consider problems with the detector alignment.

**If the  $\chi^2$  values are very high**, (> 10 or so for default values of the error parameters) **something is seriously wrong** with the indexing, refinement, or the detector. This will be apparent from **visual inspection of the displayed image**, in any case.

Since the goal of `refine` is to produce a list of `hkl`s and unscaled intensities, it does not matter that the errors in the positions differ a bit from the standard expected error. On the other hand,  $\chi^2$  values substantially different from 1 in `refine` should be investigated, because this directly affects a critical result of the scaling procedure, namely the  $\sigma$  value assigned to each scaled intensity. So that's where you should pay attention to finer deviations of  $\chi^2$  values from 1, not in `refine`.

## Finishing Up

Once your first frame is indexed and you've adjusted the mosaicity, spot, and box sizes, it's time **to process the rest of your frames**. This can be done as a batch job or as a hands-off interactive session. Wladek and Zbyszek recommend that you take half an hour and watch your frames refine and integrate interactively, because this is the best way to spot any problems. For example, if you have misindexed the first frame by one (due to an incorrect beam position), you are likely to see it very quickly this way.

Since you still have your *DenzoWindow* and *ImageWindow* open, just continue on from there using the following refinement macro:

`refine.dat`

---

```

FIX all
START REFINEMENT
PRINT no shifts no profiles
                                (see below)

RESOLUTION LIMITS 100 2.8
FIT crystal rotx roty rotz
GO GO GO
RESOLUTION LIMITS 100 2.2
FIT ALL
GO GO GO GO GO
PRINT profiles 1 1
CALCULATE GO

```

---

This macro will refine all of the subsequent frames specified by the **SECTOR** in the macro `myexperiment.dat` (p. 26). You can just sit back and watch them zip along on the screen. The result will be a series of `.x` files called `xtal0##.x`.

Alternatively, you can process everything in batch mode using the following command file:

`denzo.com`

---

```

setenv cnverr yes
cd /people/mydir/DENZO
/usr/local/bin/DENZO << STOPINPUT >> xtal.log
@site.dat
@myexperiment1.dat
@refine.dat

```

---

---

```
END OF PACK
END OF JOB
STOPINPUT
```

---

Note that the file `myexperiment1.dat` has been modified from `myexperiment.dat` to include the final refined values of everything determined in the interactive refinement of the first frame, done above. The easiest way to do this is to examine the `xtal001.x` file produced by the `CALCULATE GO` statement. At the end of this file are all the final parameters determined from refinement. Cut these out and just paste them into the end of `myexperiment.dat` and rename as `myexperiment1.dat`. Since you are not running autoindexing in the batch mode, you have to take the results and paste them in by hand.

Warning: for unusual instruments (e.g. R-Axis with non-vertical spindle, synchrotrons) make sure that you have the correct value of **FILM ROTATION**. The defaults in the formats are only for the standard configuration of each instrument.

All that remains is to submit the job and examine the log file for the results. The processed frames in this case have the name `xtal0##.x`. (The `##` wild cards have been replaced by the **SECTOR** argument.) These ".x" files are ready to be scaled in .

`findbeam.dat`

---

```
BOX 5.5 5.5
SPOT elliptical 2.0 10.0 0.0
BACKGROUND radius 1.5
USE BEAM only
X BEAM 100.0 Y BEAM 102.0
FIT x beam y beam
GO GO GO
BOX 2.5 2.5
SPOT radius 1.0
BACKGROUND radius 1.0
GO GO GO
SPOT elliptical 0.6 0.7 0.0
BACKGROUND elliptical 0.7 0.8 0.0
USE NO BEAM
```

USE BEAM

RESOLUTION LIMITS

only

refine.dat

---

Another problem sometimes encountered with off-line scanners involves not the scanner, but rather the carousel on which the imaging plate cassettes are mounted during data collection. Sometimes these old war-horses, carrying their heavy load of IP cassettes, do not stop precisely at the zero position such that the imaging plate is perpendicular to the beam axis. In this case, care should be taken to refine the **CASSETTE** `roty` value.

## Understanding the DENZO Log File

The log file can be examined to see if the data processing went well. It is very long, since it contains a record of every refinement cycle. The interesting part comes after the **CALCULATE** statement at the end of the refinement. After printing out the final refined parameters, the following output is given. Notes are in green:

1. These numbers below are the maximum **h**, **k**, **l** expected for a given resolution range, the number of pixels needed to save the spot profile, and two computer memory statements.

```

hklmax= 99 99 99
93 pixels needed to save spot profile
maximal usage of the common /sfprof/ :283044

```

2. Below is a complete list of the suspect, but not rejected, reflections. The columns contain, in order: index, x and y coordinate of the reflection centroid, the oscillation angle at which maximum diffraction occurs (within the oscillation range), the measured intensity of the reflection, the normalized intensity of the reflection, the error on I, and then some flags which describe the problem with the reflection. There are only two types of problems flagged: negative intensity, and positional error. This list is useful because you can go back to the image and see what the problem is for each reflection.

h	k	l	x	y	osclang	intens	scalI	errI	xdel	errx	ydel	erry
-27	-84	-5	p	60.25	31.81	0.02	-797.9	-0.29	129.6	0.24	1	negative
-4	-28	2	p	96.13	91.63	0.50	1696.6	0.03	114.8	0.115	0.033	-0.274 0.036 0.0 1 position
15	2	-1	p	109.66	126.93	0.47	5601.3	0.03	122.1	-0.019	0.031	-0.129 0.031 0.3 1 position

3. Below is the mosaicity histogram for the reflections of the frame. The reflections are sorted into 20 zones, which range from minus the input mosaicity/2 to plus the input mosaicity/2. Each zone represents the shortest angular distance of the center of the reflection from the surface of the Ewald sphere at the end of the oscillation range. Only reflections that are single partials are included in the analysis. A reflection in a negative zones means that the center of the reflection has already passed through the Bragg condition. The fourth column lists the average observed partiality of the reflections in each zone, and the histogram is just a plot of the average observed partiality. In all cases, no matter what mosaicity you have input to Denzo, the histogram should pass through 50% in the zero zone. If the mosaicity you guessed/input was too small, then the histogram will not descend to zero at +mos/2. If the mosaicity you guessed/input was too large, then you will see a histogram which looks like a step function, falling to zero just past the zero zone. In the example below, the mosaicity was chosen correctly ( $2 \times 0.06^\circ = 0.12^\circ$ ), or slightly overestimated (which is the preferred side to err on) but there is some diffuse scattering which accounts for the small tail on the histogram. Note also that this is a particularly nice example because there are many contributing reflections to each zone. If there are only a few reflections, the histogram may be quite choppy. If the histogram does not resemble any of the shapes mentioned here, it may be indicative of motor, spindle, or shutter problems.

# refl	zone	av. part.	-10..0..10..20..30..40..50..60..70..80..90..100..110
169	-0.060	-0.054	0.927 *****
179	-0.054	-0.048	1.019 *****
155	-0.048	-0.042	0.929 *****
161	-0.042	-0.036	0.897 *****
164	-0.036	-0.030	1.005 *****
165	-0.030	-0.024	1.039 *****
176	-0.024	-0.018	1.040 *****
175	-0.018	-0.012	0.898 *****
176	-0.012	-0.006	0.700 *****
160	-0.006	0.000	0.607 *****
195	0.000	0.006	0.409 *****
192	0.006	0.012	0.217 *****
188	0.012	0.018	0.142 *****
163	0.018	0.024	0.091 ****
188	0.024	0.030	0.072 ***
162	0.030	0.036	0.055 **
193	0.036	0.042	0.056 **
186	0.042	0.048	0.042 **
167	0.048	0.054	0.041 **
181	0.054	0.060	0.041 **

4. The table below gives a summary of the types of reflections measured from the image. The first column is the type of reflection. The next three columns sort the reflections by strength: above the weak level, saturated (overflows), and below the weak level. Note that weak reflections are NOT ignored or omitted; they are just not used in the calculation of the average profile. Overlap tabulates the number of reflections that are flagged as overlapping (depending on whether you specified overlap spot, guard, or box), background overflow is the number of reflections with unmeasurable backgrounds due to saturation or excessive background editing, and others represent reflections which sit on the edge of the image and are lost but not outside the detector. Outside is a calculation of how many reflections are missing because they landed outside of the scanned image area (i.e. off the IP), for the specified resolution range. Remeasured is the same as total on the last cycle.

type	good	intns	ovfl	weak	overlap	backgr	ovfl	others	total	outside	remeasured
fiducials	0		0	0	0		1	0	1	0	1
partials	2004		60	1556	0		1	4	3625	850	3625
whole	2739		52	740	0		0	0	3531	273	3531
total	4743		112	2296	0		2	4	7157	1123	7157

#### 5. These are the final statistics for the positional refinement

spots refined: 4855  $\chi^2/\text{\#refl}$ : x 0.70 y 0.78 pred. decrease: 0.000 \* 4855 = 0.0

partiality refinement: 3620  $\chi^2/\text{\#refl}$ : 0.80 pred. decrease: 0.000 \* 3620 = 0.0

## Understanding the *DENZO* .x file

The product of *DENZO* is a file called the .x file, named for the suffix it is usually given. Each file represents the distillation of a single image. The .x file is organized into three consecutive parts: header, listing of reflections, and refined parameters.

Here is an annotated example:

ajit alpha toxin						title of the frame						
a matrix representation of orientation						u matrix orientation						
-0.00265575		0.00167628		-0.00863879		-0.069715		-0.985799		-0.152772		
0.008777937		-0.00028329		-0.00281712		-0.289696		0.166551		-0.942517		
-0.00067600		-0.00903865		-0.00142214		0.954577		-0.021450		-0.297193		
osc. start	osc. end	dist. in pixel units		wavelength	cryst rotz	roty	rotx	mosaicity				
0.00000	1.50000	1332.82227		1.54180	72.1579	178.6024	170.2019	0.2500				
these are the actual reflections now												
1	2	3	4	5	6	7	8	9	10	11	12	13
-27	7	5	1	249.8	231.0	1.82	20.4	0.919	29.6	684.4	0.172	1529.5
-27	7	4	1	68.2	50.5	1.69	19.6	0.920	37.6	702.4	0.208	1119.9
-23	5	15	1	125.2	208.5	1.73	25.5	0.922	49.9	462.7	0.261	1412.3
-22	5	17	1	306.9	301.4	1.65	35.9	0.919	53.8	414.9	0.354	1029.9
-23	5	14	1	1413.6	1409.2	3.28	68.4	0.924	56.7	481.6	0.224	1397.3
(... some lines skipped ...)												
▶ 0 -12 -32 1 118141 5399 24.77 -10031 0.997 1372.8 1196.3 0.191 3094.7 .												
1,2,3 h,k,l not reduced 4 flag 0 full, 1 partial 5 I by profile fitting 6 I by profile summation 7 χ² of profile fitting 8 σ of I 9 cosine of incidence angle at detector 10 x coordinate of predicted centroid, in pixel units 11 y coordinate of predicted centroid, in pixel units 12 Lorentz, polarization, obliquity factor 13 strength of averaged profile, arbitrary units												
In columns 5,6, and 8 the intensity of very strong reflections (i.e. overflows) will be written as an integer rather than as a floating point number - see example on line ▶ above. Scalepack understands this convention.												
999 17 17 999 → flag for end of reflection list, 17, 17 → size of the integration box (left)												
111111111111111111												
111111111111111111												
111111111111111111												
111111111111111111												
11111100200111111												
11111022222011111												
11110222222011111												
11110222222011111												
11112222222211111												
11110222222011111												
11110222222011111												
this is a saved spot profile for use in the display program if you want to overlay the final predictions on the original diffraction image. The "rectangle" is the box, 1's represent the background area 0's the guard area, and 2's the spot area												

```
11110222222201111
11111022222011111
11111100200111111
11111111111111111
11111111111111111
11111111111111111
11111111111111111
11111111111111111
```

---

```
error density 0.750
error positional 0.030
sector 1
raw data file 'alpha2_###.image'
profile fitting radius 40.00
overlap spot
monochromator 0.111
oscillation start 0.00 end 1.50
wavelength 1.54180
mosaicity 0.250
resolution limits 100.0 3.8
spindle axis 0 0 1 vertical axis 1 0 0
space group h32
unit cell 108.740 108.740 108.740 89.433 89.433 89.433
crystal rotx 170.202 roty 178.602 rotz 72.158
cassette rotx -0.16 roty 0.00 rotz 0.00
distance 199.92
x beam 89.194 y beam 89.582
film rotation 0.000
radial offset -0.202 angular offset -0.144
crossfire y 0.138 x 0.132 xy 0.000
```

these are the final parameters  
as refined from this image

useful for restarting data processing.

if clipping out, may start from here

# *Input Formats and Conventions*

The input to `shelx` consists of:

- **the digitized diffraction image** - a disk file of optical density readings from a detector. Data are organized as unformatted files. Headers, if any, are ignored. One record corresponds to a vertical or horizontal stripe (depending on the detector) one pixel wide, with one or two bytes per pixel. Pixel values are interpreted as integers which have no sign to them. The orientation conventions used by various detectors and the `ORIENT` convention are discussed below.
- **the control data** - a series of keywords followed by modifying words and/or some numbers and a few command words. These are usually written in a command file (macro file), although they can be typed from the keyboard.

All data are read in free format. Capital letters are converted to lower case unless in single quotation marks, so one should be aware when inputting case-sensitive UNIX file names. Some keywords contain a space in their name, for example `Y SCALE`. This space must be maintained and cannot break the line. Where a word or a number modifies a keyword, a space, comma, left or right parenthesis, carriage return or equal sign must separate the two. These characters are recognized as `SPACE` in `shelx`. Comments can be put in square brackets [as in this example]. If you wish to include a `CR` in a datum, for example in a title or in the name of a file, the whole datum must be put within single quotes - the title 'my crystal' must be enclosed in single quotes due to the space between the words "my" and "crystal". Similar problems crop up with file names. When in doubt or if the program says it can't read the data file, put the file names in single quotes.

A different type of parser pitfall is discussed for the keywords `END FIT` and `END FIX`. See the documentation under the keyword `FIT`.

Most of the **KEYWORDS** in `shelx` require that a modifier and/or a value follow. For example, in `CRYSTAL rotx -6.012` the keyword `CRYSTAL` is followed by the modifier `rotx` which is in turn followed by the value `-6.012`. Floating point and integer data should follow the FORTRAN convention. Text strings are limited to 80 characters. When an uninterpretable input is encountered, the program produces an error message and continues. Omitting a necessary input causes the program to write a warning to the output file and although the program may continue it is likely to stop with an error message or maybe crash.

Although there are a lot of keywords in `shelx`, the program provides defaults for most of the critical ones that are not specific to the crystal itself. These are listed in the documentation for each keyword in this Chapter and a more concise tabulation is presented in

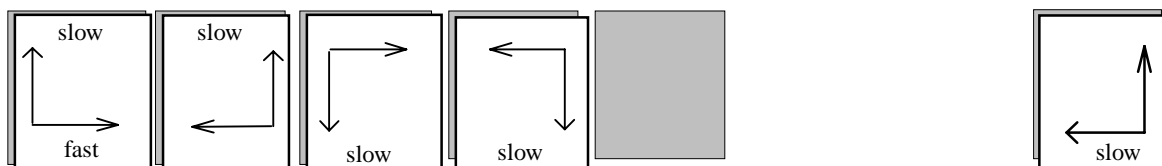
appendix, just after the keywords list. Keep in mind that **just because *Denzo* provides a default doesn't mean that it is best to use that default**. The sample command files in Chapter 2 and the listing in appendix give a listing of the keywords you should ordinarily state in your command files.

## Denzo Orientation Conventions

The standard display and processing convention for `denzo` is to show the diffraction pattern such that as you view the image from the crystal side of the detector, the spindle axis is horizontal and coming in from the right hand side. What follows is an explanation of how this corresponds to the detector data storage convention.

### Detector Conventions

A detector writes data to a file as a series of lines. The `line` coordinate is defined as the line number times the pixel (**RASTER**) size and the `col` coordinate is defined as the pixel position within the line times the pixel size in the `col` direction. The outer corner of the first pixel in the data file begins at `line = 0`, `col = 0`. Unfortunately, the data file convention doesn't correspond to the laboratory frame of reference because it is not standardized among detector manufacturers. Thus, there is a total of 8 possible combinations of `line` and `col` conventions. They are shown below:





the conventions. These are grouped as {1, 4, 5, and 8} and {2, 3, 6, and 7}. To use an X-ray film analogy, this is equivalent to looking at the film either from the back side (i.e. towards the X-ray source) or from the crystal side. handles this aberration by changing the sign of the **Y SCALE** keyword from positive to negative. If this is not done, then the assignment of the **Friedel mates (Bijvoet differences)** will be inverted, and thus the hand of the structure will be similarly inverted.

**Spiral scanners** Obviously, spiral scanners like the Mar scanner or some of the Mac Science scanners do not scan the IP in an orthogonal pattern. However, the controllers for these detectors convert the spiral data into Cartesian coordinates by using an interpolation algorithm. The conventions listed in the table above for these detectors are based on how the data is written in the file.

## Camera Conventions

It is inconvenient and non-intuitive to work with images in the data conventions, because we are used to thinking about diffraction geometry in terms of the x-ray beam, the spindle axis, and the laboratory frame of reference (horizontal, vertical). There are two possible conventions within this "intuitive" set:

### Gravity/beam system

In this system, the X-ray beam is defined as being perpendicular to the direction of the force vector due to gravity ("down"). The beam is the **beam** axis, gravity is parallel to the **vertical** axis, and the **horizontal** axis is perpendicular to both of these. In terms of crystal orientation rotations

- **rotz** would denote rotations of the crystal around the **beam** axis,
- **roty** would denote rotations of the crystal about the **vertical** axis and
- **rotx** would denote rotations of the crystal about the **horizontal** axis.

This convention has the advantage that it does not change with different camera spindle geometry. In addition, it corresponds directly to the common orthogonal detector directions, in that the detector directions x and y on the film or IP are perpendicular to the x-ray beam. A disadvantage of this system is that the x axis may or may not correspond to the spindle axis, depending upon the  $\chi$  (chi) and  $\Omega$  (omega) setting angles of the 3-axis goniostat. For the purposes of aligning your crystal, this makes life more complicated, as we are used to thinking about aligning crystals by moving the arcs on a goniometer head. For  $\chi$  equal 90 or 270 degrees and  $\Omega$  equal to zero, however, the spindle axis is parallel to the x axis and thus **rotx** corresponds to rotations about the spindle axis.

Although this system will not (formally) work in Outer Space (where the gravity vector is essentially zero), nor in the event that the x-ray beam is vertical, the fact remains that (so far) all macromolecular crystallography is done with horizontal x-ray beams on the surface of good old planet Earth, and thus it is a viable option. While does not use this convention, does. **Hey, life is messy...**

### Spindle/beam system

The spindle/beam system is the crystal and cassette orientation convention.

In this system,

- ⇒ the **z** axis is again parallel to the **beam**,
- ⇒ the **x** axis is parallel to the **spindle** axis and
- ⇒ the **y** axis, also termed the **"vertical"** axis (for want of a better word), is perpendicular to the spindle and beam axes.

In terms of crystal orientation rotations:

- **rotz** would again denote rotations of the crystal around the **beam** axis,
- **rotx** would denote rotations of the crystal about the **spindle** axis and
- **roty** would denote rotations about the axis perpendicular to the beam and the spindle.

The advantage of this system is that it is very intuitive to the crystallographer. In addition, with Eulerian  $\chi$  values of 0, 90, 180 or 270 degrees the spindle and vertical axes are parallel to the x or y axes of the detector. The main disadvantage of this system is that it is dependent on knowing the geometry of the camera.

- ➡ While the crystal and cassette orientations follow the spindle/beam convention for  $\chi$ , the beam, box, box printout, spot, margin, film width and length in the  $\chi$  log file follow the data convention.

### ***Determining the values of FILM ROTATION and Y SCALE***

As discussed above, the sign of **Y SCALE** depends upon the data storage convention of the instrument. As was shown in the table under the Figure 5 in **Detector Conventions** analysis, to date about 50% of the detectors have a negative values for **Y SCALE**. The default values in  $\chi$  are good for all automatic detectors (MacScience DIP2000 series, R-axis, Mar) in their standard configurations.

**FILM** rotation depends on the spindle orientation. The standard  $\chi$  convention is to display the image such that looking along the x-ray beam, the spindle axis is horizontal and comes in from the right hand side (Eulerian  $\chi=90$ ). In such situation **FILM** rotation = 0. If the spindle is at any other  $\chi$  value **FILM** rotation =  $\chi - 90$

- ➡ On the Raxis, the data is  $\chi$  by 90 degrees when written to the data file. This is a convention specific to this machine. Thus, for Raxis images, **FILM** rotation =  $\chi - 180$
- ➡ If the film or IP is rotated about the beam axis prior to scanning, then the value of film rotation will have to be determined by noting the angle of rotation.

# Alphabetical Denzo Keyword Descriptions

This is a listing of all of the

keywords and their modifiers, if any:

AIR ABSORPTION LENGTH	radius	solver; no solver
ANGULAR OFFSET	crossfire x y xy	statistics; no statistics
BACKGROUND	all	PROFILE FITTING RADIUS
radius	FORMAT	QUIT
elliptical	see <i>Detector Specific Info</i>	RADIAL OFFSET
BAD VALUE	GO	RASTER
BOX	IBOX	RAW DATA FILE
CALCULATE	IGNORE	RECORD LENGTH
CASSETTE	IMAGING PLATE	REFINE PARTIALITY
flat	INPUT / @	REFINE NO PARTIALITY
radius	INTENSITY AVERAGING WINDOW	REJECT
rotx, roty, rotz	LAMBDA	fraction
CELL	LATTICE	cutoff
COUPLING	LEFT MARGIN	slope
CROSSFIRE	LIST	systematic
x, y, xy	LONGEST VECTOR	RESOLUTION LIMITS
CRYSTAL	MONOCHROMATOR	SECTOR
rotx, roty, rotz	filter	SKEW
DI STANCE	graphite	SPACE GROUP
EIGENVALUE FILTER	MOSAICITY	SPINDLE AXIS
END FIT	NO BEAM	SPOT
END OF JOB	NUMBER OF RECORDS	radius
END OF PACK	OMIT	elliptical
ERROR	radius	START REFINEMENT
density	elliptical	STOP
overload	ORIENTATION AXIS 1	SWAP BYTES (obsolete)
partiality	ORIENTATION AXIS 2	TITLE
positional	OSCILLATION	TOP MARGIN
systematic	start	UNIT CELL
y/x factor	end	USE PARTIALS POSITION
FIDUCIALS	range	USE BEAM
FILM	step	USE FIDUCIALS
absorption	OVERLAP	USE NO FIDUCIALS
length	spot	USE PARTIALS POSITION
output file	guard	USE PARTIALS NO POSITION
rotation	none	VERTICAL AXIS
width	OVERLOAD VALUE	WAVELENGTH
FIT/FIX	PLATE ROTATION	WEAK LEVEL
a* b* c*	PEAK SEARCH FILE peaks.file	WRITE PREDICTIONS
alpha*, beta*, gamma*	POLARIZATION	WRITE NO PREDICTIONS
crystal rotx roty rotz	PRINT	X BEAM
cassette rotx roty rotz	chi square; no chi square	Y BEAM
distance	fitted	Y SCALE
x beam	hkl	YORK OUTPUT FILE
y beam	memory; no memory	
y scale	profiles; no profiles	
skew	shifts; no shifts	
film rotation		

**AIR ABSORPTION LENGTH**

Length of air or helium, in mm, over which the scattered X-rays are attenuated by a factor of  $e$  (2.718)

---

**AIR ABSORPTION LENGTH** *value\_(mm)*

**AIR ABSORPTION LENGTH** 860

**AIR ABSORPTION LENGTH** 200 [for 2.5 Å X-rays]

---

**ANGULAR OFFSET**

This is one part of the alignment correction needed for the spiral MacScience and MAR scanners. It is the distance, in mm, that the reading head is offset in the non-radial scanning direction from the center of the imaging plate. See also **RADIAL OFFSET**.

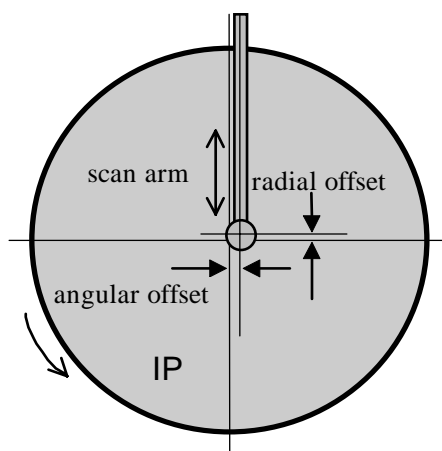
---

**ANGULAR OFFSET** *value\_(mm)*

**ANGULAR OFFSET** 0.0

**ANGULAR OFFSET** 1.2

---

**BACKGROUND**

Defines the shape and size of the background area where the background measurement takes place. Usually defined to be outside of the spot area. However, when no background is specified, the default area is the area just outside of the spot border.

---

**radius** defines the background to be circular with a radius value, in mm. The center of the background area and of the box coincide.

**elliptical** defines the background to be ellipsoidal with characteristic major and minor semi-axis lengths (in mm), along with the clockwise angle between a horizontal line and the major semi-axis. Refer to the diagram under the keyword **SPOT**.

---

**BACKGROUND** *radius value\_(mm)*

**BACKGROUND** *elliptical value-major\_(mm) value-minor\_(mm) value-angle\_(degrees)*

a ring just outside of the spot border

**BACKGROUND** *radius* 0.45

---

**BACKGROUND** elliptical 0.80 0.70 0.0

---

## BAD VALUE

The value in the image file which does not represent a measurement. Typically 0. When set to 40000 it disables 0. Used only on DIP series scanners to process small molecule data.

## BOX

Defines the size of the box where both the spot and the background are measured, given as two lengths, in mm. If the **BOX** size is larger than the average separation between spots (i.e. the boxes overlap), then the program will be slowed down by the spot overlap calculation. The current limit on the box size is 90 pixels in each direction, for a total of 8100 pixels.



The box size can alternatively be specified by the **I BOX**, where the size is given in pixels instead of mm.

---

print	tells the program to print the spot definition at the beginning of every refinement cycle (i.e. for every <b>GO</b> statement)
-------	--

---

**BOX** *value1\_(mm)* [slow scan direction] *value2\_(mm)* [fast scan direction]

**BOX** print *value1 value2*

none

**BOX** 2.4 2.4

---

## CALCULATE

This command tells the program that it is the last cycle, when the spot intensities are calculated and written to the film output file. It is followed by the command **GO**. After the cycle both the **RAW DATA FILE** and the **FILM OUTPUT FILE** are closed. Signals the **SECTOR** argument to be incremented by 1.

## CASSETTE

Defines the orientation and the shape of the detector, historically, a film cassette.

---

rotz, roty, rotx	the three angular deviations from zero
------------------------	--

---

radius radius of the film in a curved cassette



Radius of curvature does not have to be equal to the film-to-crystal distance, and may be different than the radius of the cassette due to the thickness of the film or imaging plate.

---

flat	turns off radius
------	------------------

---

**CASSETTE** rotx *value1* roty *value2* rotz *value3\_(degrees)*

**CASSETTE** radius *value1\_(mm)*

all angles = 0.0

**CASSETTE** rotx 2.0 roty -1.0

---



You rarely have to enter these values since the program can refine them once the crystal parameters are established. However, some computer time can be saved if previously

determined values are used. Usually they change significantly only after the camera is realigned. There is no limit on values.

➡ **The cassette rotations are the primary means by which a non-zero  $2\theta$  angle is specified** (along with **DI STANCE**, **X BEAM** and **Y BEAM**). For example, a  $2\theta$  setting of, say, +20 degrees (R-Axis convention), would be described by **CASSETTE rotx -20.0** (note change in sign). The **DI STANCE** will have to be changed as well, since it is the distance from the crystal to the point where it intersects the IP surface. This can be calculated by the formula: Distance ( ) = distance (R-Axis) /  $\cos(2\theta)$ . **X BEAM** and **Y BEAM** are the position of the direct beam on the IP during exposure, not at  $2\theta = 0$ . For large values of  $2\theta$  the **Y BEAM** may be off the detector. The correct value may be negative or more than 200 in such cases.

➡ **CRYSTAL rotz** and **CASSETTE rotz** are strongly correlated and **cannot be refined simultaneously**.

## CELL

Alias for the keyword **UNIT CELL**

## COUPLING

This value reflects how fast the detector moves during exposure for Weissenberg images. Given as mm per degree. This is the **inverse** of the Photon Factory parameter degree/mm, and is only useful for Weissenberg images.

---

**COUPLING** *value\_(mm/degree)*

0.0

**COUPLING** 0.057

---

## CROSSFIRE

Crossfire is a measure of the X-ray beam divergence and focusing as it leaves the collimator and illuminates the crystal. Crossfire, being a symmetric tensor, has x, y, and xy components. It affects the prediction of partial reflections and their positions, not their angular width. It is expressed as angular divergence of the beam. The default value is zero crossfire, i.e. a perfectly parallel beam or a beam focused on the detector.

---

x	angular spread of the beam in the x direction
---	---

y	angular spread of the beam in the y direction
---	---

xy	correlated xy component of the beam spread. This tends to be zero within error.
----	---

---

**CROSSFIRE** *x value1\_(arbitrary\_units)*

**CROSSFIRE** *y value2\_(arbitrary\_units)*

**CROSSFIRE** *xy value3\_(arbitrary\_units)*

0 0 0

**CROSSFIRE** *x* 0.023

**CROSSFIRE** *y* 0.345

**CROSSFIRE** *xy* 0.0004

**CROSSFIRE** *x* 0.023 *y* 0.345 *xy* 0.0004

---

**CRYSTAL**

Defines the crystal orientation for the calculation of predicted spots.

---

rotz,	the three angular deviations from the reference orientation specified
roty,	by <b>SPINDLE AXIS</b> and <b>VERTICAL AXIS</b> . See the diagram and further
rotx	description in .

---

**CRYSTAL** rotz value1 roty value2 rotx value3\_(degrees)

0 0 0

**CRYSTAL** rotx 2.34 roty -4.566 rotz 34.560

---

**DISTANCE**

The distance from the crystal to the beam spot on the film, in mm, along the X-ray beam path, regardless of the  $2\Theta$  angle.

---

**DISTANCE** value\_(mm)

**DISTANCE** 75.0

**DISTANCE** 155.453

---

**EIGENVALUE FILTER (USE IT WITH GREAT CAUTION)**

When attempting to simultaneously refine two or more highly correlated parameters with , an eigenvalue filter is triggered and an error message appears in the log file stating that the parameters being fit were too highly correlated to be fit simultaneously. The eigenvalue filter value provides a way of defeating this check, by lowering the value. Unfortunately, just because you change the value doesn't mean that the correlated parameters are going to become uncorrelated, so changing the value is unlikely to help you and may lead to some horrible result.

---

**EIGENVALUE FILTER** value

0.001

**EIGENVALUE FILTER** 0.0005

---

**END OF JOB**

Signals the end of the job. Placed at the end of the command file or used to exit from during manual fitting and indexing.

**END OF PACK (REDUNDANT)**

Signals the **SECTOR** argument to be incremented by 1. The **CALCULATE GO** statement does now the previous work of **END OF PACK**.

**ERROR**

The keyword **ERROR** and its modifiers defines the model for the error in the measurement of the spot positions. **The default values should be adequate in most cases** and you should not worry too much about them. The error model does, however, directly affect the  $\chi^2$  values printed out at the end of each cycle of refinement. Since this is usually the only check you can reasonably make as to the success of your data reduction in , it is wise to know

something about how these are calculated. The names of the modifiers contributing to the error definition should not necessarily be taken literally, since they are short hand notation for complicated effects on the accuracy of the measurement. The model has been developed and black boxed (i.e. given the Z.O. Good Housekeeping seal of approval) to give a realistic estimate of the weight each spot should contribute to the refinement of the crystal and camera coordinates. This model gives  $\chi^2$  values close to one for good (non-mosaic) crystals. For mosaic crystals either error positional needs to be increased or higher values than 1 of  $\chi^2$  should be expected.

---

positional	The <b>minimum</b> error, in mm, in the positional measurement. It is usually a fraction of the pixel size, typically one half or one quarter of a pixel. Its squared value is added to the other components of the error estimate. Enlarging this value will lower your $\chi^2$ values, but will not necessarily improve your data. ► if you have to increase this value much for a well-tested instrument like MacScience or R-axis, check the instrument alignment.
density	Models the error measurement of the optical density of the film. The error equals the square root of the corrected reading times the density value. This is used only to estimate the error in spot positions and in the definition of strong reflections. ► the error density is affected by the voltage on the photomultiplier. The higher the voltage, the higher the error density, with a very steep dependence.
overload	Refers to the increase in the modeled positional error due to the existence of pixels with intensity overflow. For every overflowed pixel the modeled error squared is increased by the overload value squared.
y/x factor	This reflects the increase in the positional error of the spot due to its being close to the oscillation axis. The error estimate is by 1 + the ratio of y/x factor to the absolute value of the tangent of the angle between the oscillation axis and the line connecting the spot with the beam spot. This prevents the spot distortion which occurs close to the oscillation axis from affecting the refinement very much.
partiality	the expected error of the partiality determination The value for the error is expressed as a decimal, e.g. $\pm 10\%$ would be written 0.1.
systematic	the weight given the positional refinement compared to the partiality refinement when refining crystal orientation parameters. The higher the value, the more weight is given to the positional refinement. Typical values range from 1 to 20. A value of 5 seems to work almost all the time.

---

**ERROR** modifier1 value modifier2 value ...

overload 0.005

density

positional 0.030

y/x factor 0.1

partiality 0.10

systematic 5.0

**ERROR** systematic 5.00 partiality 0.100 positional 0.050

---



- ➡ **Avoid modifying the error parameters other than positional and density.** Moreover, error density should be constant for a detector (site) and changed only with caution.

## FIDUCIALS

Defines the position of the fiducial marks on the film relative to the beam spot, before **FILM** rotation and **Y SCALE** are applied. Given as a series of , coordinates, in mm. These days this is only used with Photon Factory Weissenberg images.

---

**FIDUCIALS** *xvalue\_(mm) yvalue\_(mm) xvalue\_(mm) yvalue\_(mm) ...*

none

**FIDUCIALS** -90.0 -90.0 90.0 -90.0

---

## FILM

This keyword tells the program about the characteristics of the data collection medium. Identical to the keyword **IMAGING PLATE**.

---

absorption	Total absorption of the film. Default is 100 for image plates
length	Distance from the top of the detector to the last useful pixel in the <b>fast</b> direction. Depending on the detector, on the computer display it could be the apparent width of the detector.
width	Width of the detector in the <b>slow</b> direction. Depending on the detector, on the computer display it could be the apparent height of the detector.
rotation	Describes the orientation of the <b>spindle</b> relative to the detector. The value of the <b>FILM rotation</b> is function of which file storage convention (see figure 5 in <i>Detector Conventions</i> ) is being used.  For manual scanner, adding 180° to film rotation can compensate for putting plate upside-down into scanner. Small changes (refined by ) to film rotation can compensate for small rotation slop that is inherent when manually handling plates.
output file	This gives the name of the processed output file

---

**FILM** modifier1 value ...

output file format is a string of characters

absorption 100

length *detector dependent*

rotation *detector dependent*

output file (none)

**FILM** rotation 270.0

**FILM** absorption 2

---

***FIT, FIX, END FIT, END FIX***

These command words tell the program to **FIT** (refine) or **FIX** (not refine) a parameter specified by a keyword. The **FIT** command is followed by the command **GO**, which calls one cycle of refinement. One **GO** command calls one cycle of refinement, 5 **GO** commands call for 5 cycles of refinement, etc. If a keyword parameter is fixed, it will retain its current value through the refinement. **The program assumes that all of the keyword parameters are fixed at the beginning of every new frame.**

The effect of these commands is cumulative and carries over through subsequent cycles of refinement. For example, if the command **FIT CRYSTAL rotx roty rotz GO** is given, then on every subsequent refinement cycle (called by the **GO** statement), the **CRYSTAL rotx roty** and **rotz** parameters will be refined. Typically a refinement starts by fitting just a few parameters, like **CRYSTAL rotx roty rotz**, allowing the refinement to proceed towards convergence for a few cycles, and then fitting more parameters a few at a time. The keyworded parameters that can be **FIT** are given below.

- All of the keyworded parameters can be refined simultaneously, so care should be taken not to fit parameters which are too strongly correlated at low resolution. For example, the crystal to film distance and the unit cell lengths are highly correlated, so one should be allowed to refine before refining the other (usually the unit cell lengths are refined first).

The **END FIT** and **END FIX** command words mark the end of a list of fitted parameters. This is especially useful for inputting more parameters after you have called a **FIT** command.

- Consider the following case: you want to refine the crystal rotation values, so you type **FIT CRYSTAL rotx roty rotz**. Now let's say that you also want to enter a new value for the distance before giving a **GO** command. You might think to try **FIT CRYSTAL rotx roty rotz DISTANCE 200 GO**, but this won't work! Program will not recognize the 200, and it will give you an error message, plus it will try to refine the distance value as well. The solution is to use the **END FIT** command, as follows: **FIT CRYSTAL rotx roty rotz END FIT DISTANCE 200 GO**. The same goes for the **END FIX** command.

The option **FIT cell**, when used in conjunction with the **SPACE GROUP** or **LATTICE** command, tells the program to fit the unit cell lengths and angles. The program is smart enough to know which lengths and angles are to be fit, and which are to be fixed. For example, it knows enough to keep all angles at 90 degrees in orthorhombic space group. Likewise, it also knows enough to keep  $a = b$  in tetragonal space groups, etc. The **LATTICE** or **SPACE GROUP** must be declared before **FIT CELL** can be used.

In practice **most often option used is FIT all**. This will fit all parameters that can be sensibly fit given the circumstances (space group, detector). Fitting all parameters is now stable even at low resolution.

<b><i>FIT- FIX</i></b>	all, a*, b*, c*, alpha*, beta*, gamma*, cell, <b>CRYSTAL</b> rotx roty, <b>CASSETTE</b> rotz rotx roty, radius, <b>DISTANCE</b> , <b>X BEAM</b> , <b>Y BEAM</b> , <b>Y SCALE</b> , <b>SKEW</b> , <b>FILM</b> rotation, <b>CROSSFIRE</b> x, y, xy
	<b>FIT</b> <i>KEYWORD</i> modifier
	all keyword parameters fixed (i.e. type <b>GO</b> and nothing will be refined)
	<b>FIT CRYSTAL</b> rotx roty rotz
	<b>FIT X BEAM Y BEAM Y SCALE SKEW DISTANCE GO GO GO</b>

❖ **Do not run alternative cycles of fit and fix commands!** This is the old way and does not take advantage of the fact that the refinement is very stable and can refine many parameters simultaneously. Simultaneous refinement is the best way to avoid getting into a false minima.

## FORMAT

A critical keyword. The modifier following this keyword tells the program about a series of defaults to use for the detector parameters. This keyword can thus substitute for many of the parameters called by, for example, the **FILM** (or **IMAGING PLATE**) keyword.

A detailed table of all the keywords specified by the format keyword can be found in the and appendixes.

---

see appendix

**FORMAT** modifier

this is a required input!

**FORMAT** dip2000

---

## GO

This command tells the program to execute one iteration cycle of refinement. After the cycle the fitted parameters will have new values. You can specify more than one **GO** in one line. This will tell to print results of the fit only for the last cycle from the series of **GO**s. See also **FIT/FIX**.

## IBOX (REDUNDANT)

**IBOX** is exactly the same as the keyword **BOX**, except that instead of giving the dimensions of the box in mm, the dimensions of the box are given in pixels. The **I** in **IBOX** stands for . The maximum dimensions of the box are 90 pixels in each direction.

---

**IBOX** *integer integer\_(pixels)*

none

**IBOX** 25 25

---

## IGNORE

Tells the program to ignore a specific region of the X-ray image. For example, this is useful for excluding the shadow due to the beam stop. Don't try to use this for masking out ice rings. That is accomplished using the **REJECT** fraction keyword. It is useful to think of **IGNORE** as describing a masked-out area of the film. The mask is provided by a series of geometrical descriptions, followed by their coordinates. For example, the beam stop shadow could be masked out by a combination of a quadrilateral and a circle (i.e. a lollipop shape).

---

quadrilateral A quadrilateral as defined by four , coordinates, in mm, picked in a **counterclockwise** (if **Y SCALE** > 0) or **clockwise** (if **Y SCALE** < 0) direction. The best way to do this is to use the cursor to pick the points off of the display window. The coordinates in mm are listed in the bottom row of the green

---

	box of the display.
circle	defined by the coordinates of the center, and the radius of the circle, all in mm.
triangle	defined by three coordinates. Use the display program to pick these.
<b>I G N O R E</b>	quadrilateral $x_1$ $y_1$ $x_2$ $y_2$ $x_3$ $y_3$ $x_4$ $y_4$ (all values in mm) circle $x_5(mm)$ $y_5(mm)$ $r_(mm)$ triangle $x_6$ $y_6$ $x_7$ $y_7$ $x_8$ $y_8$ (all values in mm)
	nothing ignored
<b>I G N O R E</b>	quadrilateral 50 0 52 0 52 100 50 100

---

**IMAGING PLATE (REDUNDANT)**

This substitutes 1 for 1 with the keyword **F I L M**. **F I L M** for all the details.

**INPUT / @**

This keyword is the same as the **@** keyword. Signals to the program to change reading of the control data to an auxiliary file. At the beginning of the program the auxiliary file is closed, so the keyword **I N P U T** has to be followed by the filename of the auxiliary file. The auxiliary file is read until a semicolon is reached (or an end of file mark is detected). The control is then returned to the original input stream. In the case of reaching a semicolon, the auxiliary file remains open, and the next time the keyword **I N P U T** appears it must not be followed by a file name or else the program will become confused, since it still wants to read from the open auxiliary file. On the other hand, if control is returned due to an end of file mark, then the next time input is encountered it must have an auxiliary file name next to it. If you give it the same auxiliary file name as before, then it will begin reading the auxiliary file from the beginning. Nesting of auxiliary files is not allowed. Auxiliary files are useful to keep the data common to many films in a separate file.

---

<b>I N P U T</b> <i>filename</i>
none
<b>I N P U T</b> subbatch1.dat
<b>@</b> refine

---

**INTENSITY AVERAGING WINDOW**

Sets the number of reflections in a resolution shell used to define that resolution shell for the purpose of refining partiality. Default is 100.

**LAMBDA (REDUNDANT)**

X-ray wavelength. Same as keyword **WAVELENGTH**.

---

wavelength in Ångstroms	
1.5418	[Cu K $_{\alpha}$ ]

---

---

LAMBDA 1.5418

---

## LATTICE (REDUNDANT)

Lattice type. Note that this command is superseded by the **SPACE GROUP** command for most routine use.

---

single letter or single letter with name. Can also specify lattice type. Possible choices are: P, C, A, B, F, I and R

P triclinic

LATTICE P monoclinic

---

## LEFT MARGIN

The left margin, in mm from the left edge, of useful data. It is defined in the raw data convention, not in the display program convention.

---

LEFT MARGIN *value\_(mm)*

0.0

LEFT MARGIN 2.0

---

## LIST

This command is normally used in the manual mode. It tells the program to provide a list of the current parameters. This is convenient if you forget where you were in the refinement, or you want to copy down the latest parameters. To use, just type **LIST**.

## LONGEST VECTOR

In the autoindexing subroutine of , **LONGEST VECTOR** denotes the longest real-space vector, in Å, the program expects to find in the reduced, primitive, unit cell. If the autoindexing fails, and you are confident of your **X BEAM**, **Y BEAM** and **DISTANCE** values, then this is the next parameter to try changing. Typically you would either input a value close to or somewhat larger (say  $\approx 50\%$ ) than what you really suspect is the longest vector in the unit cell.

---

LONGEST VECTOR *value\_( $\approx$ )*

LONGEST VECTOR  $\approx \frac{\lambda \cdot \text{distance}}{2 \cdot \text{spot radius}}$

LONGEST VECTOR 300

---

**MONOCHROMATOR**

Defines the polarization of the X-ray beam. Same as the keyword **POLARIZATION**. This correction is not a big deal except for synchrotron radiation, where it is a minor deal unless you are collecting high angle data. The definition of polarization is as follows:

$$\frac{\begin{pmatrix} - \\ + \end{pmatrix}}{\begin{pmatrix} - \\ + \end{pmatrix}}$$

where  $I_{\parallel}$  is the intensity of the electric field squared of the X-ray beam along the spindle axis, and  $I_{\perp}$  is the intensity perpendicular to the spindle axis. Values range from +1.0 to -1.0.

Generally speaking, this is not a number you can measure very easily, and is often calculated theoretically. It is really the responsibility of beam line operators to know this value for their beam lines. Note that the **sign of the correction depends on whether the spindle is horizontal or vertical**. As a rule of thumb, if the polarization correction is  $> 0$ , then the intensity of the X-ray scattering is stronger perpendicular to the axis of the spindle than it is to the axis parallel to the spindle. You can check the background of your image to confirm this. Typical values range from  $\pm 0.11$  for graphite to  $+ 0.9$  for synchrotron beams (the spindle axis is typically oriented at synchrotrons to make the polarization positive). For a **laboratory source**, use modifiers: `graphite` (if graphite monochromator is installed) or `filter` (if X-ray beam is filtered or mirrors are installed).

<code>filter</code>	No polarization, value = 0
<code>graphite</code>	Reflecting plane perpendicular to the plane formed by the spindle and the beam, CuK $\alpha$ and MoK $\alpha$ radiation. Value of $\pm 0.11$ for CuK $\alpha$ or $\pm 0.03$ for MoK $\alpha$ .
<b>MONOCHROMATOR</b> <code>modifier</code>	
<code>0.0</code>	
<b>MONOCHROMATOR</b> <code>graphite</code>	
<b>POLARIZATION</b> <code>-0.11</code>	

➡ The old keyworded values for the synchrotrons are probably wrong now. **Always ask the synchrotron staff what the correct number is.** A good guess for a bending magnet source is polarization 0.9. For a wiggler source it can vary a lot from station to station, in the range of 0.0 to 0.9.

**MOSAICITY**

Defined in `mosaicity` as the rocking angle, in degrees, in both the vertical and the horizontal directions which would generate all the spots seen on a still diffraction photograph. It includes contributions due to X-ray bandwidth, beam crossfire, etc.

<b>MOSAICITY</b> <code>value_(degrees)</code>
this is a required input!
<b>MOSAICITY</b> <code>0.5</code>

**NO BEAM**

Tells the program not to use the beam spot in the refinement. If the option **USE BEAM** has been invoked earlier in the refinement, it should be revoked with the **NO BEAM** keyword in the later stages.

---

**NO BEAM**

this is the default option. see **USE BEAM**

---

**NUMBER OF RECORDS**

Specifies the size of the data file. Equal to the number of vertical (fast scanning direction) lines. The default value is now 'hard wired' into the **FORMAT** modifier. In other words, once you specify the detector, the **NUMBER OF RECORDS** value is input from a look-up table.

---

**NUMBER OF RECORDS** *integer*

comes along with whatever **FORMAT** is specified

---

**OMIT**

Defines the area **outside** a specified circle or ellipse, but **within** the box, to be omitted from any use. By default there is no omitted area.

---

radius	or	Defines the omit boundary to be circular with a radius value, in
rad		mm. The center of the omit boundary and that of the box coincide

elliptical	Defines the omit boundary to be ellipsoidal with characteristic
or ell	major and minor semi-axis lengths (in mm), along with the
	clockwise angle between a horizontal line and the major semi-axis.
	Refer to the diagram under the keyword <b>SPOT</b> .

---

**OMIT** radius *value*

**OMIT** elliptical *valuemajor valuemminor valueangle*

no area omitted

**OMIT** rad 0.35

**OMIT** ell 0.70 0.60 0.0

---

**OSCILLATION**

This keyword tells the program that the oscillation method is being used.

---

start	Starting oscillation angle
-------	----------------------------

end	Ending oscillation angle
-----	--------------------------

step	Angle between 1st oscillation start and 2nd oscillation start
------	---

range	The actual oscillation angle
-------	------------------------------

---

**OSCILLATION**      modifier1      value1\_(degrees)      modifier2  
value2\_(degrees) ...

end = start + range

---

step = range

**OSCILLATION** range 1.0 start 30.0

**OSCILLATION** start 30.0 end 31.0

**OSCILLATION** range 1.0 step 0.8 start 30.0

- ➡ the first two examples describe the same thing. The third describes a series of 1 degree oscillations with ranges which overlap by 0.2 degrees.

## OVERLAP

Overlap defines which pixels of any spot are checked for overlap with the pixels of a neighboring reflection box. The choices are the **spot** pixels, the **guard** pixels (found between the spot circle and the background circle) or **none**. For example, if you specify **OVERLAP spot** and if the spot pixels overlap with the spot pixels of another reflection, then the two reflections are flagged as overlaps.

---

spot

guard

none

---

**OVERLAP** modifier

none

---

**OVERLAP** spot

---

## OVERLOAD VALUE

The lowest optical density reading considered unreliable due to saturation of the film or image plate. The default value corresponds to the full dynamic range being useful with only the highest value being considered overflow. A value that is higher than the maximum possible in the data tells the program to ignore the intensity overflow. The range of the values themselves are determined by the analog to digital converter used to write the data. For 8-bit ADC's (found on film scanners) the highest value is  $2^8-1 = 255$ . For the 16-bit ADC's found on most on-line detectors the range is from 1 to  $2^{16}-1 = 65535$ . The particular **OVERLOAD VALUE** that is considered unreliable has been determined for several detectors and is input as the default when you specify the detector using the **FORMAT** keyword. applies the overload table in Mar scanners so that the overload value should represent the highest reliable entry in the Mar overload table.

---

**OVERLOAD VALUE** *integer* (units are function of detector type)

separate values for different IP formats, contained in **FORMAT** defaults

**OVERLOAD VALUE** 255 (film)

---

- ➡ One-of-a-kind detectors as well as the very old Mar scanners may have the wrong default value.
- ➡ The overload value is a data representation value before applying the data format dependent correction for strong intensities.
- ➡ Default values assume the full dynamic range of the instruments (Mac Science, Mar, R-Axis).



**PEAK SEARCH FILE *peaks.file***

Tells the program to perform **autoindexing**. Get slightly confused?

**PLATE ROTATION (REDUNDANT)**

This is exactly the same as the keyword **FILM rotation**. I recommend you stick with **FILM rotation**, to avoid getting confused with too many aliases.

**POLARIZATION (REDUNDANT)**

Same as **MONOCHROMATOR**.

**PRINT/ END PRINT**

The **PRINT** command tells the program to print the output specified by subsequent keywords to the log file. It does not affect the results of the refinement. Default values of the print sub options can be of several types. Some options are printed by default at every cycle, some are printed on the last cycle of every film of the refinement, and some are never used unless specified. See specifically the various print descriptions below.

---

chi square	print the $\chi^2$ statistics, including the number of spots fitted, the $\chi^2$ of x and y (a measure of how well the spots were fitted based on what sort of error you expected - see also <b>ERROR</b> ), and the predicted decrease in the $\chi^2$ 's after the current cycle of refinement. For those of you who just want numbers, $\chi^2$ 's <b>should be close to 1.0</b> (between 0.5 and 2.0), and the predicted decrease should be near zero, showing that the refinement has converged.
fitted	Print to the log file the observed and fitted profiles of those spots having the specified value. No density readings in the background will be printed. By default, no fitted spots printed.
hkl	Print to the log file the density readings and other statistics of those spots having the specified value. By default, no s printed.
memory	Print statistics of the usage of variables.
profiles	The <b>profiles</b> modifier tells the program to print the average spot profile in a specified number of regions to the log file.  It is best to refrain from printing profiles until the very last cycle of refinement or else your log file will become truly gargantuan. You can also specify <b>PRINT no profiles</b> at the beginning of your refinement, just to make sure it is <b>off</b> until you want to turn it <b>on</b> .
shifts	<b>PRINT shifts</b> tells the program to print the value, the change from the previous value, and the estimated error of all the keyworded parameters fitted on a current cycle of refinement. <b>PRINT no shifts</b> turns this off.
solver	This little-used command prints out correlation matrices calculated in the refinement.
statistics	Print the statistics, including the number of whole and partially recorded reflections matched with preds.

**zones** Print out a list of equivalent zones, given as alternative values for **CRYSTAL** **rotx** **roty** and **rotz**.  
One advantage to invoking this option is to use **rotx**, **roty** and **rotz** values that are more convenient to remember or type. For example, instead of using crystal rot values of, say, -213.6, 137.4, -13.27, you could use the equivalent values of, say, (I'm just making these up) 18.2, -15.4, 63.9. If you wanted to orient your crystal for a particular data collection scheme, these alternative zones could make life easier. **PRINT zones** lists all of these equivalent values.

---

**PRINT** **chi square** /no **chi square**  
**PRINT** **fitted** *integer integer integer*  
**PRINT** **hkl** *integer integer integer*  
**PRINT** **memory** /no **memory**  
**PRINT** **profiles** *integer integer* (the first integer specifies the number of equal sized horizontal regions to divide the frame into, the second specifies the number of equal sized vertical regions. The total number of regions is the product of the first and second integers) /  
**PRINT** **no profiles**  
**PRINT** **shifts** /no **shifts**  
**PRINT** **solver** /no **solver**  
**PRINT** **statistics** /no **statistics**

---

**chi squares** printed after every cycle of refinement  
**no fitted spots** printed  
**no hkl**s printed  
**print** memory statistics on the last cycle  
**no profiles** printed except on the last cycle, when a 3 x 3 grid is printed  
**all shifts** printed  
**no solver**  
**statistics** turned on on the last cycle only  
**PRINT** **fitted** 12 8 0  
  
**PRINT** **profiles** 3 3 (tells the program to divide the frame into 9 equal regions and print out 9 average spot profiles).  
\* the more profiles you specify, the larger the log file will be  
  
**PRINT** **hkl** 12 8 0

---



➡ You can easily generate gigabytes of log file if you use some of the print options, so be careful!

## PROFILE FITTING RADIUS

The radius of the area around each particular spot, in mm, containing neighboring spots, used to calculate the average spot profile. The spot in question is fitted to the average profile of all the spots within the specified radius. Generally, **the radius is set so that spots on roughly 3-5% of the area of the detector are included in the averaging**. Only on the last cycle of the refinement is the profile fitting applied, via the keyword **CALCULATE**.

The particular size of the area chosen depends on how much the profiles vary across the film and the density of the spots on the film. For example, if the spot profiles vary a lot across the film, then you would choose a smaller radius. If you have a small lattice and the spots are widely separated, you might want to choose a larger radius.

➔ The calculation of the average profile is a time consuming task, proportional to the number of spots in the profile fitting radius circle, but too small a radius will not capture enough spots and lead to a noisy average profile. Too large a radius will average out significant profile variations.

To see the **PROFILE FITTING RADIUS** in *ImageWindow* of . A thin white circle corresponding to the profile fitting radius will appear and will move around with the cursor; only those spots whose intensity is above the weak level will have predictions displayed in this mode and only these reflections are used in the calculation of the average spot profile (to return to the normal mode, click  again).

A good rule of thumb is that the **PROFILE FITTING RADIUS** should enclose about 10 to 50 spots. Useful increments of the profile fitting radius are 2.5% of the detector size (e.g. 5 mm for a 200 mm wide IP).

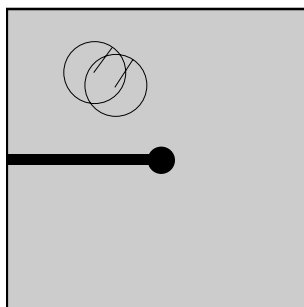
---

**PROFILE FITTING RADIUS** *value\_(mm)*

20.0

**PROFILE FITTING RADIUS** 30.0

---



The circles in the upper left quadrant are two 20 mm profile fitting radii centered on adjacent spots. Each of the hundreds of spots on the film would have its own radius on a real image.

## QUIT

Exit from program . Same as **STOP** or a semicolon (;).

## RADIAL OFFSET

This is one part of the alignment corrections needed for the spiral MacScience and MAR scanners. It is the radial departure, in mm, of the scanning heads from perfect concentricity with the imaging plate. A perfectly aligned scanner would have a **RADIAL OFFSET** of 0.0. The radial offset is perpendicular to the angular offset. A diagram of these offsets is shown under the **ANGULAR OFFSET** keyword.

---

**RADIAL OFFSET** *value\_(mm)*

0.0

**RADIAL OFFSET** -0.43

---

## RASTER

Pixel size in the horizontal (slow scanning) dimension

---

real number in mm

defined by the values in the **FORMAT** statement

**RASTER** 0.1

---

## RAW DATA FILE

This defines the detector data file name template. If the file name contains a run of # characters, then they will be substituted with the numerical value of the **SECTOR** argument. This allows for easy processing of a large series of sequential images.

---

**RAW DATA FILE** *file\_name*

none

**RAW DATA FILE** 'bigdisk:[dir.frames]xtal001.osc'

**RAW DATA FILE** /vms/data/dir/xtal###.osc



in the first example the single quotes are necessary, or else the directory names in square brackets will be ignored (commented out).



You must not repeat this command before the next **GO** statement.

---

## RECORD LENGTH

Defines the number of pixels in the data record. It is equal to the number of horizontal (slow scanning direction) pixels.

---

**RECORD LENGTH** *value\_(pixels)*

**FILEM** length / (**RASTER** / **Y SCALE**) (this is defined by the **FORMAT** statement)

**RECORD LENGTH** 2000

---

## REFINE PARTIALITY / REFINE NO PARTIALITY (DO NOT USE!)

Tells the program to determine the fraction partiality of the reflections on the image. It is critical that this be on during the refinement of the crystal rotation parameters, since **CRYSTAL** **rotx** and **roty** cannot be determined without knowing the partiality of the reflections on the image.

---

**REFINE PARTIALITY**

**REFINE PARTIALITY**

---

➡ **Do not use REFIN NO PARTIALITY !** (see theoretical explanation in chapter).

## REJECT

Describes a condition whereby a reflection is rejected by `REJECT`. This is especially useful for eliminating spurious measurements due to ice rings.

<code>fraction</code>	The fraction of <b>acceptable</b> pixels in the background, which are needed for the reflection to be accepted. Default = 0.75 (i.e. 75%). <b>Very sensitive to small changes</b> , so increment by 1-2% if you are messing with this. To get rid of spots in an ice ring, increase this value by 1-2% until the preds in the ice ring are colored red. After all, the defining feature of an ice ring is a horrible background. The acceptability of a background pixel is determined by the next three modifier words.
<code>cutoff</code>	The <b><math>\sigma</math> deviation</b> from the best fit to the background plane at which a pixel has a 50% chance of being incorrect and thus rejected. Default is $3\sigma$ .  The background of a spot can be described by a plane and should be uniform and flat. Sometimes it is not; for example, it may have a slope to it, which can be calculated. If a pixel is still out of the best-fit background plane by (default) $3\sigma$ , then it is flagged as a reject. It is best not to mess with this unless you really know what you are doing.
<code>slope</code>	The slope of the background plane above which the entire background and spot are considered unacceptable and rejected. Default is 50 units. Flat = 0, vertical = infinite. It is sometimes useful to increase this to 100 for very low background images.
<code>systematic</code>	This is a way of changing the background rejection criteria when you have a major problem with <b>diffuse scattering</b> . Default is 0.01. The range is from 0 to 1, with 0 meaning rely totally on the cutoff criteria, and 1 meaning don't reject anything. As you can see from the default, you normally use the $3\sigma$ cutoff criterion.

---

```
REJECT fraction value
REJECT fraction 0.75
REJECT cutoff 3.0
REJECT slope 50
REJECT systematic 0.01
REJECT fraction 0.80
```

---

## RESOLUTION LIMITS / RES

Defines the resolution range used in the program, in Ångströms.

---

```
RESOLUTION LIMITS valuelower valueupper_(
none
RESOLUTION LIMITS 20.0 2.5
```

---

---

**RES** 20.0 2.5

---

## SECTOR

This is the **counter and range delimiter** for a series of frames when processing in batch mode. The number in the sector argument is substituted for the ### (or #, or ##, ...) variable in the **RAW DATA FILE** and **FILM output file** commands. Make sure there are enough # marks to accommodate all of the digits in the sector argument, e.g. if you have sector values in the 100's then there must be at least three # marks in the file names specified in **RAW DATA FILE** and **FILM output file**. The sector value can start at any non-negative integer, and is incremented by 1 every time an **END OF PACK** statement is reached. The first number in the sector argument is the first frame and the last number is the last frame. See also **START REFINEMENT**.

---

**SECTOR** *integer to integer*

none

**SECTOR** 1 to 50

---

## SKEW

**SKEW** refers to the **non-orthogonality** of the vertical and horizontal scanning directions, measured in radians. For a perfectly orthogonal system **SKEW** is defined to be zero. Generally **SKEW** values will be very close to this number. It can be fitted in the later rounds of parameter refinement.

---

**SKEW** *value\_(dimensionless)*

0.0000

**SKEW** 0.00169 (value taken from the earlier refinement)

---



Spiral scanners and CCD detectors should use the default value of zero, which should not be refined.

## SPACE GROUP

The space group symbol or number as given in the **International Tables**. Note that for rhombohedral space groups autoindexing only works for the primitive choice, i.e. H3 or H32, not R3 or R32. The **SPACE GROUP** must be declared before the **FIT cell** command can work.



does not actually care about the **SPACE GROUP** per se. It indexes films by Bravais lattice and does not take into account systematic absences arising from two-folds, screws, etc. These are accounted for in Scalepack, when the data are reduced. So for the purposes of , the resulting .x file will be the same whether you declare the space group to be P222 or P212121, or, P3 and P6122.

---

**SPACE GROUP** *modifier*

P1

**SPACE GROUP** P212121

---

**SPINDLE AXIS / ORIENTATION AXIS 1**

The reciprocal lattice vector parallel to the spindle axis. Same as keyword **ORIENTATION AXIS 1**

3 integers representing the reciprocal vector

0 0 1 (c\* axis)

**SPINDLE AXIS 1 0 0** (a\* axis)

**SPINDLE AXIS 1 0 1** (a non-major axis)

**SPOT**

Defines the shape and area of the region in the box where the spot is to be found.

**radius** Defines the spot to be circular with a radius value, in mm. The center of the spot area and of the box coincide

**elliptical** Defines the spot to be ellipsoidal with characteristic major and minor semi-axis lengths (in mm), along with the clockwise angle between a horizontal line and the major semi-axis.

**SPOT radius value\_(mm)**

**SPOT elliptical valuemajor valuemajor\_(mm) valueangle\_(degrees)**

none

**SPOT radius 0.35**

**SPOT elliptical 0.70 0.60 0.0**

➡ See diagram (fig. 4) in section

**START REFINEMENT**

**START REFINEMENT** denotes the start of the series of commands given to index the diffraction image. The end of the **START REFINEMENT** loop is denoted by the **CALCULATE GO** statement. The **START REFINEMENT** statement must come after the **SECTOR x to y** statement, and the loop is run for every sector value.

**START REFINEMENT**

this is not a default

➡ see the example command file in section

**STOP (REDUNDANT)**

This command stops the job. Same as **QUIT** or a semicolon (;).

**SWAP BYTES (OBSOLETE)**

All programs now automatically recognize swapped data.

**TITLE**

Data set title

---

string. If it contains any spaces, commas, equal signs, or brackets ( - recognized separators), it must be enclosed in single quotes.

none

**TITLE** 'my crystal'

---

**TOP MARGIN**

The top margin, in mm from the top edge, of useful data, i.e. the beginning of every record. Do not set the top margin to be larger than the film length. Note that commands like top margin are working in the data coordinate system. It is very important to check the results of these commands in the display window.

---

**TOP MARGIN** *value\_(mm)*

0.0

**TOP MARGIN** 2.0

---

**UNIT CELL**

Unit cell parameters in either real or reciprocal space.

---

**UNIT CELL** *a\* b\* c\*  $\alpha^*$   $\beta^*$   $\gamma^*$*  or **UNIT CELL** *a b c  $\alpha$   $\beta$   $\gamma$* . Lengths in Ångstroms or reciprocal Å, angles in degrees.

none

**UNIT CELL** 120.92 120.92 131.65 90 90 90

---



If the space group has been declared, the program is smart enough to know which cell lengths should be equivalent, and which angles are variable, and it will hold those lengths equal or fixed throughout the refinement.

**USE BEAM**

Tells the program to use the beam spot in the refinement. This keyword has the opposite effect of the keyword **NO BEAM**. This is useful in the initial steps of the refinement, when not too many spots are matched by preds.

---

*only* Tells the program to use only the beam spot in the refinement. Has the

---



---

effect of turning off the **RESOLUTION LIMITS** keyword.

**USE BEAM**

**USE BEAM** only

not the default option. See **NO BEAM**.

---

## **USE FIDUCIALS / USE NO FIDUCIALS**

Tells the program to use the **fiducials** in the refinement. Turned off with **USE NO FIDUCIALS**. Fiducial positions are entered with the keyword **FIDUCIALS**.

---

only      Tells the program to use only the fiducial positions in the refinement.

**USE FIDUCIALS**

**USE FIDUCIALS** only

not the default. Default is **USE NO FIDUCIALS**.

---

## **USE PARTIALS POSITION / USE PARTIALS NO POSITION**

Normally, data reduction programs use only fully recorded reflections in the refinement of the crystal and detector positional parameters. The keyword **USE PARTIALS POSITION** tells the program to use a reflection that is predicted to be only partially recorded in the positional refinement. A spot is predicted to be partially recorded based on the values entered for the oscillation range, as well as the crystal mosaicity and beam crossfire. This is a very useful feature of **REFMAC5**, because it allows one to refine the orientation of the crystal even if none of the reflections are fully recorded. This occurs if one collects either very thin oscillation sectors (in an attempt to minimize background) or has a crystal of relatively high mosaicity, or some combination of both. **USE PARTIALS NO POSITION** is mostly of historical interest.

---

**USE PARTIALS POSITION**

**USE PARTIALS POSITION**

**USE PARTIALS NO POSITION**

---

## **VERTICAL AXIS / ORIENTATION AXIS 2**

The reciprocal lattice vector perpendicular to both the spindle axis and the X-ray beam. Equivalent to **ORIENTATION AXIS 2**.

---

3 integers representing the reciprocal vector

1 0 0 (a\* axis)

**VERTICAL AXIS** 0 1 0 (b\* axis)

---

## **WAVELENGTH**

X-ray wavelength. Same as **LAMBDA**.

---

**WAVELENGTH** value\_ (Ångstroms)

1.5418 (CuK $\alpha$ )

**WAVELENGTH** 1.5418

---

**WEAK LEVEL**

This is the intensity rejection level, expressed as an  $I/\sigma$  ratio. Only strong spots with an integrated intensity (i.e.  $I$ ) above the **WEAK LEVEL** times the intensity error measurement (i.e.  $\sigma$ ) are used in the positional refinement and as reference profiles in profile fitting.



The intensities of all reflections determined to be  are still measured and included in the .x file. **They're just not used in the profile fitting.** You can use **Prof fit R** in the  window to see which reflections are above the weak level.

---

**WEAK LEVEL** *value\_(dimensionless)*

5.0

**WEAK LEVEL** 5.00

---

**WRITE PREDICTIONS / WRITE NO PREDICTIONS**

This tells  to write a file called hklpredictions. This file contains the positions of the predicted reflections, and is read by  to update the predictions in the display. Once the flag is turned on, the file is updated every time a new cycle of refinement is executed. This way you can follow the course of the refinement in the display program. Turned off with **WRITE NO PREDICTIONS**.

---

**WRITE PREDICTIONS**

**WRITE NO PREDICTIONS**

**WRITE PREDICTIONS**

---

**X BEAM**

Defines the distance from the edge of the data to the beam spot, in mm. You can use coordinates of the cursor in the green box of the  window to find the beam position.

---

**X BEAM** *value\_(mm)*

**FILM** width/2

**X BEAM** 98.961

---

**Y BEAM**

Defines the distance from the edge of the data to the beam spot, in mm. Again, use coordinates of the cursor in the green box of the  window to find the beam position.

---

**Y BEAM** *value\_(mm)*

**FILM** length/2

**Y BEAM** 100.146

---

## Y SCALE

**Y SCALE** in `SCALE` serves two purposes. First, it is a dimensionless correction factor for anisotropy in the pixel dimensions. It is applied to the fast scanning, or `FAST`, dimension of the pixel. If the pixel is larger in `FAST` than in `SLOW`, then **Y SCALE** is smaller than 1.

- ➡ It is possible to create a data file which has the inverted direction of scanning along one of the axis - similarly as in case of film scanned from back side. `SCALE` handles this aberration by changing the sign of the **Y SCALE** keyword to negative. If this is not done, then the assignment of the Friedel mates (Bijvoet differences) will be inverted, and thus the hand of the structure will be similarly inverted - see *Determining the values of FILM ROTATION and Y SCALE* section in '*Denzo Orientation Conventions*' chapter.
- ➡ The sign of **Y SCALE** is a detector property, while the absolute value of **Y SCALE** seems stable for all detectors. Once the correct sign has been determined, the calibration of **Y SCALE** is done by refining with a high symmetry crystal (trigonal, tetragonal, hexagonal, or cubic).
- ➡ For CCD and spiral scanners the value of **Y SCALE** should be exactly  $\pm 1$  and not refined !

---

**Y SCALE** *value\_(dimensionless)*

detector dependent, defined by the **FORMAT** statement

**Y SCALE** 1.03152

---

## YORK OUTPUT FILE

Substitutes for the **OUTPUT FILE** command, and tells the program to write the `YORK` file in the **Eleanor Dodson** York file format. Note that Scalepack reads the `YORK OUTPUT FILE` as format `denzo_york1`.

# Denzo Appendixes

## Keywords by Category

---

### 1. Film Pack Parameters: Defined at the beginning of each job

---

TITLE	FILM. . .
UNIT CELL	CRYSTAL
LATTICE	CASSETTE
OSCILLATION	RADIUS
MOSAICITY	DISTANCE
WAVELENGTH/LAMBDA	Y SCALE
SPINDLE AXIS	AIR ABSORPTION LENGTH
VERTICAL AXIS	FIDUCIALS
MONOCHROMATOR	FORMAT
RASTER	

### 2. Film Parameters: Defined at the beginning of every frame

---

FORMAT or	RAW DATA FILE
FILM. . .	SECTOR
RECORD LENGTH	X BEAM
NUMBER OF RECORDS	Y BEAM
LEFT MARGIN, TOP MARGIN	

### 3. Refinement Parameters: Defined at the beginning of every iteration step

---

BOX	CALCULATE
SPOT	PRINT
BACKGROUND	ERROR
OMIT	WEAK LEVEL
RESOLUTION LIMITS	GO
USE BEAM/NO BEAM	END OF PACK
OVERLOAD VALUE	END OF JOB
PROFILE FITTING RADIUS	INPUT/@
FIT/FIX	

## Default Parameters

**Format - independent keywords** with predefined default values:

AIR ABSORPTION LENGTH	860	LATTICE	P
ANGULAR OFFSET	0.0	MONOCHROMATOR	graphite
BACKGROUND	lowest reading in spot area	NO BEAM	
CASSETTE rotx roty	0.0	OVERLAP	none
COUPLING	0.0	PLATE ROTATION	0.0
CROSSFIRE x y	0.0 0.0	PRINT chi square	
CRYSTAL rotz roty rotx	0.0 0.0 0.0	PRINT no profiles	
DISTANCE	75.0	PRINT no solver	
EIGENVALUE FILTER	0.001	PRINT shifts	
ERROR		PRINT statistics	
overload	0.005	PROFILE FITTING RADIUS	20.0
positional	0.012	RADIAL OFFSET	0.0
y/x factor	0.1	RADIUS	flat
partiality	0.1	SPINDLE AXIS	0 0 1
systematic	5.0	VERTICAL AXIS	1 0 0
FILM rotation	0.0	WAVELENGTH	1.5418
FIX	all	WEAK LEVEL	5.0
I/O BUFFER	full file size	X BEAM	FILM width/2
LAMBDA	1.5418	Y BEAM	FILM length/2

**Format - Specific Keywords.** Theirs default values can be found in section:

ERROR density	FILM width	RASTER
FILM absorption	NUMBER OF RECORDS	RECORD LENGHT
FILM length	OVERLOAD VALUE	Y SCALE

**Keywords with no defaults:**

► without which the program absolutely will not run:

BOX	OSCILLATION or PRECESSION	RESOLUTION LIMITS
FORMAT	OVERLAP	SPOT
MOSAICITY	RAW DATA FILE	UNIT CELL

► required for **interactive processing**: FIT . . . GO

► required for **batch processing**:

FIT . . . GO	END OF PACK	SECTOR
CALCULATE	FILM OUTPUT FILE	TITLE
END OF JOB	INPUT/@	

► which may or may not be required, depending on the circumstances:

FIDUCIALS	PRINT hkl	READ UMAT
IGNORE	PRINT profiles	UNIX I/O
OMIT	PRINT solver	LONGEST VECTOR
PRINT fitted	READ AMATRIX	

## Recommended Hardware

The operative philosophy behind the programs , , and is that computer hardware is cheap compared to all of the other expenses of crystallography. Modern platforms are supported; old ones are gradually rendered obsolete by their speed, discontinuation, or by other hardware requirements, like memory. Here's a list of current hardware recommendations, but note that this is not meant to be a complete list of computers on which the program runs. It's more to give you an idea of the direction your crystallographic computing should be heading.

CPU	DEC Alpha, SGI, Pentium III,IV
Operating System	OSF, UNIX, Linux
Memory	512 MB or more
Display	17-21", 1280x1024 or more; 16 bit or more; workstation, not X-terminal
Disk	<ul style="list-style-type: none"><li>➤ much more than one data set, local to workstation</li><li>➤ preferably more than 20GB</li></ul>
Swap Space	>1.5 Gb (very important!)



You should never economize on swap space because otherwise the programs may get killed randomly by the operating system.



OSF-Digital Unix and IBM AIX set default datasize to unrealistically low values. Datasize should be set to unlimited, when the HKL programs are run.

## Manual Indexing of Diffraction Images

Manually indexing and orienting diffraction patterns is a tricky business, more so with small unit cells, centered space groups, or lattices containing pseudosymmetry, and less so with high resolution data and simple lattices. Suffice it to say that the more you know about your crystal and detector parameters the easier the job will be. For the purposes of this section, I will assume that you know the following:

- detector type (duh!)
- wavelength
- crystal to film distance within 1% or so
- x beam and y beam within a pixel or two each (take a quickie; direct beam shot if unsure!)
- detector  $2\theta$  angle within 1 degree
- film rotation exactly (depends on detector and  $\chi$  value of spindle)
- crystal unit cell lengths within a couple of Ångströms, cell angles within a few degrees
- space group/lattice type
- oscillation range
- effective mosaic spread of the reflections. If in doubt, enter in a higher number. A good choice is 0.1 to 0.3 degrees for capillary mounts, and 0.5 to 1.0 degrees for frozen crystals.
- Furthermore, I will assume that the non- $2\theta$  dependent cassette rotation parameters are near zero.

All of the non-trivial parameters listed above can be refined if they are close enough and the crystal orientation is figured out to a good first approximation. You will know you have successfully indexed your crystal when the predicted reflections (“preds”) line up on the observed diffraction spots.

- ➡ It is indeed possible to manually index crystals whose lattice type and unit cell parameters are not known. However, this significantly increases the number of variables you will have to worry about.

The indexing proceeds in two steps. The first is to determine the spindle and vertical axes. Then the crystal rotation angles are determined.

### Determining the spindle axis and vertical axis

The **SPI NDLE AXIS** is the unit cell vector which is parallel to the axis of rotation of the camera spindle. It is given in  $h k l$  notation, so that non-major axes can also be specified. For example, the axis 0 0 1 means that the  $c^*$  axis is parallel to the spindle axis. The **VERTI CAL AXIS** is the unit cell vector which is perpendicular to both the **SPI NDLE AXIS** and the beam axis. In practical terms, this is simply the axis running in the vertical direction on the screen.

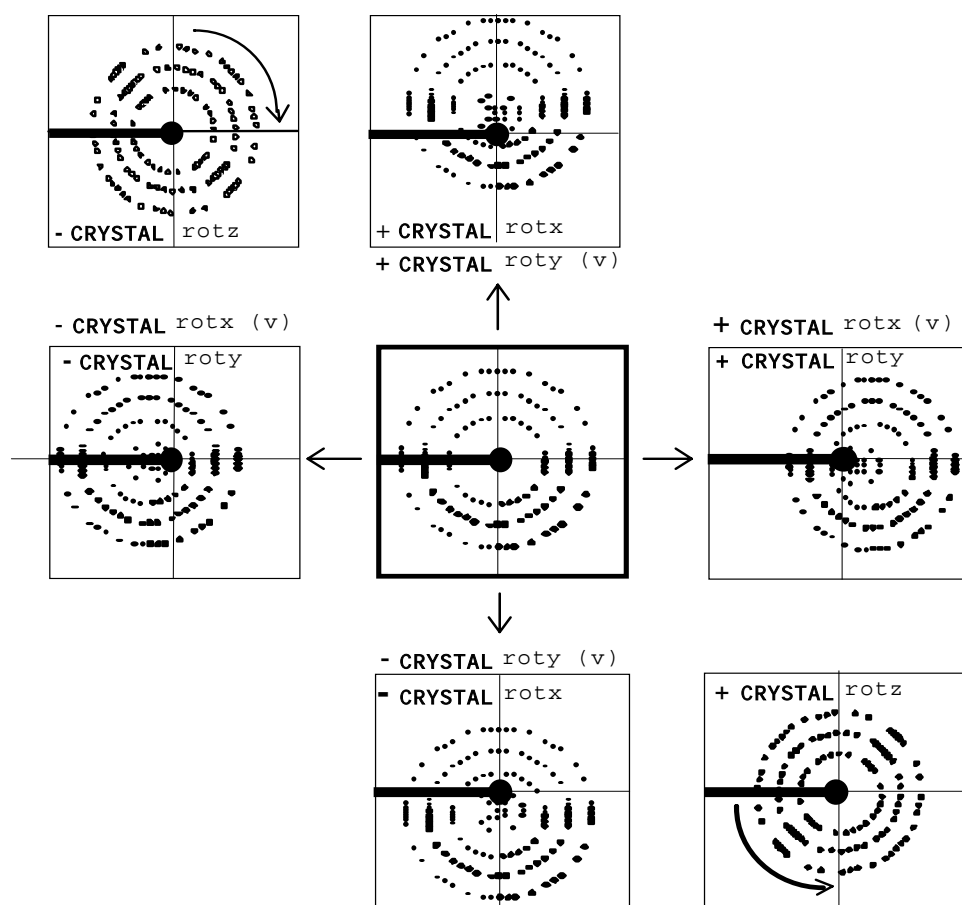
The way to determine each axis is to look at your diffraction pattern for two lines of spots running perpendicular to each other. Call the one more parallel to the spindle axis the **SPI NDLE AXIS**, and measure the spacing between the spots using the mouse pointer. If you are lucky, you will have measured spacing corresponding to a major axis, for example the  $c^*$  axis, and you can enter the command **SPI NDLE AXIS 0 0 1**. The same holds true for the **VERTI CAL AXIS**.

But let's say you are unlucky, or you have a centered space group and you can't see a major zone. Here's what you can do:

1. Page through your frames until you find one that does have a recognizable major zone. Depending on your oscillation range, this could mean looking at every 5<sup>th</sup> or 10<sup>th</sup> frame, which is not such a bad investment of time.
2. If no major zone is apparent, you will have to figure out which pseudo zone you are looking at. For example, let's say you have a tetragonal, body centered unit cell with  $a=b=120\text{\AA}$ ,  $c=130\text{\AA}$ . You measure a spacing of about  $90\text{\AA}$ . This is clearly the  $1\ 0\ 1$  or  $0\ 1\ 1$  axis. There is another axis  $90$  degrees away from this one, with spacing  $62\text{\AA}$ . This must be the  $0\ 2\ 0$  or  $2\ 0\ 0$  axis, and can be entered for the vertical axis. Once the axes have been determined, enter them using the commands `SPINDLE AXIS 1 0 1 VERTICAL AXIS 0 2 0 GO`.

### Determining crystal orientation angles: `CRYSTAL rotz roty rotx`

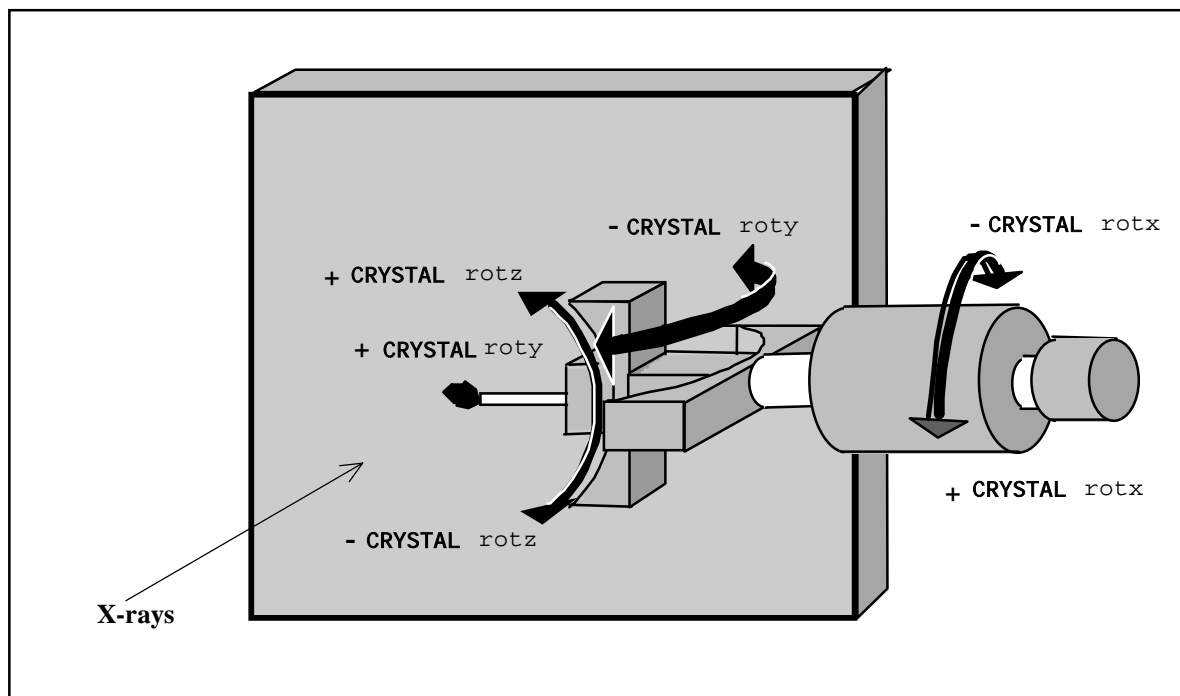
The convention used in is as shown in the pictures on the next two pages.



**Figure 2:** Denzo rotation conventions. A simulated array of predicted reflections is shown in the center box, along with the operations that shift the position of the preds to those seen in the other boxes. The operation used to move the preds away from the pattern given in the center display (heavy lines) is given above each box. The spindle axis is assumed here to be horizontal,  $180$  degrees away from the beam stop axis (which is indicated by its shadow). In the case of vertical spindles, the subsequent effects of crystal rotations are given in parentheses, and denoted with a 'V'.



The rotations **CRYSTAL** rotz, **CRYSTAL** roty, and **CRYSTAL** rotx are best thought of as rotations, expressed in degrees, applied to the predicted reflections (preds) from the initial conditions of **CRYSTAL** rotz 0.0 roty 0.0 rotx 0.0. Values are entered as absolute rotations from zero. For example, if you have **CRYSTAL** rotz 23.7, and then you subsequently enter **CRYSTAL** rotz -15.2, the applied value of **CRYSTAL** rotz is -15.2, *not* ( $23.7 - 15.2 = 8.5$ ). Another way to picture these rotations is as the rotations of the arcs on a goniometer head mounted on the spindle of a camera:



**Figure 3:** Another way of thinking about Denzo rotation conventions: arcs on a goniometer  
*Tip: to interpret this diagram for non-horizontal spindles, simply align the axis of your spindle with that of the spindle in the diagram. All the rotations line up then as well.*

It is easiest to change the crystal rotations one at a time in the following order: **CRYSTAL** rotx (spindle), **CRYSTAL** roty and then **CRYSTAL** rotz. To enter a new value, type it in the command window as: **CRYSTAL** rotx -22.0 **GO**. You can change more than one value at a time, for example: **CRYSTAL** rotx -20 rotz 5.0 roty 5.5 **GO**. In the manual part of the indexing, it is best to change the rotations in increments of no **less** than 0.5 degrees, since finer adjustments can be done automatically by the program. can recapitulate the current parameters with the command **LIST**.

The best advice is to think clearly, keep track of your movements and their results, and be patient. If you do manage to get a reasonable agreement with the diffraction pattern, then proceed with refinement.

## ***Film Cassette and Off-line Image Plate Users***

The success of indexing with `SCALE` (or any program, for that matter) depends critically on knowing or finding the position of the direct X-ray beam for each frame. With automatic image plate scanners like the MacScience, R-axis, or Mar, this is generally not a problem because the detector and the beam do not move much during data collection, and the scanning mechanism is accurate enough to scan the IP in a reproducible fashion. Thus, once the position of the direct beam is established for one frame, it is pretty much the same for all the frames of the series. Cassette-based systems like the Fuji scanners found at several synchrotrons (CHESS F1, Photon Factory, Brookhaven X25, X4, etc.) do not have this reproducibility, however. This is due to the slight slop inherent in the positioning of the IP in the cassette, the attachment of the cassette onto the camera carousel, and the positioning of the IP in the scanner magazine. Thus, the `x BEAM` and `y BEAM` positions on each frame can vary beyond the normal radius of convergence for refinement in `SCALE`. One way to overcome this is simply to examine each frame one by one and note down the position of the direct beam in each case, and enter this value into the command file for batch processing.

Another way is to index the first frame and refine each subsequent frame interactively, one by one after determining the direct beam position, again, by inspection of each image. With a fast computer this is a viable, albeit tedious, option. Note however that you should *not* keep re-running the autoindexing routine. This is because the autoindexing routine will not use the same values for crystal `rotx`, `roty`, and `rotz` each time it is run, but instead will randomly choose between equivalent values. This does not matter the first time these values are chosen, but if each image is indexed with different values of `CRYSTAL rotx`, `roty`, `rotz`, then the original index of all the reflections will be inconsistent. This means that you will not be able to add partially recorded reflections correctly, or even scale the data, which is a major drag.

A third way is to incorporate a small "beam search" routine into the `refine.dat` macro file above. An example of such a search routine is shown below. Essentially, what this routine does is create a large zone within which the direct beam may be found, and then exclusively search for this single "reflection." Note that you must actually have a spot on the frame to find; if the beam stop does not have a thinned center through which the attenuated direct beam can pass, then this search routine is pointless and will not find what is not there to be found. Also, it helps if the starting values of `x BEAM` and `y BEAM` are reasonably good average values. If this search is going to fail, it is usually because the starting point for the direct beam search was too far away from the actual point. It has also been known to fail if there are other small intense noise spots within the beam spot search zone. Once the beam is found, the box and spot sizes are reset to normal values, and the refinement proceeds normally.

## Detector Specifications

Detector Description	Modifier: XdisplayF Ver 1.5 and lower	Modifier: XdisplayF Ver 1.96 and higher	Denzo Format Modifier
fuji 10 bit scanner 100 micron raster	INST_FUJI	fuji 10bit	fuji 10bit
Film scanner (Purdue type) 25 micron scan	INST_FILM25	film 25	film 25
Film scanner (Purdue type) 50 micron scan	INST_FILM50	film 50	film 50
Film scanner (Purdue type) 100 micron scan	INST_FILM100	film 100	film 100
Film scanner (Uppsala type) 50 micron scan		film lars	film lars
Film scanner (Weizmann type) 50 micron scan		film is	film is
CCD Bruker quad detector - IMCA beam line		ccd bruker quad	ccd bruker quad
CCD Bruker quad detector (binned mode) - IMCA beam line		ccd bruker quad binned	ccd bruker quad binned
CCD SMART-6000 Bruker detector		ccd smart6000	ccd smart6000
CCD Bruker single detector (older)		ccd bruker	ccd bruker
CCD SMART-6000 Bruker detector in binned mode		ccd smart6000 binned	ccd smart6000 binned
CCD APS-1 detector at SBC		ccd aps1	ccd aps1
CCD APS-1 detector at SBC in binned mode		ccd aps1 binned	ccd aps1 binned
CCD SBC-2 detector at SBC		ccd sbc2	ccd sbc2
CCD SBC-2 detector at SBC in binned mode		ccd sbc2 binned	ccd sbc2 binned
Nonius KappaCCD detector		ccd nonius	ccd nonius
Nonius KappaCCD detector dezingering-on		ccd nonius dz	ccd nonius dz
Nonius KappaCCD detector binned		ccd nonius binned	ccd nonius binned
Nonius KappaCCD detector binned dezingering-on		ccd nonius binned dz	ccd nonius binned dz
Nonius KappaCCD-2000 detector		ccd noniusb	ccd noniusb
Nonius KappaCCD-2000 detector dezingering-on		ccd noniusb dz	ccd noniusb dz
Nonius KappaCCD-2000 detector binned		ccd noniusb binned	ccd noniusb binned
Nonius KappaCCD-2000 detector binned dezingering-on		ccd noniusb binned dz	ccd noniusb binned dz
CCD (Brandeis University) at BNL X12C		ccd brandeis	ccd brandeis
CCD Brandeis 2k CCD single chip detector at X12C		ccd brandeis-2k	ccd brandeis-2k

Detector Description	Modifier: XdisplayF Ver 1.5 and lower	Modifier: XdisplayF Ver 1.96 and higher	Denzo Format Modifier
CCD Brandeis 2x2		ccd unsupported-b4	ccd unsupported-b4
CCD (ESRF) detector BL19 at ESRF-raw data		ccd bl19	ccd bl19
CCD (ESRF) detector BL19 at ESRF>1200x1200 pxl		ccd esrf fit2d transformed	ccd esrf fit2d transformed
CCD (ESRF) detector BL19 at ESRF>1200x1200 pxl		ccd esrf fit2d transformed wrong distance	ccd esrf fit2d transformed wrong distance
CCD (ESRF) detector 1024*1024		ccd esrf 1024	ccd esrf 1024
CCD (ESRF) detector 1152x1242		ccd esrf 1242	ccd esrf 1242
CCD 1k 55mm [original Sol Gruner] detector at CHESS	INST_CCD	ccd chess 55mm	ccd chess 55mm
CCD 2k detector CHESS 2048x2048 image	INST_CCD2x2	ccd chess 80mm	ccd chess 80mm
CCD 2k detector CHESS 1024x1024 (binned mode)	INST_CCD2048b	ccd chess 80mm binned	ccd chess 80mm binned
MAC Science DIP-100	INST_DIP100	dip 100	dip 100
MAC Science DIP-2000 12-bit A/D converter	INST_DIP2000	dip 2020	dip 2020
MAC Science DIP-2000 16-bit A/D converter		dip 2020b	dip 2020b
MAC Science DIP-2030 12-bit A/D converter		dip 2030	dip 2030
MAC Science DIP-2030 16-bit A/D converter		dip 2030b	dip 2030b
MAC Science DIP-2040 16-bit A/D converter		dip 2040b	dip 2040b
MAC Science DIP-3000 cylindrical Weissenberg	INST_DIP3000	dip 3000	dip 3000
MAC Science DIP-320 Weissenberg scanner	INST_DIP320	dip 320	dip 320
MAC Science DIP-320 Weissenberg scanner 16bit		dip 320b	dip 320b
MAC Science DIP-3200 Weissenberg scanner		dip 3200	dip 3200
MAC Science DIP-3200-2 Weissenberg scanner (long)		dip 3200l	dip 3200l
MAC Science DIP-420 off-line scanner		dip 420	dip 420
MAC Science DIP scanner for neutrons 0.2mm pixel		dip nf	dip nf
MAC Science DIP scanner for neutrons 0.4mm pixel		dip nc	dip nc
ADSC Quantum4 CCD detector (raw data)		ccd adsc quantum4	ccd adsc quantum4
ADSC Quantum4 CCD detector (raw data) dezingering-on		ccd adsc quantum4 dz	ccd adsc quantum4 dz
ADSC Quantum4 CCD detector (raw data) binned		ccd adsc quantum4 binned	ccd adsc quantum4 binned
ADSC Quantum4 CCD detector (raw data) binned dezingering-on		ccd adsc quantum4 binned dz	ccd adsc quantum4 binned dz

ADSC Quantum4 CCD detector (corrected data)		ccd adsc unsupported-q4	ccd adsc unsupported-q4
Detector Description	Modifier: XdisplayF Ver 1.5 and lower	Modifier: XdisplayF Ver 1.96 and higher	Denzo Format Modifier
ADSC Quantum4 CCD detector (corrected data) binned		ccd adsc unsupported-q4 binned	ccd adsc unsupported-q4 binned
ADSC Quantum210 CCD detector (corrected data)		ccd adsc unsupported-q210	ccd adsc unsupported-q210
ADSC Quantum210 CCD detector (corrected data) binned		ccd adsc unsupported-q210 binned	ccd adsc unsupported-q210 binned
ADSC Quantum315 CCD detector (corrected data)		ccd adsc unsupported-q315	ccd adsc unsupported-q315
ADSC Quantum315 CCD detector (corrected data) binned		ccd adsc unsupported-q315 binned	ccd adsc unsupported-q315 binned
CCD (ADSC) detector single - CHESS		ccd adsc	ccd adsc
red EMBL scanner in Hamburg	INST_HAMBURG	mar_corr	mar_corr
MAR scanner @ LBL -LMB prototype 1187 pixel format		mar lbl	mar lbl
MAR scanner in MRC-LMB compressed 'pck' mode		mar compressed	mar compressed
MAR scanner in small plate mode	INST_HAMBURGN	mar 18cm	mar 18cm
MAR scanner in large plate mode	INST_HAMBURGD	mar 30cm	mar 30cm
MAR 345mm scanner 180mm scan 100 micron size		mar345 180mm 100	mar345 180mm 100
MAR 345mm scanner 180mm scan 100 micron size compressed		mar345 180mm 100 marpck	mar345 180mm 100 marpck
MAR 345mm scanner 180mm scan 150 micron size		mar345 180mm 150	mar345 180mm 150
MAR 345mm scanner 180mm scan 150 micron size compressed		mar345 180mm 150 marpck	mar345 180mm 150 marpck
MAR 345mm scanner 240mm scan 100 micron size		mar345 240mm 100	mar345 240mm 100
MAR 345mm scanner 240mm scan 100 micron size compressed		mar345 240mm 100 marpck	mar345 240mm 100 marpck
MAR 345mm scanner 240mm scan 150 micron size		mar345 240mm 150	mar345 240mm 150
MAR 345mm scanner 240mm scan 150 micron size compressed		mar345 240mm 150 marpck	mar345 240mm 150 marpck
MAR 345mm scanner 300mm scan 100 micron size		mar345 300mm 100	mar345 300mm 100
MAR 345mm scanner 300mm scan 100 micron size compressed		mar345 300mm 100 marpck	mar345 300mm 100 marpck
MAR 345mm scanner 300mm scan 150 micron size		mar345 300mm 150	mar345 300mm 150
MAR 345mm scanner 300mm scan 150 micron size compressed		mar345 300mm 150 marpck	mar345 300mm 150 marpck
MAR 345mm scanner 345mm scan 100 micron size		mar345 345mm 100	mar345 345mm 100
MAR 345mm scanner 345mm scan 100 micron size compressed		mar345 345mm 100 marpck	mar345 345mm 100 marpck
MAR 345mm scanner 345mm scan 150 micron size		mar345 345mm 150	mar345 345mm 150

MAR 345mm scanner 345mm scan 150 micron size compressed		mar345 345mm 150 marpck	mar345 345mm 150 marpck
MAR CCD detector 165 mm		ccd unsupported-m165	ccd unsupported-m165
Detector Description	Modifier: XdisplayF Ver 1.5 and lower	Modifier: XdisplayF Ver 1.96 and higher	Denzo Format Modifier
MAR CCD detector 130mm		ccd unsupported-mar	ccd unsupported-mar
RAXIS-IIc Rigaku scanner 105 micron raster scan	INST_RX	raxis	raxis
RAXIS-IIc Rigaku scanner 210 micron raster scan	INST_RXS	raxis 210	raxis 210
RAXIS-IIc Rigaku scanner 105 micron scan (Japanese version)		raxis2n	raxis2n
RAXIS-IIc Rigaku scanner 210 micron scan (Japanese version)		raxis2n 210	raxis2n 210
RAXIS-IV Rigaku (prototype) 100 micr. raster scan		raxis4 old	raxis4 old
RAXIS-IV Rigaku scanner 50 micron raster scan		raxis4 50	raxis4 50
RAXIS-IV Rigaku scanner 100 micron raster scan		raxis4 100	raxis4 100
RAXIS-IV Rigaku scanner 200 micron raster scan		raxis4 200	raxis4 200
STOE scanner		stoe	stoe
CCD detector EEV chip 576 pixels		ccd eeV 576	ccd eeV 576
CCD detector EEV chip 576 pixels double image mode		ccd eeV 576 double	ccd eeV 576 double
CCD detector EEV chip 620 pixels		ccd eeV 620	ccd eeV 620
CCD detector EEV chip 621 pixels		ccd eeV 621	ccd eeV 621
CCD detector EEV chip 625 pixels		ccd eeV 625	ccd eeV 625
CCD detector EEV chip 625 pixels binned		ccd eeV 625 binned	ccd eeV 625 binned
CCD detector EEV chip 1152 pixels		ccd eeV	ccd eeV
Photon Factory - Amemiya drum scanner		pf beam line 14a	pf beam line 14a
Photon factory fuji BA-100 scanner NL format		fuji nl	fuji nl
Photon factory fuji BA-100 scanner NL format (Weisenberg exp)		weis nl	weis nl
Photon factory fuji BA-100 scanner SL format (Weisenberg exp.)	INST_WEIS	weis sl	weis sl
Photon Factory beam line 18b whole image		pf beam line 18b	pf beam line 18b
Photon Factory beam line 18b center		pf beam line 18b center	pf beam line 18b
Photon Factory beam line 18b beam region		pf beam line 18b beam	pf beam line 18b
Photon Factory beam line 18b bottom		pf beam line 18b bottom	pf beam line 18b

Photon Factory beam line 18b top		pf beam line 18b top	pf beam line 18b
----------------------------------	--	----------------------	------------------

# THE *SCALEPACK* MANUAL

## About the Scalepack manual ( from Edition 4 )

The manual is organized somewhat differently than the *Scalepack* manual. Compared to film or IP processing, scaling is complex. There's a lot more to scaling data than just plugging in a bunch of *.x* files and getting a merged data set out the other side. In particular, scaling is where you make explicit estimates of the errors associated with your experiment, and these error estimates are crucial in determining the statistics by which you judge the quality of your data. In addition, *Scalepack* is also used for many other manipulations of the diffraction data. As a result, we decided to make the heart of the manual a series of scenarios, each accompanied by a sample command file, which correspond to many of the typical applications of *Scalepack*. If you do not find your particular problem addressed by these scenarios, we hope that they, along with the keyword documentation, will provide enough guidance so that you can construct your own.



# Program Description

The scaling algorithm used in is described in: Fox G.C., Holmes K.C., 1966 "An Alternative Method of Solving the Layer Scaling Equation of Hamilton, Rollet & Sparks" Acta Cryst. 20:886-891

The scaling and merging of different data sets as well as the global refinement of crystal parameters (postrefinement) is performed by . Unlike in other procedures, in the **estimated error** of the measurement is enlarged by a fraction of the **expected**, rather than the observed, intensity. **This algorithm reduces the bias towards reflections with an integrated intensity below the average.**

Due to the correlation between crystal and detector parameters the values of the unit cell parameters refined from a single image may be quite imprecise. This lack of precision is of little significance to the process of integration, as long as the predicted positions are on target. There is no contradiction here, because at some crystal/detector orientation the positions of the reflections may only weakly depend on the value of a particular crystal parameter. At the end of the data reduction process one would wish to get precise unit cell values. This is done in a procedure referred to as a global refinement or **postrefinement**. The implementation of this method in allows for separate refinement of the orientation of each image, but with the same unit cell value for the whole data set. In each batch of data (a batch is typically one image) different unit cell parameters may be poorly determined. However, in a typical data set there are enough orientations to determine precisely all unit cell lengths and angles. Global refinement is also more precise than the processing of a single image in the determination of crystal mosaicity and the orientation of each image.

allows you to analyze the **quality of the diffraction data** from your crystals. The quality of the diffraction data is judged by many standards, ultimately including the quality of the electron density map which results, but a widely used objective standard is based on the statistics which result from scaling and merging the raw diffraction data. This program calculates **single, isotropic scale** and **B factors** for each of the "films" or "frames" of processed data that are input. Scaled, merged data are output.

**A multiplicative correction factor**

$$= \frac{2 \left( \frac{\sin \theta}{\lambda} \right)^2}{\sigma^2 + \langle \sigma^2 \rangle}$$

is applied to I and  $\sigma(I)$ . Merging of redundant measurements, which occurs throughout the refinement process, is done as follows. The multiplicative factor for each measurement is calculated from the scale and B of its film and applied to I to give Icorr.  $\langle \sigma^2 \rangle$  is a solution to the following equations:

$$= \frac{1}{(\sigma^2 + E_1)^2 + (\langle \sigma^2 \rangle + E_2)^2}$$

where  $E_1$  and  $E_2$  are the input variables **error scale factor** and **estimated error**, respectively, and

$$\langle \quad \rangle = \frac{\sum}{\sum}.$$

The output  $\sigma(I)$  is

$$\frac{1}{\sqrt{\sum}}$$

During **parameter refinement** the scale (and B, if requested) for all refinable films are shifted simultaneously to minimize the difference between the  $\langle \quad \rangle$ 's and  $\quad$ 's for individual measurements, summed over all reflections on all films.  $\langle \quad \rangle$ 's are redetermined in each cycle as described above. One (or more) of the "batches" that is input is designated as the reference batch, and its scale and B factors are not refined. There is room for many reference batches; this program will also work with no reference batch in which case **eigenvalue filtering** stabilizes the shifts in scale and B factors.

Initial default scale and B factors are applied to each film, this can be replaced by explicit input of initial values. Initial scale factors should be around 1.0 (or larger for very strong data), so the output values are not too large, otherwise they can overflow fixed formats.

The user controls whether **anomalous pairs** are treated as equivalents or not and whether they are merged. The program rejects measurements that fail an  $I/\sigma(I)$  test when they are read in, and these do not contribute to the output. Partially recorded reflections can be added or not, according to the control input. Individual measurements can be rejected explicitly in the control input.

**Control input** is like that for  $\quad$ ; each record has a title and the order matters only to some degree. The title tells what input is expected to follow. All **file names** must be 80 characters or less and it is usually safer to enclose all file names in single quotes ('myxtal##.x'). Usually input variables follow in the same record as the title, except where the input is a table or a list. Anything between [ and ] in the command line is considered a comment by the program. A semicolon (;) terminates input. Unlike  $\quad$ ,  $\quad$  is usually run as a batch job because there is no point in watching your frames scale and because the  $\quad$  log file is where the statistics associated with scaling and merging are presented.

## Statistics and Scalepack

The quality of X-ray data is initially assessed by statistics. In **small molecule** crystallography there is almost always a large excess of strong data, so this allows the crystallographer to discard a substantial amount of suspect data and still accurately determine a structure. Compared to small molecules, however, **proteins** diffract poorly. Moreover, important phase information comes from weak differences and we must be sure these differences do not arise from the noise caused by the limitations of the X-ray exposure and detection apparatus. As a result, we cannot simply throw away or statistically down-weight marginal data without first making a sophisticated judgment about which data is good and which is bad.

To accurately describe the structure of a protein molecule, we ordinarily need higher resolution data than the crystal provides. That is life. One of the main judgment the crystallographer makes in assessing the quality of his data is thus the **resolution** to which his crystal diffracts. In making this judgment, we wish to use the statistical criteria which are most discriminatory and which are the least subjective. In practice, there are two ways of assessing the high resolution limit of diffraction. The first is **the ratio of the intensity to the error of the intensity**, i.e.  $I/\sigma$ . The second way, which is traditional but inferior, is the agreement between symmetry related reflections, i.e.  $R_{\text{merge}}$ .

From a statistical point of view,  **$I/\sigma$  is a superior criterion**, for two reasons. First, it defines a resolution “limit” since by definition  **$I/\sigma$  is the signal to noise** of your measurements. In contrast,  $R_{\text{merge}}$  is not directly related to signal to noise.

Second, the  $\sigma$  assigned to each intensity derives its validity from the  $\chi^2$ 's, which represent the **weighted** ratio of the difference between the observed and average value of  $I$ ,  $\langle I \rangle$ , squared, divided by the square of the error model, the whole thing times a factor correcting for the correlation between  $I$  and  $\langle I \rangle$ . Since it depends on an explicit declaration of the expected error in the measurement, the user of the program is part of the Bayesian reasoning process behind the error estimation.

The essence of **Bayesian reasoning** in  $\chi^2$  is that you bring  $\chi^2$  (or technically speaking, the goodness-of-fit, which is related to the total  $\chi^2$  by a constant) close to 1.0 by manipulating the parameters of the error model.  $R_{\text{merge}}$ , on the other hand, is an **unweighted** statistic which is independent of the error model. It is sensitive to both intentional and unintentional manipulation of the data used to calculate it, and may not correlate with the quality of the data. An example of this is seen when collecting more and more data from the same crystal. As the redundancy goes up, the final averaged data quality definitely improves, yet the  $R_{\text{merge}}$  also goes up. As a result,  $R_{\text{merge}}$  is only really useful when comparing data which has been accumulated and treated the same. This will be discussed again later.

In short,  **$I/\sigma$  is the preferred way** of assessing the quality of diffraction data because it derives its validity from the  $\chi^2$  (likelihood) analysis. Unless all of the explicit and implicit assumptions which have been made in the calculation of an  $R_{\text{merge}}$  are known, this criterion is less meaningful. This is particularly true when searching for a “number” which can be used by others to critically evaluate your work.

There are two modes of analysis of data using  $\chi^2$ 's. The first mode keeps the  $\chi^2$  (or more precisely, the goodness-of-fit) constant and compares the error models. Basically, this means that you are adjusting your estimates of the errors associated with the measurement until the deviations within observations agree with the expectation based on the error model.

The second mode keeps the error model constant and compares  $\chi^2$ 's. This mode is computationally much faster and is used in refinement procedures. Of the two modes, the first is more informative, because it forces you to consider changes in the error model. Which mode you use generally depends on what you are comparing. When assessing the general quality of your detector, the first mode is used. When comparing a **derivative** to a **native**, the second mode is used due to an incomplete error model which does not take into account important factors like non-**isomorphism**. Thus, the  $\chi^2$  of scaling between a native and a derivative provides a measure of non-isomorphism, assuming the detector error is accurately modeled for both samples.

$R_{\text{merge}}$  was historically used as an estimate of the non-isomorphism of data collected on film using several different crystals, and for this purpose it still has some validity because we do not account for non-isomorphism in our error model. It is not so important now that complete x-ray data sets are collected from single, frozen crystals.

One of the drawbacks of using  $R_{\text{merge}}$  as a measure of the quality of a data set is that it **can be intentionally and unintentionally manipulated**. Unintentional factors which can artificially lower  $R_{\text{merge}}$  generally have the effect of reducing the redundancy of the data or eliminating weaker observations. In crystallography, the greater the redundancy of the data, the worse the  $R_{\text{merge}}$ , because of the correlation between  $I$  and  $\langle I \rangle$  which reduces the  $R_{\text{merge}}$ . The greater the redundancy, the lower the correlation. For two measurements with the same  $\sigma$ , the correlation is 50%, so  $R_{\text{merge}}$  is underestimated by  $\sqrt{2}$  compared to the case of no correlation. Known unintentional factors which lower  $R_{\text{merge}}$  include the following:


1. Data collected so that lower resolution shells, where the data is strong, have a higher redundancy than the higher resolution shells, where the data is generally weaker. This can be accomplished by collecting data on detectors where  $2\theta \neq 0$ , or including data from the corners of rectangular or square image plates. There is nothing wrong with using this data; it will just artificially lower the  $R_{\text{merge}}$ .
2. Inclusion of single measurements in the calculation of  $R_{\text{merge}}$  in one widely used program, which is why a table using this erroneous calculation used to be presented in the output. Although the bug in the widely used program was unintentional, it nonetheless reduced the  $R_{\text{merge}}$  and this may have accounted for its longevity. A second, more subtle bug that reduced  $R_{\text{merge}}$  prompted the introduction of the keyword **BACKGROUND CORRECTION**. Fortunately, both bugs have now been fixed, but the point is that errors of this type can persist.
3. Omission of negative or weak reflections from the calculation of  $R_{\text{merge}}$ . This is often undocumented behavior of crystallographic data scaling/merging software. Examples include:
  - a) elimination of a reflection from both the data set and the calculation of  $R_{\text{merge}}$  if  $I < 0$
  - b) conversion of  $I < 0$  to  $I = 0$  before the calculation of  $R_{\text{merge}}$  and inclusion of this reflection in the data set (the statistics of such a type are included in the Scalepack output for reasons of comparison. This is the first  $R_{\text{merge}}$  table in the log file, not the final one).
  - c) omitting reflections with  $\langle I \rangle < 0$  from the calculation of  $R_{\text{merge}}$  but inclusion of these reflections in the output data set.
4. Default sigma cutoffs set to unreasonable values, like 1. This is in fact the default of the R-Axis software.
5. Use of unreasonable default/recommended rejection criteria in other programs. These eliminate individual  $I$ 's which should contribute to  $R_{\text{merge}}$  and yet are still statistically sound measurements.
6. Use of the eigenvalue filter to determine the overall B factor of a data set collected on a non-frozen, decaying crystal. In this case, the eigenvalue filter will calculate an overall B factor which is appropriate for the middle of the data set, yet apply this to all data. As a result, the high resolution data will be down weighted compared to data processed with the first, least decayed frame as the reference. The high resolution data is generally weaker than the low resolution data, and as a result is more likely to result in higher  $R_{\text{merge}}$ . By down-weighting the high resolution data, the  $R_{\text{merge}}$  is artificially lowered. Any program which does not allow the option of setting the reference frame will have this problem. Of course, there is no problem with non-decaying crystals.

There are also intentional ways of lowering your  $R_{\text{merge}}$ . Like those ways listed above, they generally result from the statistically invalid elimination of weak reflections, reduction of the redundancy of the data, or de-emphasis of weak data. The difference between these methods and those listed above is that they are generally under the control of the user.

1. Use of an unreasonable **SIGMA CUTOFF** (e.g.  $\geq 0$ ). The rejection of weak data will always improve  $R_{\text{merge}}$ . There is a further discussion of **SIGMA CUTOFF** in the section.
2. Use of a resolution limit cutoff. Again, the omission of weak data will improve  $R_{\text{merge}}$ . A reasonable resolution cutoff is the zone where  $I/\sigma < 2$ .
3. Combining into a single zone for the purposes of calculations those resolution shells where  $R_{\text{merge}}$  is rapidly changing. In this case, the shell will be dominated by the strong data at the low resolution end of the zone and give the impression that the high resolution limit of the zone has better statistics than it really does. For example, if you combined all your data into a

single zone, the  $R_{\text{merge}}$  in the        shell would be pretty good ( $=R_{\text{merge}}$  overall), when in fact it was substantially worse. It is more sensible to divide your zones into equal volumes and have enough of them so that you can accurately monitor the decay with resolution.

4. Omitting partially recorded reflections. This has the effect of a) reducing the redundancy, and b) eliminating poorer reflections. Partially recorded reflections will always have a higher  $\sigma$  associated with them because they have a higher total background, due to the inclusion of background from more than one frame in the reflection.
5. Scaling I+ and I- reflections separately in the absence of a legitimate anomalous signal (**SCALE ANOMALOUS**) . This has the effect of reducing the redundancy.
6. Ignoring overloaded reflections using the **IGNORE OVERLOADS** in Scalepack. The intensity of overloaded or saturated, reflections cannot be directly measured, because obviously some of the pixels are saturated. Profile fitting only measures these reflections indirectly, by fitting a calculated profile to the spot using the information contained in the        or        of the spot. Ignoring the inaccuracies inherent in this method by ignoring overloads may have a dramatic effect on  $R_{\text{merge}}$ .

 **IGNORE OVERLOADS** is often a useful tool, however. For example, when calculating anomalous differences you do not want to use overloaded reflections because you are looking for very very small differences and want to use only the most accurate data. Another time you might ignore overloads is when you collect multipass data. In this case, a crystal is exposed twice, once for a short time, the other for a longer time. The longer exposure is to sufficiently darken the high resolution reflections, but will result in saturated low resolution reflections. Since the low resolution reflections can be obtained from the short exposures, the overloaded ones can be ignored in the long exposures.

See: Press, William H., Teukolsky Saul A., Vetterling William T., Flannery Brian P. "Numerical Recipes in C, The Art of Scientific Computing", Second Edition, Cambridge University Press, 1992

# Implementation

## Limits on variable dimensions

The current standard version (scalepack) allows up to 2,000,000 total observations, 1,000,000 unique reflections, and 2000 films (frames) or batches. The program will refine up to 3000 parameters (e.g. scale factor, B factor) simultaneously. This should be enough for the most uses of the program. If the number of observations, number of unique reflections, etc. exceeds the standard scalepack limits, you may use a bigger version of scalepack. This is usually only a problem in virus crystallography or in case of very high resolution data. For this purpose three other versions of `scalepack` have been made: `scalepack-2m` (limit  $2 \times 10^6$  unique reflections), `scalepack-3m` ( $3 \times 10^6$  unique reflections) and `scalepack-9m` ( $9 \times 10^6$  unique reflections). Below the maximum values of several parameters are shown for different scalepack versions:

Scalepack	Scalepack- 8m

# Scenarios

allows you to analyze the quality of the diffraction from your crystals providing the statistics which result from scaling and merging the raw diffraction data. Scenarios presented in this chapter are designed to show you how to get accurate data and statistics in situations typically encountered in protein crystallography. These scenarios consist of input command files, followed by some comments about the statistics which result.

Scalepack is run in batch mode, unlike *Scale* which has a graphical display component and is best run interactively. All of the scenarios are thus given as command files. For clarity the first two or three lines of each command file which set the directory and call the program have been omitted but are given here:

```
setenv cnvrr yes
cd /frames/mydir/Denzo
Scalepack <<STOPINPUT>> scale.log
...
STOPINPUT
```

It is necessary to run the program this way, rather than making each of the scenarios an input file called by the @ command, e.g. `@scalesimple.inp`, because the Scalepack parser does not recognize nested @ commands. This is important for Scalepack because the list of rejected reflections is usually called with the `@reject` command. *Just one of those things.*

## List of Scenarios

1. Scale data from one crystal processed with
2. Scale data from one crystal not processed with
3. Heavy atom search: Scale native + a few frames of a potential derivative
4. Scale multiple native data sets together. Data from
5. Reindexing
  - a. Switch *a* and *c* in P2<sub>1</sub> or P2
  - b. Change the order of *a*, *b*, *c* to *b*, *c*, *a* in orthorhombic
  - c. Convert *a*, *b*, *c* to *c*, *a*, *b* in orthorhombic
  - d. Reindex P23, switching *a*, *b*
  - e. P321 or P312: reindex data and combine with another data set
  - f. P6, P4, or H3: reindex data and combine with another data set
  - g. Reindex from H32 to R32
  - h. Reindex two H3 data sets, each with inconsistent index, and combine
  - i. Reindex four P3<sub>1</sub> data sets, each with inconsistent index, and combine
6. Comparing two data sets from Scalepack
7. Compare I<sup>+</sup> and I<sup>−</sup> : detection of an anomalous signal
8. Fitting polarization
9. High resolution + low resolution data sets from one crystal
10. Thin frames or few frames (**SCALE RESTRAIN** and **B RESTRAIN**)
11. Space group determination

## Scenario 1: Scale Data from One Crystal Processed with Denzo

[scalesimple.com](http://scalesimple.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	estimated error for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.3	Multiplicative factor applied to input $\sigma$
<b>REJECTION PROBABILITY</b> 0.0001 [@reject]	expected fraction of outliers in the data. [read in rejected reflections]
<b>WRITE REJECTION FILE</b> 0.9	observations with greater than 90% chance of being outliers are written to the reject file
<b>SPACE GROUP</b> P21	
<b>OUTPUT FILE</b> mydata.sca	output file containing $h, k, l, I, \sigma_I$
<b>REFERENCE BATCH</b> 1	reference batch in scaling procedure. Can also leave out and let eigenvalue filter set reference scale and B
<b>ADD PARTIALS</b> 1 to 90	partials in batches 1 through 90 summed
<b>POSTREFINE</b> 10	number of cycles of postrefinement
<b>FIT</b> crystal a* 1 to 90	refine unit cell parameters based on all data from batches 1 through 90.
<b>FIT</b> crystal b* 1 to 90	
<b>FIT</b> crystal c* 1 to 90	► space group dependent! e.g. don't refine alpha for monoclinic space groups.
<b>FIT</b> crystal beta* 1 to 90	
<b>FIT</b> batch rotx 1 to 90	crystal orientation of spindle axis angle will be refined for each batch (frame) separately. This allows you to spot slippage & other problems.
<b>FIT</b> batch roty 1 to 90	one mosaicity value will be fit
<b>FIT</b> crystal mosaicity 1 to 90	format of the input intensity data
<b>FORMAT</b> denzo_ip	these values will be substituted in for the ###
<b>SECTOR</b> 1 to 90	batch numbers will start at 1 for file myxtal001.x and increase to 90 for file myxtal090.x.
<b>FILE</b> 1 '/frames/denzo/myxtal###.x'	

➡ Proper scaling requires multiple rounds of **SCALEPACK**, as the statistics may improve as outliers are rejected. On the second and all subsequent rounds of Scalepack the brackets would be removed from the **@REJECT** command on line 9 and outliers rejected in previous rounds would be removed from the data prior to refinement of scale and B factors. It usually takes about 1 to 3 iterations before no more reflections are rejected and the process converges.

The log file should be examined after each iteration. In particular, the errors,  $\chi^2$  and both R factors should be checked. You can adjust the error estimates in the command file to improve the accuracy of the statistics. Here are some helpful guidelines:

1. Examine the difference between the total error and the statistical error in the final table of statistics. The difference between these numbers tells you what contribution statistical error makes to the total error ( $\sigma$ ). If the difference is small, then reasonable changes in the values following the **ESTIMATED ERROR** keyword will not help your  $\chi^2$  much. This is because the estimated errors represent your guess / knowledge of the contribution of systematic error to the total error and a small difference indicates that estimated error is not contributing much. If the difference between total and statistical error is significant, and the  $\chi^2$  are far from 1, then consider adjusting the **ESTIMATED ERROR** values in the affected resolution shells.



2. Examine the difference between the linear and square  $R_{\text{merge}}$ s. For this comparison to be valid, you cannot have strong overloads in your data, so the command **IGNORE OVERLOADS** must be given for this analysis. If  $R_2$  is larger than  $R$ , then the distribution of the deviations ( $I - \langle I \rangle$ ) is skewed. This implies that you have too many outliers in your data, so you should reject more reflections. The way to increase rejections is to choose a higher value of **REJECTION PROBABILITY** or write rejections to the reject file which have a lower probability of being outliers. Useful increments are factors of 10 in **REJECTION PROBABILITY** and replacing default value of **WRITE REJECTION FILE** (0.9) with 0.5. If after one round this does not bring  $R_2$  close to  $R$  (e.g.  $R_2$  is more than 50% greater than  $R$ ), then you should consider whether you have detector problems. Likely culprits include the shutter and the spindle. Note that bad crystals (split, twinned) and ice rings can also produce this phenomenon, as can the inclusion of a few frames of data from a badly decayed crystal.
3. If  $R \approx R_2$  and your  $\chi^2$  are high, then your crystal may have suffered radiation damage which cannot be corrected by scale and B factors. This type of damage may not even increase B factors or decrease scale factors. This is frequently seen with non-frozen crystals.
4. The  $\chi^2$  is your best indication of whether your error model accurately reflects the actual errors in your data. Since the errors in your data are expressed in the  $\sigma$  value you attach to every I, it is important that your error model be correct.  $\chi^2$  values of close to 1 indicate that your model is pretty good. If your  $\chi^2$  values deviate much from 1, you can fix this by adjusting the value of the error scale factor. Note that you don't have to hit 1.0, since  $\chi^2$  goes as the square of your error parameters, so 20% deviations from 1.0 represent only a 10% adjustment to  $\sigma$ , which is OK for most practical purposes. Even a  $\chi^2 = 2$  is unlikely to seriously affect your subsequent calculations.

## Scenario 2: Scale Data from One Crystal not Processed with Denzo

[scalexengen.com](http://scalexengen.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.0	Multiplicative factor applied to input $\sigma$ s
<b>REJECTION PROBABILITY</b> 0.0001	expected fraction of outliers in the data. Here expect 1/10,000 will be outliers
<b>SPACE GROUP</b> P21	
<b>UNIT CELL</b> 39 94 67 90 96.1 90	<b>UNIT CELL</b> needed because not in header
<b>OUTPUT FILE</b> mydata.sca	<b>OUTPUT FILE</b> containing $h\ k\ l$ , $I$ , $\sigma_I$
<b>REFERENCE BATCH</b> 101	<b>REFERENCE BATCH</b> in scaling procedure. Can also leave out and let eigenvalue filter set reference scale and B factor
<b>[POSTREFINE 10]</b>	can't <b>POSTREFINE</b> because not Denzo data
<b>@reject</b>	file with list of rejected reflections from previous round of
<b>WRITE REJECTION FILE</b> 0.9	observations with greater than 90% chance of being outliers are written to the reject file
<b>FORMAT</b> xengen urf	<b>FORMAT</b> of the input intensity data
<b>SECTOR WIDTH</b> 5.0	must specify these two because not in header
<b>FRAME WIDTH</b> 0.2	
<b>PROFILES SUMMED</b>	since didn't run , this may be better than <b>PROFILES FITTED</b>
<b>FILE</b> 1 ctbn1.urf	urf files are an intermediate file format created by
<b>FILE</b> 101 ctbn2.urf	
<b>FILE</b> 201 ctbn3.urf	this would be data from 4 crystals or perhaps 4 sweeps of
<b>FILE</b> 301 ctbn4 urf	the same crystal. A sweep can be a full data set.

➡ Because the file format is not from , you will not be able to **POSTREFINE** your data to get better values of the cell constants. Partial sums are not summed in these cases because the (or , or ) software has already done these tasks. You would perform the iterative refinement with this data just as described above for Scenario 1.

➡ Some of the programs produce binary files that may not be able to read-in if runs on a different computer architecture than data reduction program that produced binary file.

**Scenario 3: Heavy Atom Search: Scale native plus a few frames of a potential derivative**

[scalepotderiv.com](http://scalepotderiv.com)

[***The Native Dataset***]	
NUMBER OF ZONES 10	number of resolution shells for statistics
ESTIMATED ERROR 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	ESTIMATED ERROR for each resolution shell. Set to zero because native has already been scaled
[ERROR SCALE FACTOR 1.0]	don't need here because data already scaled
SPACE GROUP P3121	SPACE GROUP of native & derivative duh!
FORMAT Scalepack	native data has been processed already and is in
FILE 101 native.sca	format.
REFERENCE BATCH 101	REFERENCE BATCH in scaling procedure. We will compare derivative to native, so native is reference
[***Potential Derivative***]	
RESOLUTION 3.5	better to search at medium to low resolution
ESTIMATED ERROR 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05	derivative is new crystal so need new ESTIMATED ERROR
ERROR SCALE FACTOR 1.6	we need this here because new data
REJECTION PROBABILITY 0.0001	expected fraction of outliers in the data. This is critical because the presence of be indicative of a potential derivative!
IGNORE OVERLOADS	this removes overloaded reflections from statistics. Overloads may cause large differences but are not indicative of a derivative
ADD PARTIALS 1 to 3	partials in derivative will be summed
FORMAT denzo_ip	data format of potential derivative
[HKL MATRIX -1 0 0 0 -1 0 0 0 1]	only reindex if native and derivative have different indices. See scenario #5
SECTOR 1 to 3	3 frames of a potential Hg derivative e.g. 3 x 1 degree oscillations.
FILE 1 'NewHg###.x'	
OUTPUT FILE 'junk.sca'	not interested in .sca file, only the log

➡ The goal here is to scale a few frames of a potential derivative with a native data set. Here you only do a single round of refinement, not the iterative cycles done in Scenario 1 because you are not doing any cycles of rejection. In this case three frames of the derivative, say as much as 3 degrees, will be compared with the native to a resolution of 3.5 Å.

➡ This scenario is only valid for a small wedge because otherwise you will be looking at  $R_{\text{merge}}$ s which result from reflections combined with the derivative data set as well as with those of the native and so the statistics will be less indicative.

We are only interested in the last page of the Scalepack log file, where we will examine the  $\chi^2$  values and  $R_{\text{merge}}$ . Here are some general rules for assessing your potential derivative.

- ➡ you should pay most attention to the low resolution shells, where the measurements are most accurate.
1. If  $\chi^2 \approx 1$  then your crystal scales well with the native so it is not a useful derivative. Remember, derivatives are different!
  2. If  $\chi^2$  = very large (say 50) then you have either severe non-isomorphism (sorry), your have very heavy substitution (but it is rare to have such a strong isomorphous difference signal), or your derivative is misindexed relative to the native. Reindexing is only a potential problem in the case of all trigonal, P6, R3, P4, I4, I23, and P23 Laue groups, and in cases where there are accidental metric tensor symmetries (e.g. monoclinic  $b \approx 90^\circ$ , or orthorhombic  $a = b$ ). To reindex, apply an **HKL MATRIX** to test the other possibilities ( ) and see which one gives the lowest  $\chi^2$  and  $R_{\text{merge}}$ . That is then the correct index.
  3. If  $\chi^2 \approx 10$  and the detector is not a problem (as indicated by  $R_2 \approx R$ ), then this indicates a potential derivative because you know that the differences are much larger than the expected measurement error.
  4. If  $\chi^2 \approx 2$  you either have a poorly substituted derivative or a slight detector problem. Either way this is not going to be your best derivative. The resolution shell where the  $\chi^2$  drops to about 2 represents the useful limit of your derivative measurements.
- ➡ Instead of three 1 degree photographs, in many cases you can get away with a single 3 degree oscillation.

## Scenario 4: Scale Multiple Native Data Sets Together. Data from Denzo

There are two ways to scale multiple native data sets together. The first and preferred way is to scale all of the .x files together. If you do it this way, you get to do postrefinement over all of the raw data, and the scaling is applied at the level of the individual frames, rather than over the whole data set. This is probably a little more accurate. Also it saves you from having to run twice: once for the second native data set and then to scale the two .sca files together.

[scaletwo.com](http://scaletwo.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.3	Multiplicative factor applied to input $\sigma$ s
<b>REJECTION PROBABILITY</b> 0.0001	expected fraction of outliers in the data.
<b>WRITE REJECTION FILE</b> 0.9	reflections with $\geq 90\%$ probability as outlier are written to the rejection file
<b>SPACE GROUP</b> P21	
<b>REFERENCE BATCH</b> 1	may use the first frame as reference for scaling or can omit this and use eigenvalue filter
[@reject]	[reject outliers from previous rounds]
<b>POSTREFINE</b> 10	10 cycles of postrefinement
<b>FIT</b> crystal a* 1 to 327	
<b>FIT</b> crystal b* 1 to 327	
<b>FIT</b> crystal c* 1 to 327	
<b>FIT</b> crystal beta* 1 to 327	
<b>FIT</b> batch rotx 1 to 327	
<b>FIT</b> batch roty 1 to 327	
<b>FIT</b> batch mosaicity 1 to 327	
<b>ADD PARTIALS</b> 0 to 103 201 to 327	the consecutive sequences must be declared here, otherwise program will try to <b>ADD PARTIALS</b> from non-consecutive frames
<b>FORMAT</b> denzo_ip	
[****Data Set #1****]	
<b>SECTOR</b> 1 to 103	first data set has 103 frames
<b>FILE</b> 1 'my_crystal1###.x'	
[****Data Set #2****]	
<b>SECTOR</b> 1 to 127	second data set sectors in batches starting with 201
<b>FILE</b> 201 'my_crystal2###.x'	
<b>OUTPUT FILE</b> 'my_crystal12.sca'	

► note which parameters are fit over the two data sets and which are for each frame

Alternatively, if you have already processed the two data sets and have two .sca files you want to scale and merge them, you can use the following command file.



Note that **you cannot POSTREFINE or ADD PARTIALS between the two data sets**. You are basically scaling and merging one set of processed and outlier-rejected native data to another.

[scaletwosca.com](http://scaletwosca.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	<b>ESTIMATED ERROR</b> for each resolution shell. Set to zero because native has already been scaled
<b>ERROR SCALE FACTOR</b> 1.0	Multiplicative factor applied to input $\sigma$ 's expected fraction of outliers in the data.
<b>FORMAT</b> Scalepack	
[****Data Set #1****]	
<b>FILE</b> 1 'my_crystal1.sca'	
[****Data Set #2****]	
<b>FILE</b> 2 'my_crystal2.sca'	
<b>OUTPUT FILE</b> 'my_crystal12.sca'	

## Scenario 5: *Reindexing*


Reindexing involves reassigning indices from one unit cell axis to another. This becomes an important issue when comparing two or more data set which were collected and processed independently. This is because *STAD*, when confronted with a choice of more than one possible indexing convention, makes a random choice. This is no problem, except that if it makes a different choice for a second data set, the two will not be comparable without reindexing one or the other. One cannot distinguish non-equivalent alternatives without scaling the data, which is why this is not done in *STAD*. You can tell if you need to reindex a data set if the  $\chi^2$  values upon merging the two are very high (e.g. 50). This makes sense when

Change order of **a, b, c** to **b, c, a** in orthorhombic

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.4	multiplicative factor applied to input $\sigma$
<b>REJECTION PROBABILITY</b> 0.0001	
<b>WRITE REJECTION FILE</b> 0.9	
@reject	
<b>RESOLUTION</b> 40 3.3	
<b>SPACE GROUP</b> P21212	can also use for P2221
<b>REFERENCE BATCH</b> 1	
<b>ADD PARTIALS</b> 1 to 180	
<b>POSTREFINE</b> 10	good idea to postrefine independent of reindexing
<b>FIT</b> crystal a* 1 to 180	
<b>FIT</b> crystal b* 1 to 180	
<b>FIT</b> crystal c* 1 to 180	
<b>FIT</b> batch rotx 1 to 180	
<b>FIT</b> batch roty 1 to 180	
<b>FIT</b> batch mosaicity 1 to 180	
<b>FORMAT</b> denzo_ip	
<b>HKL MATRIX</b> 0 1 0 0 0 1 1 0 0	reindexing matrix: h ► k k ► l l ► h
<b>SECTOR</b> 1 to 180	
<b>FILE</b> 1 '/frames/myxtal###.x'	
<b>OUTPUT FILE</b> 'myxtal.sca'	<b>OUTPUT FILE</b> contains reindexed reflections



Convert *a*, *b*, *c* to *c*, *a*, *b* in orthorhombic

NUMBER OF ZONES 10	number of resolution shells for statistics
ESTIMATED ERROR 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	ESTIMATED ERROR for each resolution shell
ERROR SCALE FACTOR 1.4	Multiplicative factor applied to input $\sigma$
REJECTION PROBABILITY 0.0001	
WRITE REJECTION FILE 0.9	
@reject	
RESOLUTION 40 3.3	
SPACE GROUP P21212	
REFERENCE BATCH 1	
ADD PARTIALS 1 to 180	
POSTREFINE 10	
FIT crystal a* 1 to 180	
FIT crystal b* 1 to 180	
FIT crystal c* 1 to 180	
FIT batch rotx 1 to 180	
FIT batch roty 1 to 180	
FIT batch mosaicity 1 to 180	
FORMAT denzo_ip	
HKL MATRIX 0 0 1 1 0 0 0 1 0	reindexing matrix: 
SECTOR 1 to 180	
FILE 1 '/frames/Denzo/myxtal###.x'	
OUTPUT FILE 'myxtal.sca'	OUTPUT FILE contains reindexed reflections

Reindex P23, switch **a, b**

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.4	multiplicative factor applied to input $\sigma$
[ <b>REJECTION PROBABILITY</b> 1.0e-4] [ <b>WRITE REJECTION FILE</b> 0.9] [@reject]	commented out to spare log file in case this reindexing is incorrect
<b>RESOLUTION</b> 40 3.3	
<b>SPACE GROUP</b> P23	
<b>REFERENCE BATCH</b> 1001	
<b>ADD PARTIALS</b> 1 to 180 1001 to 1180	
<b>POSTREFINE</b> 10	
<b>FIT</b> crystal a* 1 to 180	
<b>FIT</b> batch rotx 1 to 180	
<b>FIT</b> batch roty 1 to 180	
<b>FIT</b> batch mosaicity 1 to 180	
<b>FORMAT</b> denzo_ip	
<b>HKL MATRIX</b> 0 1 0 -1 0 0 0 0 1	reindexing matrix: h ► k k ► -h also works for I23 and F23
<b>SECTOR</b> 1 to 180	
<b>FILE</b> 1 '/frames/myxtal2###.x' [*** Unit Matrix ***]	this data will be reindexed by the matrix above
<b>HKL MATRIX</b> 1 0 0 0 1 0 0 0 1	data set #1 to be scaled with reindexed data
<b>SECTOR</b> 1 to 180	
<b>FILE</b> 1001 '/frames/myxtal1###.x'	this data will not be reindexed
<b>OUTPUT FILE</b> 'myxtal.sca'	output combines 2 data sets with consistent indexes

## P321 or P312. Reindex data and combine with another data set

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.4	multiplicative factor applied to input $\sigma$
[ <b>REJECTION PROBABILITY</b> 0.0001]	commented out to spare log file in case
[ <b>WRITE REJECTION FILE</b> 0.9]	reindexing is incorrect.
[@reject]	
<b>RESOLUTION</b> 40 3.3	
<b>SPACE GROUP</b> P321	
<b>REFERENCE BATCH</b> 101	
<b>ADD PARTIALS</b> 1 to 60 101 to 160	
<b>POSTREFINE</b> 10	good idea to <b>POSTREFINE</b>
<b>FIT</b> crystal a* 1 to 160	
<b>FIT</b> crystal c* 1 to 160	
<b>FIT</b> batch rotx 1 to 160	
<b>FIT</b> batch roty 1 to 160	
<b>FIT</b> batch mosaicity 1 to 160	
<b>FORMAT</b> denzo_ip	
<b>HKL MATRIX</b> -1 0 0 0 -1 0 0 0 1	reindexing matrix: h ► -h k ► -k also works for P312, P3121, P3221, P3112 and P3212
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 1 '/frames/myxtal2###.x'	this data will be reindexed by matrix above
<b>[***Unit Matrix***]</b>	
<b>HKL MATRIX</b> 1 0 0 0 1 0 0 0 1	data set #1 to be scaled with reindexed data
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 101 '/frames/myxtal1###.x'	this data will not be reindexed
<b>OUTPUT FILE</b> 'myxtals12.sca'	output combines 2 data sets with consistent indexes

P6, P4 or H3: Reindex data and combine with another data set

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.4	multiplicative factor applied to input $\sigma$
[REJECTION PROBABILITY 1.0e-4]	
[WRITE REJECTION FILE 0.9]	
[@reject]	
<b>RESOLUTION</b> 40 3.3	
<b>SPACE GROUP</b> P6	
<b>REFERENCE BATCH</b> 1	
<b>ADD PARTIALS</b> 1 to 60 101 to 160	
<b>POSTREFINE</b> 10	good idea to <b>POSTREFINE</b> because this will affect partials
<b>FIT</b> crystal a* 1 to 160	
<b>FIT</b> crystal c* 1 to 160	
<b>FIT</b> batch rotx 1 to 160	
<b>FIT</b> batch roty 1 to 160	
<b>FIT</b> batch mosaicity 1 to 160	
<b>FORMAT</b> denzo_ip	
<b>HKL MATRIX</b> 0 1 0 1 0 0 0 0 -1	reindexing matrix: h ► k k ► h l ► -l, also works for p4 and h3
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 1 '/frames/myxtal2###.x'	this data will be reindexed by matrix above
[*** <b>Unit Matrix</b> ***]	
<b>HKL MATRIX</b> 1 0 0 0 1 0 0 0 1	data set #1 to be scaled with reindexed data
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 101 '/frames/myxtal1###.x'	this data will not be reindexed
<b>OUTPUT FILE</b> 'myxtals12.sca'	output combines 2 data sets with consistent indexes

Reindex from H32 to R32

NUMBER OF ZONES 10					number of resolution shells for statistics
ESTIMATED ERROR 0.04 0.04					ESTIMATED ERROR for each resolution shell
0.04	0.04	0.04	0.04	0.04	
0.04	0.04	0.04			
ERROR SCALE FACTOR 1.4					multiplicative factor applied to input $\sigma$
REJECTION PROBABILITY 1.0e-4					
WRITE REJECTION FILE 0.9					
@reject					
RESOLUTION 40 3.3					
SPACE GROUP R32					after reindexing this will be the space group
REFERENCE BATCH 1					
ADD PARTIALS 1 to 90					
POSTREFINE 10					good idea to <b>POSTREFINE</b> because this will affect partials
FIT crystal a* 1 to 90					
FIT crystal c* 1 to 90					
FIT batch rotx 1 to 90					
FIT batch roty 1 to 90					
FIT batch mosaicity 1 to 90					
FORMAT denzo_ip					
HKL MATRIX -1 0 1					
1 -1 0					
1 1 1					
SECTOR 1 to 90					
FILE 1 '/frames/myxtal###.x'					this data will be reindexed by matrix above
OUTPUT FILE 'myxtals12.sca'					output combines 2 data sets w/ same index

**Reindex two H3 data sets, each with inconsistent index, and combine**

This scenario shows how to reindex two H3 data sets into R3 and combine them. This is

## Reindex four P3<sub>1</sub> data sets, each with inconsistent index, and combine

This scenario shows how to reindex four P3<sub>1</sub> data sets and combine them. This is slightly more complicated than the scenarios above. Again, the procedure is as follows: choose one of your 4 data sets as the **reference** set. This reference is only for the purposes of indexing, not for finding scale factors. Then what you do is to take each of the other 3 data sets, one by one, and scale them with the reference data set. Each time, you will have to try a different **HKL MATRIX** for the second data set until you find the one which is consistent with the reference data set. Make a note of this **HKL MATRIX**, and move on to the next data set. When you have figured out the proper **HKL MATRIX** for each of the three other data sets, then you can run this command file to reindex them and scale them all together. This command file assumes that you were unlucky enough to have all four data sets in four inconsistent indices.

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
0.04 0.04 0.04 0.04 0.04	
<b>ERROR SCALE FACTOR</b> 1.4	multiplicative factor applied to input $\sigma$
[ <b>REJECTION PROBABILITY</b> 1.0e-4]	don't reject until you know all data sets are in a consistent index
[ <b>WRITE REJECTION FILE</b> 0.9]	
[@reject]	
<b>RESOLUTION</b> 40 3.3	
<b>SPACE GROUP</b> P31	
<b>REFERENCE BATCH</b> 201	
<b>ADD PARTIALS</b> 1 to 60 101 to 160	
201 to 260 301 to 360	
<b>POSTREFINE</b> 10	good idea to <b>POSTREFINE</b> because this will affect partials
<b>FIT</b> crystal a* 1 to 360	
<b>FIT</b> crystal c* 1 to 360	
<b>FIT</b> batch rotx 1 to 360	
<b>FIT</b> batch roty 1 to 360	
<b>FIT</b> batch mosaicity 1 to 360	
<b>FORMAT</b> denzo_ip	
<b>HKL MATRIX</b> -1 0 0	reindexing first matrix: h ► - h
0 -1 0	k ► - k
0 0 1	
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 1 '/frames/myxtal1###.x'	this will be reindexed by matrix 1 above
<b>HKL MATRIX</b> 0 -1 0	data set #2 to be scaled with reindexed data
-1 0 0	
0 0 -1	
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 101 '/frames/myxtal2###.x'	this data will be reindexed by matrix 2 above
<b>HKL MATRIX</b> 0 1 0	data set #3 to be scaled with reindexed data
1 0 0	
0 0 -1	
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 201 '/frames/myxtal3###.x'	this data will be reindexed by matrix 3 above
<b>HKL MATRIX</b> 1 0 0	original unit matrix
0 1 0	
0 0 1	
<b>SECTOR</b> 1 to 60	
<b>FILE</b> 301 '/frames/myxtal4###.x'	this is the reference index data set
<b>OUTPUT FILE</b> 'myxtals12.sca'	

## Scenario 6: Comparing two data sets from Scalepack

[scalenatder.com](http://scalenatder.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	Set to zero because native & derivative have already been assigned proper errors.
<b>[ERROR SCALE FACTOR</b> 1.0]	don't need here because data already scaled
<b>SPACE GROUP</b> I422	<b>SPACE GROUP</b> of native & derivative (optional)
<b>FORMAT</b> Scalepack	data has been processed already and is in Scalepack <b>FORMAT</b>
<b>FILE</b> 1 native.sca	
<b>REFERENCE BATCH</b> 1	<b>REFERENCE BATCH</b> in scaling procedure. We will compare derivative to native, so native is reference
<b>[***Derivative***]</b>	
<b>RESOLUTION</b> 2.8	use highest resolution from the worse data set
<b>[HKL MATRIX</b> 0 0 1 0 -1 0 1 0 0	only reindex if native and derivative have different indices
<b>FILE</b> 2 'NewHg.sca'	derivative data set
<b>OUTPUT FILE</b> 'NL:junk.sca'	not interested in .sca file, only the log
print total chi2 1000 print single chi2 1000	these are set super high so that basically no outliers are printed to log file

You would use this scenario if you wanted to see how well a native and derivative data set scaled together. It is different from Scenario 3 because both data sets are complete and are being read in as .sca files. This is a more accurate way of identifying your derivatives but it requires enough data from the derivative so that it can be internally scaled.



## Scenario 7: Compare $I+$ and $I-$ : detection of anomalous signal

This involves a 2-step scaling procedure. In the first step the data are scaled and postrefined along the lines of [Scenario 1](#) with one exception, noted below, namely that the **ANOMALOUS** flag has been set. This tells the program to output a .sca file where the  $I+$  and  $I-$  reflections are separate. Then, in the second step, a clever trick is used: the .sca file is read in, but the **ANOMALOUS** flag is turned off. In this case Scalepack will treat the  $I+$  data and the  $I-$  data as two separate measurements within a data set and the statistics that result from merging the two will reflect the differences between the  $I+$  reflections and the  $I-$  reflections. Notice that there is no need to go through a lot of jiffies which separate the  $I+$  and  $I-$  data and reformat them, etc.

### Step 1 [scaleano.com](http://scaleano.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
0.04 0.04 0.04 0.04 0.04 0.04	
<b>ERROR SCALE FACTOR</b> 1.3	multiplicative factor applied to input $\sigma$
<b>SPACE GROUP</b> P212121	
<b>REFERENCE BATCH</b> 1	
<b>ADD PARTIALS</b> 1 to 180	
<b>POSTREFINE</b> 10	
<b>FIT</b> crystal a* 1 to 180	
<b>FIT</b> crystal b* 1 to 180	
<b>FIT</b> crystal c* 1 to 180	
<b>FIT</b> batch rotx 1 to 180	
<b>FIT</b> batch roty 1 to 180	
<b>FIT</b> crystal mosaicity 1 to 180	
<b>REJECTION PROBABILITY</b> 0.0001	expected fraction of outliers in the data; if your
<b>WRITE REJECTION FILE</b> 0.9	

**Step 2** [scalecompareano.com](http://scalecompareano.com)

```

NUMBER OF ZONES 10
ESTIMATED ERROR 0.0 0.0 these are zero, because we have already scaled the data
0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0
ERROR SCALE FACTOR 1.0
NUMBER OF ITERATIONS 0 no scaling done, we aren't interested in .sca, only stats in log
OUTPUT FILE 'NL:junk.sca' we will be reading in a .sca file from step 1 above
FORMAT Scalepack
FILE 1 'mydatano.sca'

```

Step 2 reads in the .sca file, which contains both I+ and I-. Obviously, for centric reflections there is no I-, so the merging statistics will only reflect the non-centric reflections. You can tell what percentage of your data is being used to calculate the merging statistics by examining the redundancy table near the end of the log file. Under the column of redundancy > 2 you will find out what percentage of the data is being compared. Since you only have I+ and I- you will never have a redundancy of more than 2.

The presence of an anomalous signal is detected by examining the  $\chi^2$  values in the log file from Step 2. Assuming that the errors in the Step 1 cycles were reasonable, a  $\chi^2 \approx 1$  in the log file of Step 2 means that there is no useful anomalous signal in your data. On the other hand, if  $\chi^2 > 1$  in Step 2, this may indicate the presence of an anomalous signal. Of course, it may also indicate absorption problems, detector problems, etc. But at the very least you can examine the  $\chi^2$  as a function of resolution and decide where to cut off your resolution for the purposes of calculating an anomalous difference Patterson (use the shells where  $\chi^2 \geq 2$ ). This whole analysis assumes that the error model is reasonable and gives you a  $\chi^2$  close to 1 when the **SCALE ANOMALOUS** flag is used.

🔴 You can only use **SCALE ANOMALOUS** when you have enough redundancy to treat the F+ and F- completely independently.

## Scenario 8: *Fitting polarization*

```

NUMBER OF ZONES 10                                number of resolution shells for statistics
ESTIMATED ERROR 0.04 0.04 0.04 0.04 0.04 0.04    ESTIMATED ERROR for each resolution shell
0.04 0.04 0.04 0.04 0.04 0.04
ERROR SCALE FACTOR 1.3                            multiplicative factor applied to input  $\sigma$ 
REJECTION PROBABILITY 0.0001
SPACE GROUP P422
OUTPUT FILE mydata.sca
REFERENCE BATCH 1
ADD PARTIALS 1 to 90
@reject
[WRITE REJECTION FILE 0.9]                        commented out here because we do not want to increase the
                                                  reject list
FORMAT denzo_ip
POLARIZATION denzo 0.9 corrected POLARIZATION correction is made here.
0.8
SECTOR 1 to 90
FILE 1 '/frames/Denzo/myxtal###.x'

```

Polarization correction testing is done in two stages. First, you scale the data as in Scenario 1, writing a rejection file. Then, second, you run the command file above, which changes the **POLARIZATION**. The log file is then examined to see if the  $\chi^2$  crystal have decreased, compared to the previous round when no polarization correction was done. If the  $\chi^2$  have improved (a significant improvement may be as low as 0.01), then you know that this is a better value for the polarization. If there is no change in the  $\chi^2$ , then you cannot establish whether the polarization was right or wrong. This requires a tetragonal crystal, trigonal crystal, or cubic crystal, exposed in the proper orientation. It is mostly for synchrotron beamline staff, not ordinary users.

## Scenario 9: High resolution + low resolution data sets from one crystal

This scenario is useful if you have made multiple passes of data collection from the same crystal, where one of the passes had short exposures to keep the low resolution reflections from saturating, and the other pass was a longer exposure to get the high resolution data. In the high resolution, long exposure pass, the low resolution reflections are mostly saturated. Note the more complicated batch numbering scheme. This is often typical of data collected at synchrotrons where there are gaps in the file numbering for various reasons. This is particularly applicable to CCD detectors at CHESS.

[scalehilo.com](http://scalehilo.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each res. shell. Set to zero because native has already been scaled
<b>ERROR SCALE FACTOR</b> 1.5	good for CCD at CHESS
<b>REJECTION PROBABILITY</b> 1. e-4	expected fraction of outliers in the data
<b>WRITE REJECTION FILE</b> 0.5	
<b>SPACE GROUP</b> P21	
<b>UNIT CELL</b> 50 93 95 90 89.9 90	
<b>RESOLUTION</b> 1.4	one res. range for both high & low res.
<b>REFERENCE BATCH</b> 1	use the first frame as reference for scaling
@reject	reject outliers from previous rounds
<b>POSTREFINE</b> 20	20 cycles of postrefinement
<b>FIT</b> crystal a* 0 to 1039	
<b>FIT</b> crystal b* 0 to 1039	
<b>FIT</b> crystal c* 0 to 1039	
<b>FIT</b> crystal beta* 0 to 1039	
<b>FIT</b> batch rotx 0 to 1039	
<b>FIT</b> batch roty 0 to 1039	
<b>FIT</b> batch mosaicity 0 to 1039	
<b>ADD PARTIALS</b> 0 to 180 184 to 216 217 to 238 1000 to 1039	note that the consecutive sequences must be declared here, otherwise program will try to <b>ADD PARTIALS</b> from non-consecutive frames
<b>FORMAT</b> denzo_ip	
<b>IGNORE OVERLOADS</b>	
<b>SECTOR</b> 0 to 35	
<b>FILE</b> 0 'my_crystal_hi###.x'	
<b>SECTOR</b> 39 to 183	
<b>FILE</b> 36 'my_crystal_hi###.x'	
<b>SECTOR</b> 184 to 238	
<b>FILE</b> 184 'my_crystal_hi###.x'	
<b>INCLUDE OVERLOADS</b>	
<b>SECTOR</b> 244 to 283	
<b>FILE</b> 1000 'my_crystal_lo###.x'	
<b>OUTPUT FILE</b>	
'my_crystalhilo.sca'	

### High Resolution Data

for the high res. data (long exposure)

the .x files not numbered consecutively. Note renumbering of sector 39.x to batch 36

frame 39.x is next **SECTOR** after 35.x because 36, 37, 38 were calibration frames

there was change in crystal orientation at 184.x.

Note that now we're even with batch numbering

### Low Resolution Data

now we go to low exposure low resolution data

we put low resolution sectors in batches starting

with 1000, just to keep it straight

## Scenario 10: Thin or few frames (*SCALE RESTRAIN* and *B RESTRAIN*)

Occasionally you may try to scale data which was collected in very thin sectors or in sectors whose width is less than the mosaicity or from data which consists of only a few frames. In these cases, there may not be enough intersections between frames to get accurate scale and B factors. Indeed, what you may see is both scale and B factors ranging all over the place. If things get really bad, the program may crash due to floating point arithmetic exceptions when taking the exponent of unreasonable B factors. This does not mean that the data is unusable, it simply means that the scale and B factors must be restrained. The keywords **SCALE RESTRAIN** and **B RESTRAIN** are used for this purpose, and an example of their use is shown below. Note that the restraints only apply to the frames over which you are adding up partially recorded reflections, so you must have the **ADD PARTIALS** keyword in the command file for this to make any sense.

[scalethins.com](http://scalethins.com)

<b>NUMBER OF ZONES</b> 10	number of resolution shells for statistics
<b>ESTIMATED ERROR</b> 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04 0.04	<b>ESTIMATED ERROR</b> for each resolution shell
<b>ERROR SCALE FACTOR</b> 1.3	multiplicative factor applied to input $\sigma$
<b>REJECTION PROBABILITY</b> 0.0001 [@reject]	expected fraction of outliers in the data [read in rejected reflections]
<b>WRITE REJECTION FILE</b> 0.9	obs. with greater than 90% chance of being outliers are written to the 'reject' file
<b>SPACE GROUP</b> P21	
<b>OUTPUT FILE</b> mydata.sca	<b>OUTPUT FILE</b> containing h, k, l, I, $\sigma_1$
<b>REFERENCE BATCH</b> 1	<b>REFERENCE BATCH</b> in scaling procedure.
<b>SCALE RESTRAIN</b> 0.05	scale and b factors between adjacent frames cannot vary by more than these values
<b>B RESTRAIN</b> 0.1	
<b>ADD PARTIALS</b> 1 to 90	must have this for restraints to work number of cycles of postrefinement
<b>POSTREFINE</b> 10	
<b>FIT</b> crystal a* 1 to 90	
<b>FIT</b> crystal b* 1 to 90	refine <b>UNIT CELL</b> parameters based on all data from batches 1 through 90
<b>FIT</b> crystal c* 1 to 90	
<b>FIT</b> crystal beta* 1 to 90	► Space group dependent!
<b>FIT</b> batch rotx 1 to 90	crystal orientation of spindle axis angle will be refined for each batch (frame) separately.
<b>FIT</b> batch roty 1 to 90	
<b>FIT</b> crystal mosaicity 1 to 90	one mosaicity value will be refined
<b>FORMAT</b> denzo_ip	<b>FORMAT</b> of the input intensity data
<b>SECTOR</b> 1 to 90	these values will be substituted in for the ###. Batch numbers
<b>FILE</b> 1 '/frames/Denzo/myxtal###.x'	will start at 1 for file myxtal001.x and increase to 90 for file myxtal090.x

❖ It is not correct to try to "bin" the individual frames into larger batches to try to overcome the problem of few intersections between frames. This is because you then lose the ability to **ADD PARTIALS** between the new "bins." You can, however, overcome the problem in another way, by including instead the statement **NUMBER OF ITERATIONS** 0. In this case, the data will not be scaled, but simply merged. Obviously, this has its drawbacks.

## Scenario 11: Space Group Identification

Scalepack can be used to determine the space group of your crystal. What follows is a description of how you would continue from the lattice type given by                      to determine your space group. This whole analysis, of course, only applies to enantiomorphic compounds, like proteins. It does not apply to small molecules, necessarily, which may crystallize in centric space groups. If you expect a centric space group, you use any space group which is a subgroup of the Laue category to which your crystal belongs. You also need enough data for this analysis to work so that you can see systematic absences.

To determine your space group, follow these steps:

1. Determine the lattice type in                      .
2. Scale by the primary space group in Scalepack. The primary space groups are the first Laue groups in each Bravais lattice class in the table which follows. In the absence of lattice pseudosymmetries (e.g. monoclinic with  $\beta \approx 90^\circ$ ) the primary space group will not incorrectly relate symmetry related reflections. Note the  $\chi^2$  statistics.. Now try a higher symmetry Laue group (next down the list) and repeat the scaling, keeping everything else the same. If the  $\chi^2$  is about the same, then you know that this is OK, and you can continue. If the  $\chi^2$  are much worse, then you know that this is the wrong Laue type, and the previous choice was your Laue class. The exception is primitive hexagonal, where you should try  $P6_1$  after failing  $P3_121$  and  $P3_112$ .
3. Examine the bottom of the log file for the list of candidate systematic absences. If this was the correct space group, all of these reflections should be absent and their values very small. Compare this list with the listing of reflection conditions by each of the candidate space groups. The set of absences seen in your data which corresponds to the absences characteristic of the listed space groups identifies your space group or pair of space groups. Note that you cannot do any better than this (i.e. get the handedness of screw axes) without phase information.
4. If it turns out that your space group is orthorhombic and contains one or two screw axes, you may need to reindex to align the screw axes with the standard definition. If you have one screw axis, your space group is  $P222_1$ , with the screw axis along *c*. If you have two screw axes, then your space group is  $P2_12_12$ , with the screw axes along *a* and *b*. If the                      indexing is not the same as these, then you should reindex using one of the two transformations described above in Scenario 5 so that you are following the standard conventions.
5. So far, this is the way to index according to the conventions of the International Tables. If you prefer to use a private convention, you may have to work out own transformations. One such transformation has been provide in the case of space groups  $P2$  and  $P2_1$ .

Bravais Lattice	Space Groups used first to id Laue group	Candidates	Reflection Conditions along screw axes
Primitive Cubic	P213  P4132	195 P23	
		198 P213	(2n,0,0)
		207 P432	
		208 P4232	(2n,0,0)
		212 P4332	(4n,0,0)*
		213 P4132	(4n,0,0)*
I Centered Cubic	I213  I4132	197 I23	*
		199 I213	*
		211 I432	
		214 I4132	(4n,0,0)
F Centered Cubic	F23 F4132	196 F23	
		209 F432	
		210 F4132	(2n,0,0)
Primitive Rhombohedral	R3 R32	146 R3	
		155 R32	
Primitive Hexagonal	P31  P3112  P3121  P61  P6122	143 P3	
		144 P31	(0,0,3n)*
		145 P32	(0,0,3n)*
		149 P312	
		151 P3112	(0,0,3n)*
		153 P3212	(0,0,3n)*
		150 P321	
		152 P3121	(0,0,3n)*
		154 P3221	(0,0,3n)*
		168 P6	
		169 P61	(0,0,6n)*
		170 P65	(0,0,6n)*
		171 P62	(0,0,3n)**
		172 P64	(0,0,3n)**
		173 P63	(0,0,2n)
		177 P622	
		178 P6122	(0,0,6n)*
		179 P6522	(0,0,6n)*
		180 P6222	(0,0,3n)**
		181 P6422	(0,0,3n)**
		182 P6322	(0,0,2n)

Primitive Tetragonal	P41	75	P4	
		76	P41	(0,0,4n)*
		77	P42	(0,0,2n)
		78	P43	(0,0,4n)*
	P41212	89	P422	
		90	P4212	(0,2n,0)
		91	P4122	(0,0,4n)*
		95	P4322	(0,0,4n)*
		93	P4222	(0,0,2n)
		94	P42212	(0,0,2n),(0,2n,0)
I Centered Tetragonal	I41	79	I4	
		80	I41	(0,0,4n)
	I4122	97	I422	
		98	I4122	(0,0,4n)
Primitive Orthorhombic	P212121	16	P222	
		17	P2221	(0,0,2n)
		18	P21212	(2n,0,0),(0,2n,0)
		19	P212121	(2n,0,0),(0,2n,0), (0,0,2n)
C Centered Orthorhombic	C2221	20	C2221	(0,0,2n)
		21	C222	
I Centered Orthorhombic	I212121	23	I222	*
		24	I212121	*
F Centered Orthorhombic	F222	22	F222	
Primitive Monoclinic	P21	3	P2	
		4	P21	(0,0,2n)
C Centered Monoclinic	C2	5	C2	
Primitive Triclinic	P1	1	P1	

➡ Note that for the pairs of similar candidate space groups followed by the \* (or \*\*) symbol, scaling and merging of diffraction intensities cannot resolve which member of the possible pair of space groups your crystal form belongs to.





# Alphabetical Scalepack Keyword Descriptions

## ADD

Increments batch numbers by a constant to every batch from this point on until another add command is read. Useful to make unique batch numbers from two or more files which have the same batch numbers inside. For example, **Denzo\_york1** format embeds the batch number in the .x file.

---

**ADD** value

nothing added

**ADD** 1000

This will add 1000 to each batch number.

---

## ADD PARTIALS

Tells the program to add partially recorded reflections among consecutive batches, even if the batches do not have consecutive numbering. **Essentially obligatory.**

---

all

partials are added over all consecutive batches

**ADD PARTIALS** start1 to end1 start2 to end2 ... etc

do not **ADD PARTIALS**

**ADD PARTIALS** 0 to 49 51 to 99 100 to 149

---



Be sure that ranges of numbers do not overlap.

## ANOMALOUS

Flag for keeping Bijovets (I+ and I-) separate in output file. If the **ANOMALOUS** flag is on, anomalous pairs are considered equivalent when calculating scale and B factors and when computing statistics, but are merged separately and output as I+ and I- for each reflection.

---

**ANOMALOUS**

not turned on (I+ and I- are combined)

**ANOMALOUS**

---

**B RESTRAIN**

Can be used to restrain B factor differences from consecutive films or batches. The value which follows the flag represents the amount in Å<sup>2</sup> you will allow the B factors to differ from consecutive frames or batches. See also the keyword **SCALE RESTRAIN**.

---

**B RESTRAIN** *value\_(Å<sup>2</sup>)*

not turned on

**B RESTRAIN** 0.5

---

**BACKGROUND CORRECTION**

The **BACKGROUND CORRECTION** command specifies amount per frame to increase background. Corrects for errors in SDMS (Hamlin) data integration in the software distributed in the late 1980's and early 1990's. The value given after the keyword increases the background by #counts/frame. Valid **for archive file format only**.

---

**BACKGROUND CORRECTION** *value\_(#counts/frame)*


not turned on

**BACKGROUND CORRECTION** 15

---

**DEFAULT B FACTOR**

Overall B used only in the absence of an **INITIAL B FACTOR**. You can apply a higher value to lower your R<sub>merge</sub>. Does not affect the quality of the data.

 Applies only to data read-in after this command.

---

**DEFAULT B FACTOR** *value*

0

**DEFAULT B FACTOR** 5

---

**DEFAULT SCALE**

Overall scale factor used in the absence of an initial scale factor. This is useful if the data are too strong, which is sometimes the case with small molecules. It will reduce the output intensities by the factor entered.


---

**DEFAULT SCALE** *value*

1

**DEFAULT SCALE** 10

---

 Useful to reduce overall scale of the data set. If the numbers in the output file are too large, **DEFAULT SCALE** 10 will reduce them 10-fold.

**DO NOT REJECT OUTLIERS**

Turns off the reject outliers flag.

---

DO NOT REJECT OUTLIERS

DO NOT REJECT OUTLIERS

OUTLIERS

REJECT

---

**END FIT**

Provides a logical end for the fit command. At the moment this has **no effect**, but may be implemented in future versions of the program.

**ERROR SCALE FACTOR**

This is a single multiplicative factor which is applied to the input  $\sigma_I$ . This should be adjusted so the normal  $\chi^2$  (goodness of fit) value that is printed in the final table of the output comes close to 1. By default the input errors are used (**ERROR SCALE FACTOR** = 1). It applies to the data which are read after this keyword, so you can apply different error scale factor to subsequent batches by repeating this input with different values.

---

**ERROR SCALE FACTOR** *value*

**ERROR SCALE FACTOR** 1

**ERROR SCALE FACTOR** 1.3

---

**ESTIMATED ERROR**

Estimate of the systematic error for each of the resolution shells. **There must be exactly the same number of error estimates here as there are NUMBER OF ZONES.** So if you have 10 zones, you need 10 numbers following the keyword estimated error - one for each zone.



The error estimates do not all have to be the same. The estimated error applies to the data which are read **after** this keyword, so you can apply different error scale factor to subsequent batches by repeating this input with different values. This is an important point if you enter data from a previous Scalepack output that does not need its  $\sigma$  to be increased.

The error estimates should be approximately equal to the R-factor in the table at the end of the output for resolution shells **where statistical errors are small** namely the earlier resolution shells where the data is strong. This is a crude estimate of the systematic error, to be multiplied by I, and is usually invariant with resolution. Default = 0.06 (i.e. 6%) for all zones.

---

**ESTIMATED ERROR** *value1 value2 value3 value4 ...etc*

each value set to 0.06

**ESTIMATED ERROR** 0.02 0.03 0.03 0.03 0.04 0.04 (6 zones)

---

**EXTEND PARTIALS / DO NOT EXTEND PARTIALS**

Some partially recorded reflections may be predicted by \_\_\_\_\_ or \_\_\_\_\_ to start or end their Bragg condition in between consecutive frames due to small variations in crystal

orientation from frame to frame. For these reflections only, there are two choices of defining where the reflection started (or ended): including the extra frame, or not. **EXTEND PARTIALS** tells to include this extra frame. It only affects a very small fraction of the reflections. Opposite of **DO NOT EXTEND PARTIALS**.

---

#### EXTEND PARTIALS

this is the default

---

## FILE

This specifies the files read by . The input has two components. The first is a number. The second is a file name, which usually contains wildcard characters (###) that are incremented by the **SECTOR** command. The number which follows **FILE** specifies the **starting batch number**. A batch, previously called a **film**, can be as small as a single .x file (or the equivalent). It can be a group of .x files, even an entire data set. The most frequent content of a batch is a single .x file.

This conversion of files into batches is particularly useful if you want to scale more than one data set together. For example, let's say you want to scale 10 oscillation frames (numbered 1 through 10) from first data set with 37 oscillation frames (numbered 1 through 37) from other one. The **FILE** statement will take each of the individual .x files and assign them a batch number. Thus, you would enter something like this: **FILE 1 'setone###.x'** and **FILE 101 'settwo###.x'** (see Scenario 5d). Thus, batch numbers 1 - 10 will correspond to files setone001.x, setone002.x, etc. (assuming you used the **SECTOR 1 to 10** command above **FILE** so that the wildcards would be substituted with numbers). The batch numbers 101 - 137 will correspond to files settwo001.x, settwo002.x, etc.

➡ the ### is replaced by the sector argument, not by the batch number.

---

**FILE** value 'filename' (see also note below)

None

**FILE 101 '/frames/scale/lysoz###.x'**

---

➡ The **FILE** must come after **FORMAT** because the syntax depends on which input format is being read. The **FILE** must not be followed by number in the case of archive, denzo\_york, and denzo\_york1 formats, where the number after the word **FILE** is not given because the batch numbers are already stored in the file. If you want to change the batch numbers in these file formats, see the **ADD** command described above.

## FIT

Tells the program what parameters to fit in **postrefinement** and specifies the group of files over which the fitting is to be performed. Postrefinement fit can be applied to an entire set of batches (one batch being the entire set of frames, for example) using the modifier *crystal*, or to each individual input file using the modifier *film* or *batch*.

---

crystal	specifies that the fit operation is over the <b>entire set</b> of frames specified by a range and restrains the fit parameter to have exactly the same value over this range.
batch	specifies that the fit operation is performed on <b>each member of the set</b> of frames specified by a range.

film                    alias for batch

a\*, b\*, c\*            unit cell lengths. Values returned are real space

alpha\*                unit cell angles. Values returned are real space.

beta\*

gamma\*

rotx                   crystal orientation parameters deduced in

roty

rotz

wavelength          incident wavelength. *Untested.*

mosaicity            mosaicity as defined in            , in degrees

fit modifier parameter1 filmnumber to filmnumber

the default is that nothing is fit unless specified

**FIT** crystal roty 1 to 137

**FIT** crystal mosaicity 1 to 5 7 to 10 102 to 104 to 137

**FIT** batch rotx 1 to 137

---

**Most mistakes** in                    can be attributed to errors in **FIT** commands because the program is very sensitive to mistakes in the batch numbers. If you input non-existent batch numbers or define overlapping ranges (e.g. 1 to 10 5 to 20), the program is likely to fail in a strange way. If you specify a range of numbers, the program will only use the batch numbers that exist within the range. For example, if your batch numbers go from 1 to 40 and 70 to 90, you can get away with saying, say, **FIT** batch (*parameter*) 1 to 90, which is the same as **FIT** batch (*parameter*) 1 to 40 70 to 90. For **FIT** crystal, these two inputs are not equivalent. In the case of **FIT** crystal (*parameter*) 1 to 90, one value will be fit for all batches. In the case of **FIT** crystal (*parameter*) 1 to 40 70 to 90, two values will be fit, one for each range. Note that different parameters may be fit over different ranges and either over batch or crystal. You can also mix batch and crystal for the same parameters.

**Do not fit unit cell parameters restrained by space group symmetry.** For example, if you have space group P3, you must not fit b\*.

**Do not FIT** batch rotx because this parameter is very poorly defined by the intensities of observed partial reflections. This is another very common mistake.

Unless you know what you are doing, **do not FIT** crystal rotx roty rotz, because if the spindle is even slightly misaligned, the assumption that there is only one crystal orientation parameter for a large sweep of data will force incorrect restraints on the crystal orientation refinement.

About **fitting rotations**: Changes in rotations, like crystal rotx roty and rotz, are expressed as small rotations, call them  $\Delta$ , about the laboratory frame of reference. These  $\Delta$ :  $\Delta_x$ ,  $\Delta_y$ , and  $\Delta_z$ , are used because to a first approximation they commute with one another (commute means that the order in which they are applied is irrelevant). This in turn is because these  $\Delta$  are small, typically less than one degree. The crystal or cassette rotations,  $R_x$ ,  $R_y$ , and  $R_z$ , on the other hand, do not commute with

rotations, what is actually happening is that the  $\Delta$  are being refined. After each refinement cycle, the  $\Delta$  are converted into changes in  $R_x$ ,  $R_y$ , and  $R_z$  by a (complicated) algebraic relation. Those of you with sharp eyes will have noticed in [Scalepack](#) that the shifts reported by the program when fitting the crystal rotations do not correspond to the changes in the rot values. This is because the shifts reported are the  $\Delta$ , not the changes in  $\text{rot}_x$ ,  $\text{rot}_y$ , and  $\text{rot}_z$ . The other reason for fitting  $\Delta$  rotations the way they are defined in [Scalepack](#) is to make them have a more intuitive correlation with the other parameters. Otherwise, changes in **crystal rotz** would not correlate with **cassette rotz**. This is of more importance in [Scalepack](#), where only  $\text{rot}_x$  and  $\text{rot}_y$ , and not  $\text{rot}_z$ , are refined.

➡ About **fitting unit cell** parameters: In both [Scalepack](#) and [Scalepack](#) the unit cell is fitted in **reciprocal space**, not real space. This means that for a non-orthogonal space group, refining the value of  $\mathbf{a}^*$  may end up changing the values of  $\mathbf{b}$  and  $\mathbf{c}$ , even though  $\mathbf{b}^*$  and  $\mathbf{c}^*$  remain the same. The same is true for the angles: fitting  $\alpha^*$  may end up changing  $\beta$  and  $\gamma$ , even though  $\beta^*$  and  $\gamma^*$  remain the same. So what you may notice sometime if you are not careful is that when you ask the program to fit crystal  $\mathbf{a}^*$ , but not  $\mathbf{b}^*$ ,  $\mathbf{c}^*$ , and the angles, then  $\mathbf{a}$  will not be a constant.  $\mathbf{a}^*$  will be constant, but when converted back to  $\mathbf{a}$  when the other unit cell parameters have not been changes,  $\mathbf{a}$  will not. The moral: **when fitting unit cells, fit all the relevant parameters.**

## **FIT B**

This flag tells the program to refine B factors of every batch from the very first cycles of refinement. This is in contrast to the default procedure, where the B factors are fit only after the convergence of the scaling. In the default procedure, if scaling does not converge in 20 (default) cycles of refinement, B factors will be not be fitted. The **FIT B** command can override this. Not to be confused with the postrefine **FIT B\*** command described above. **You cannot postrefine B (temperature) factors.**

---

**FIT B**this flag is turned **off**

---

**FIT B**

---

## **FIX B**

This flag tells the program not to fit B factors at all. Usually it is combined with the input of the B factors you want to apply but do not wish to refine anymore or it is used for frozen crystals where you do not expect significant decay.

---

**FIX B**turned **off**; B factors are fit after the scale factor refinement converges.

---

**FIX B**

---

## **FIXED WINDOW**

For **Hamlin** archive files a fixed window of 3, 5, 7, or 9 frames or the original Hamlin-determined window of frames may be used for summing a reflection. **Valid for archive file format only.**

---

**FIXED WINDOW** value

---

---

uses the **Hamlin** definition of the window

**FIXED WINDOW 7**

---

## FORMAT

This keyword specifies the format of the input and intensity data. Input data can come from any of nine types of files. This program requires this keyword to properly read the files.

---

denzo\_ip

denzo\_york1

scalepack

rigaku raxis *R-Axis*

xds profit.hkl

madnes procor intout

madnes procor ascii

xengen urf

---


Archive

*Hamlin*

---

**FORMAT** modifier

denzo\_ip

**FORMAT** denzo\_ip 

---

## FRAME WIDTH

**Only for URF files** created by . Other formats do not need this specification because can read this information off the file header. The oscillation range for each frame.

---

**FRAME WIDTH** value

there is no default, this input is required for this format

**FRAME WIDTH** 0.2

---

## HKL MATRIX


Matrix for re-indexing. Default = unit matrix. This matrix is applied to the 's as they are read in (applies to data read in after this command):

$$h' = (1)*h + (2)*k + (3)*l$$

$$k' = (4)*h + (5)*k + (6)*l$$

$$l' = (7)*h + (8)*k + (9)*l$$

Input is in the order (1....9). **HKL MATRIX** works with postrefinement.

 program will not accept a matrix which has a negative determinant



---

```
HKL MATRIX value11 value12 value13 value21 value22 value23 ...etc
```

the default is the unit matrix

```
HKL MATRIX 0 1 0 1 0 0 0 transforms into , has the effect of switching  
0 -1 and . -1 is to keep the determinant positive.
```

---

## HKL SCALE

Divides each  $a$ ,  $b$ , and  $c$  by the input values. Useful for reducing the unit cell volume, particularly after matrix transformation. Rarely needed. Example, index data originally in C222 and want you data in, say, P3, you would apply **HKL SCALE** of 2 2 1, and **HKL MATRIX** of 1 1 0 1 -1 0 0 0 -1 so that the new indices would have values of  $h + k/2$ ,  $h - k/2$ , and  $-l$ .

---

```
HKL SCALE value1 value2 value3
```

```
1 1 1
```

```
HKL SCALE 2 2 1
```

---

## HKL SHIFT

Used to make a quick test of **misindexing**. Adds the specified integer vector to each original  $a$ ,  $b$ , and  $c$ . If you are successful in using this, congratulations, you now have to reprocess your data in  $h$ ,  $k$ , and  $l$  with the correct indexing.

---

```
HKL SHIFT integer integer integer
```

```
0 0 0
```

```
HKL SHIFT 0 1 0
```

---

## IGNORE OVERLOADS

Opposite of **INCLUDE OVERLOADS**. This is **not the default**. Useful if you collect data at low and high exposures and is useful to ignore the saturated reflections at the high exposures. Applies to data read in after this command.

## INCLUDE OVERLOADS

**INCLUDE OVERLOADS** is flag for whether fitted profiles with some pixels missing (typically due to overload) should be included in the scaling. Affects only image plate output files (formats denzo\_ip, denzo\_york1). Note that for summed profiles this does not apply because only profile fitting can estimate the value of the overloaded pixels.

---

```
INCLUDE OVERLOADS
```

```
INCLUDE OVERLOADS
```

---

## INITIAL B FACTOR

Table of starting **B factors**, one per batch, beginning on the record after the title. After running the program once, the output table of B factors can be cut out and pasted in here, for example, when you have set the **FIX B** flag and are not refining B values anymore.

---

**INITIAL B FACTOR** batch no. b value batch no. b value ...

value of the default B factor, which is defaulted to 0

**INITIAL B FACTOR** 1 0.0 2 0.1 3 0.1 4 0.1 5 -0.2 ...etc

---

## INITIAL SCALE

Table of **starting scale factors**, one per batch, beginning on the record after the title. After one run of the program, the output table of scale factors can be cut out and pasted in here. This table is not required. If it is not included in the control input, then the **DEFAULT SCALE** is used (this is 1 unless otherwise specified).

➡ If the initial scale factor is set to zero, that frame is ignored in the scaling and refinement.

---

batch no. scale value batch no. scale value ...

value of the default scale factor, which is defaulted to 1

**INITIAL SCALE FACTOR** 1 1.0 2 0.9 3 0.0 4 0.85 5 1.1 ...etc.

---

## INPUT / @

Redirection. Tells the program to read the file which follows the keyword. Same as in .

---

@filename

no redirection

@reject

---

## MERGE

Flag that tells the program to merge (combine, average) reflections with the same unique index. This is the default.

---

**MERGE**

this is the default and need not be specified

---

## MOSAICITY

Allows you to input the value of the mosaicity of the data set. This is normally read from the header of the .x file.

---

**MOSAICITY** value

read from the header of the .x file

**MOSAICITY** 0.5

---

## NO ANOMALOUS

Opposite of the keyword **ANOMALOUS**. Causes I+ and I- to be merged. This is the default.

---

**NO ANOMALOUS**

---

---

this is the default

---

## NO MERGE

Flag for the output of **unmerged** (reflections with the same unique are not combined) data. Opposite of **MERGE**. This is a very handy for specialized work. This flag has subsidiary modifiers include partials, original index or as default include no partials

---

include partials		all observations, both fully and partially recorded, are included in the output. The output will consist of the unique , batch number, asymmetric unit, I, $\sigma$ , and fractionality of the reflection. There is no information about I+ and I-, although it may be possible to get this in subsequent versions of the program.
------------------	--	--

include partials	no	only fully recorded reflections and those fully recorded reflections created by the summation of partials are included in the output. Partials which cannot be summed to a fully recorded reflection are lost. The output will consist of the unique , batch number, asymmetric unit, I, and $\sigma$ . This is the <b>default</b> for <b>NO MERGE</b> . There is no information about I+ and I-.
------------------	----	---

original index		the output will also contain the original (not unique) for each reflection. This is <b>designed for MAD/local scaling</b> work. The original index modifier only works with the default <b>INCLUDE NO PARTIALS</b> . The output will consist of the original , unique , batch number, a flag (0 = centric, 1 = I+, 2 = I-), another flag (0 = reflecting above the spindle, 1 = reflecting below the spindle), the asymmetric unit of the reflection, I (scaled, Lorentz and Polarization corrected), and the $\sigma$ of I. The format is (6i4, i6, 2i2, i3, 2f8.1).
----------------	--	---

---

**NO MERGE** modifier

**NO MERGE** original index

---

➡ This command is also useful if you just want to combine all of the information contained in multiple .x files into a single file. Simply read in all of the .x files, don't do any scale or B factor refinement, and output **NO MERGE include partials**.

## NO PROFILE TEST

This is a flag which operates only on **ARCHIVE** format files. Tells that reflections with weird profiles should not be rejected. Opposite of **PROFILE TEST**.

---

**NO PROFILE TEST**

this is not the default

---

## NUMBER OF ITERATIONS

Number of **cycles for refinement** of scale and B-factors. Default value is 20 cycles. If it is set to 0, the program computes statistics, merges reflections with the same unique , and writes the output file based on the initial scale and B factors. Normally you do not have to specify this unless you want to avoid scaling.

---

**NUMBER OF ITERATIONS** integer

20 iterations

**NUMBER OF ITERATIONS** 0

---

## NUMBER OF ZONES

Number of **resolution shells** the data is divided into for the basis of calculating statistics. This input is required and must match the number of zones specified under the **ESTIMATED ERROR** keyword. Handy tip: it's nice to set up the **NUMBER OF ZONES** to equal the number of zones used by for the output of refinement statistics by shell, for when you get around to publishing your data and refinement statistics together in a paper.

---

**NUMBER OF ZONES** integer

this input is required

**NUMBER OF ZONES** 10

---

## ORIENTATION AXIS 1

Three integer vector which describes the orientation of the vertical axis of the crystal. Equivalent to and keywords **VERTICAL AXIS**. This information is usually not input by the user but subsequent versions will read it from the header of the .x files. This command does not affect scaling or postrefinement, however it does affect the values of the misorientation angles reported in the log file.

---

**ORIENTATION AXIS 1** integer integer integer

1 0 0

**ORIENTATION AXIS 1** 0 1 0

---

## ORIENTATION AXIS 2

Same as **ORIENTATION AXIS 1** except describes the spindle axis of the crystal. Equivalent to and keywords **SPI NDL E AXIS**. Otherwise the same as above.

---

**ORIENTATION AXIS 2** integer integer integer

0 0 1

**ORIENTATION AXIS 2** 1 0 0

---

## ORIGINAL WINDOW

Refers to the window size of ARCHIVE format files. Opposite of **FIXED WINDOW**.

## OUTPUT ANOMALOUS

alias for the keyword **ANOMALOUS**.

## OUTPUT FILE

Name of file for output of scaled measurements. A new file will be created if none exists, but a pre-existing file will be overwritten. Maximum of 80 characters allowed in the file name. Depending on whether **ANOMALOUS** flag is set, there are one or two sets of I and  $\sigma(I)$  per reflection. Output is , , , I,  $\sigma(I)$  (**NO ANOMALOUS** flag) or , , , I+,  $\sigma(I+)$ , I-,  $\sigma(I-)$  (**ANOMALOUS** flag) in format (3I4, 4F8.1). If the **NO MERGE** flag has been set, unmerged data are output as , , , asymmetric unit #, I,  $\sigma(I)$  in format (4I4, 2F8.1) and the **ANOMALOUS** flag has no effect. For **NO MERGE** original index the format is: original , unique , batch (film) number, a flag (0 = centric, 1 = I+, 2 = I-), another flag (0 = reflecting above the spindle, 1 = reflecting below the spindle), the asymmetric unit of the reflection, I (scaled, Lorentz and Polarization corrected), and the  $\sigma(I)$ . The format is (6i4, i6, 2i2, i3, 2f8.1). **If I, I+ or I- are larger than  $10^6$ , they are output in format I8. When you read the Scalepack output by a third party program, check, if the program uses standard fortran input f8.0, instead of non-standard f8.1.**

---

**OUTPUT FILE** name        use single quotes for safety

you must specify the output file name

**OUTPUT FILE** '/people/myname/LYSOZ.sca'

---

## POLARIZATION

This command allows you to correct a mistaken polarization/monochromator value entered when you ran . It saves you from the chore of reprocessing all of your original images once you have learned of a mistake. This is mainly for synchrotron users who collect very high resolution data.

---

**POLARIZATION** Denzo value corrected value

the polarization is read off the header of the .x file

**POLARIZATION** Denzo 0.0 corrected 0.9

---

## POSTREFINE

**Number of cycles of postrefinement** do be done. Works in conjunction with the **FIT** commands described above. If you set the number of postrefinement cycles to zero, then postrefinement will be skipped even if you have **FIT** commands in your control file.

---

**POSTREFINE** integer

postrefinement is not done unless you specify it

**POSTREFINE** 10

---

**PRINT**

This flag tells the program **to print the results** of the specified calculation **after every refinement cycle**. Useful in the case where one scales non-isomorphous data and wants to make the outlier list short by using the statement **PRINT total chi2 200 PRINT single chi2 100**.

total chi2	cutoff for printing poor reflections. Total $\chi^2$ for all measurements of a reflection. Default = 5 <sup>2</sup>
single chi2	cutoff for printing poor reflections. $\chi^2$ for any single measurement of a reflection. Default = 3.5 <sup>2</sup> .
correlation / no correlation	prints correlation matrix. Default: no correlation
solution / no solution	prints new scale and B factors. Default: no solution
shifts / no shifts	prints changes to scale and B factors.
<b>PRINT</b> modifier	
<b>PRINT</b> total chi2	
<b>PRINT</b> single chi2	
<b>PRINT</b> no correlation	
<b>PRINT</b> no solution	
<b>PRINT</b> no shifts	

New scale and B factors will always be printed if the number of fitted parameters exceeds 300.

**PROFILE TEST**

**PROFILE TEST** is a flag that can be turned on and off as desired. This is the default. Bad profiles can be used to reject a DCREDUCE summed . Valid for ARCHIVE file format only. Opposite of **NO PROFILE TEST**.

**PROFILES FITTED**

This flag tells how the profiles were treated by the indexing program (e.g. ). The two choices for profiles are fitted and summed. Valid for , \_york1, and **urf** formats. Most of the time people fit profiles, so this is the default and need not be specified in the command file.

<b>PROFILES FITTED</b>
this is the default

**PROFILES SUMMED**

Opposition of **PROFILES FITTED**. This is not the default and must be specified if true.

**RECSQ**

**Metric tensor** description of the reciprocal space unit cell. Why would you ever use this when you can specify the real space constants?

---

**RECSQ** value1 value2 ... value6

same as **UNIT CELL**

**RECSQ** 0.0001 0 0 0.0004 0 0.0009 [to describe unit cell 100 50 33 90 90 90.]

---

**REFERENCE BATCH**

Specifies which batch or film or set of batches or films will be the reference for the scaling and B refinement. The scale and B factor for these are not refined. More than one film or batch may be used as the reference. This is important only for crystals which decay during data collection. If the crystal is frozen and does not decay, then the default may be used, which is to let the eigenvalue filter define the overall scale and B factor. With a large number of batches, reliance on the eigenvalue filter is a little bit dangerous, so you should consider using a reference batch number in those cases.

This keyword is entirely equivalent to the keywords **REFERENCE BATCHES**, **REFERENCE FILM**, **AND REFERENCE FILMS**. The others exist because some people have a gramatical hangup about using the singular to describe more than one object.

---

**REFERENCE BATCH** integer integer integer ... etc.

no reference batch. **Eigenvalue filter** defines overall scale and B

**REFERENCE BATCH** 1 3 4 5

---

**REFERENCE BATCHES, FILM, FILMS**

Same as reference batch

**REJECT HKL**

Flag which tells the program to reject a list of individual       's which follow. This is useful for iterative rejection cycles, since the file containing the rejected reflections can be reread. Each record after the title contains one reflection with the variables h, k, l (original index) and film number in free format. This information is most easily edited from the printed output of a previous run, can contain the whole line. One can read it from a reject file with command @reject.

---

**REJECT HKL**

this is not the default

---

## REJECT OUTLIERS

Automatic rejection of outliers. Not yet reliably implemented. This is supposed to replace multiple rounds of running `SCALEPACK`. The idea is that when this keyword is set, `SCALEPACK` runs once, makes a `reject.dat` file, then runs again, reads the `reject.dat` file and scales the data based on the reduced set of observations.

## REJECTION PROBABILITY

Applies **Bayesian rejection of outliers**. Rejected observations are written to the `reject.dat` file (see **WRITE REJECTION FILE**). The whole `SCALEPACK` (all original `SCALEPACK` s with the same unique `SCALEPACK` ) with at least one observation with a probability of being an outlier greater than or equal to the value specified after the **WRITE REJECTION FILE** keyword (default 0.9) are written to the log file. This is an estimation on your part of how frequently you expect any observation to be an outlier. In principle, the rejection probability should be about equal to the number of outliers divided by the number of observations. A good value to use is 1/10,000, i.e. 0.0001, for normal good data. If you have a non-random signal in your background (e.g. satellite crystals, malfunctioning detector, ice rings) then you will probably have to increase the rejection probability. If you do not want to generate a reject list in the log file then omit the keyword. A comparison of  $R_{\text{linear}}$  and  $R_{\text{squared}}$  is often helpful for deciding whether to increase the rejection probability. See the discussion following `SCALEPACK` for more on this.

The rejection algorithm used in `SCALEPACK` is the most sensible and statistically sound outlier rejection algorithm. Unlike some other programs, the `SCALEPACK` outlier rejection is based on comparing differences to  $\sigma$ . So it will typically reject, say 4 or 5  $\sigma$  outliers. Some reflections with a large discrepancy from the average value of  $I$  may simply represent a lack of adequate statistics in measurement and not a "mistake" (non-random error) in measurement. If a reflection has a large discrepancy from the average and a large  $\sigma$ , its contribution to the average will be very small anyway.

Here's an example of how the outlier rejection algorithm works:

Input data (after scaling)		
Obs #	I	$\sigma$
1	10.0	0.1
2	10.1	0.1
3	9.9	0.1
4	11.0	0.1
5	20.0	10.0

Although you may think that {5} is the outlier, in fact the consistent set of observations will be {1, 2, 3, 5}, and {4} will be the outlier. Why is this? A consistent average will be 10.0 (or more exactly 10.000003), because observation 5 has a 1/1000000 lower statistical weight than the other observations, due to its large  $\sigma$ . Observation 4, on the other hand, which was measured more accurately, will be 8.7  $\sigma$  deviations from the average of the remaining observations because the expected error of the difference  $4 - \langle \{1,2,3,5\} \rangle$  is 0.115, and  $8.7 = (11.0 - 10.000003) / 0.115$ . So the observation with the largest deviation from the average is not necessarily a statistical outlier.



- ➡  $R_{\text{merge}}$  (after outlier rejection) will be very bad, 25%, even if the average is measured extremely well, with a sigma of 0.057%! This has to do with  $R_{\text{merge}}$  being an unweighted statistic, and  $I/\sigma$  being a weighted statistic. If some other program rejected observation 5 (for no good reason) and observation 4 (for good reason),  $R_{\text{merge}}$  would be 0.067%. Practice shows that most programs would reject observation 5 because many users want a good  $R_{\text{merge}}$ . And sometimes these program would not even flag the true outlier - observation 4!

---

REJECTION PROBABILITY value

0.0001

REJECTION PROBABILITY 0.0001

---

## RESOLUTION

Minimum **d-spacing** for this run. Default is the maximum resolution found in the input data. One can supply two numbers in any order and they will be minimum and maximum d-spacings.

---

RESOLUTION value [value]

highest resolution detected in .x file

RESOLUTION 10 2.2

RESOLUTION 2.2

---

## RESOLUTION STEP

Changes the number of reflections per shell for the purposes of printing out statistics. Formally, it represents the exponent of the zone volume calculation. Normally, this is 3, because the volume of a sphere goes as radius cubed, so all the statistics shells will have the same number of reflections. Changing the resolution step could be useful to prepare a table of statistics to compare with other programs which may print out the statistics differently.

---

RESOLUTION STEP value

RESOLUTION STEP 3

RESOLUTION STEP 2.5

---

## SCALE ANOMALOUS

This is the flag for keeping Bijvoets (I+ and I-) separate both in scaling and in the output file. If the **SCALE ANOMALOUS** flag is on, anomalous pairs are considered non-equivalent when calculating scale and B factors and when computing statistics, and are merged separately and output as I+ and I- for each reflection.

- ➡ This is a dangerous option because scaling may be unstable due to the reduced number of intersections between images. The danger is much larger in low symmetry space groups. **SCALE ANOMALOUS** will always reduce  $R_{\text{merge}}$ , even in the absence of an anomalous signal, because of the reduced redundancy. However,  $\chi^2$ 's will not be affected in the absence of an anomalous signal.

---

SCALE ANOMALOUS

---

---

definitely not a default

---

## SCALE RESTRAIN

Can be used **to restrain scale factor differences** from consecutive films or batches. The value which follows the flag represents the amount you will allow the scale factors to differ from consecutive films or batches. It adds a factor of  $(\text{scale1} - \text{scale2})^2 / (\text{scale restrain})^2$  to the target function minimized in scaling. This only applies to batches between which you add partials. For very thin frames, this is almost obligatory. The value should roughly represent the expected relative change in scale factors between adjacent frames.

---

**SCALE RESTRAIN** value

not turned on

**SCALE RESTRAIN** 0.01 (expect 1% change between adjacent frames)

---

## SECTOR

Substitutes for the ### wildcard to specify a group of files to be read. See **FILE** keyword.

---

**SECTOR** integer to integer

no default

**SECTOR** 1 to 40

---

## SECTOR WIDTH

Sector width can be specified in degrees. A pseudofilm is a sector width's worth of data from one detector. Valid for area detector data only, default value 5 degrees.

---

**SECTOR WIDTH** value

5 degrees

**SECTOR WIDTH** 3.0

---

## SIGMA CUTOFF

Cutoff for **rejecting measurements** on input. Default = -3.0. Be careful if you increase this!

➡ What is the rationale for using  $\sigma$  cutoff -3.0 in `SIGMA CUTOFF`? Wouldn't you want to reject all negative intensities? Why shouldn't you use a  $\sigma$  cutoff 1.0 or zero? The answer is as follows: The **best estimate of I may be negative**, due to **background subtraction** and **background fluctuation**. Negative measurements typically represent random fluctuations in the detector's response to an X-ray signal. If a measurement is highly negative ( $\leq -3\sigma$ ) then it may be more likely the result of a mistake, rather than just random fluctuation. If one eliminates negative fluctuations, but not the positive ones before averaging, the result will be highly biased. In `SIGMA CUTOFF`, `SIGMA CUTOFF` is applied before averaging. If one rejects all negative intensities before averaging a number of things would happen:

- 1) The averaged intensity would always be positive;
- 2) For totally random data with redundancy 8, in a shell where there was no signal, there would be on average 4 positive measurements, with average intensity one  $\sigma$ . This is because the

negative measurements had been thrown out. So the average of the four remaining measurements would be about  $2\sigma$ ! This would look like a resolution shell with a meaningful signal!

- 3)  $R_{\text{merge}}$  would be always less than the  $R_{\text{merge}}$  with negative measurements included
- 4) A **SIGMA CUTOFF** of 1 would improve  $R_{\text{merge}}$  even more, by excluding even more valid measurements!

Why should this worry you? Exclusion of valid measurements will deteriorate the final data set. One may notice an **inverse relationship between  $R_{\text{merge}}$  and data quality** as a function of **SIGMA CUTOFF**. So much for using  $R_{\text{merge}}$  as any criterion of success.

Even the best (averaged) estimate of intensity may be negative. How to use negative I estimates in subsequent phasing and refinement steps? The author of [this paper](#) suggests the following:

- 1) You should never convert I into F
- 2) You should square  $F_{\text{calc}}$  and compare it to I. Most, but not all of the crystallography programs do not do this. That is life. In the absence of the proper treatment one can do approximations. One of them is provided by French and also by French and Wilson and implemented in the CCP4 program TRUNCATE. A very simplified and somewhat imprecise implementation of TRUNCATE is this:

if  $I > \sigma(I)$ ,  $F = \sqrt{I}$   
 if  $I < \sigma(I)$ ,  $F = \sqrt{\sigma(I)}$

---

**SIGMA CUTOFF** value

-3

**SIGMA CUTOFF** -2.5

---

## SPACE GROUP

Space group symbol from the list below. This input is required! The space group may be entered as a name (e.g. P212121) or as a number (e.g. 19, for the same space group). Most of the numbers correspond to those of the International Tables. The numbers above 230 are non-standard definitions of space groups.

1	P1	89	P422	154	P3221	207	P432
3	P2	90	P4212	155	R32	208	P4232
4	P21	91	P4122	168	P6	209	F432
5	C2	92	P41212	169	P61	210	F4132
16	P222	93	P4222	170	P65	211	I432
17	P2221	94	P42212	171	P62	212	P4332
18	P21212	95	P4322	172	P64	213	P4132
19	P212121	96	P43212	173	P63	214	I4132
20	C2221	97	I422	177	P622	303	P2C
21	C222	98	I4122	178	P6122	305	B2
22	F222	143	P3	179	P6522	318	P21221
23	I222	144	P31	180	P6222	401	C1
24	I212121	145	P32	181	P6422	403	P21C
75	P4	146	R3	182	P6322	446	H3
76	P41	149	P312	195	P23	455	H32
77	P42	150	P321	196	F23	501	I1

78	P43	151	P3112	197	I23	503	I2
79	I4	152	P3121	198	P213	505	C21
80	I41	153	P3212	199	I213		

Notes to particular space groups:

146	R3	R3 in hexagonal setting
446	H3	R3 in primitive setting
155	R32	R32 in hexagonal setting
455	H32	R32 in primitive setting
401	C1	Non-standard, but useful to make angles close to 90.
501	I1	Non-standard, but useful to make angles close to 90.
303	P2C	P2, C axis unique
403	P21C	P21, C axis unique
305	B2	like C2, B face centered, c axis unique
503	I2	Non-standard, but useful to make beta angle close to 90.

## SPINDLE AXIS

Alias for **ORIENTATION AXIS 2**.

## UNIT CELL

Real cell specified as a, b, c, alpha, beta, gamma. **UNIT CELL** is included in the header of `_IP`, `_york1`, `ARCHIVE` and `files`, otherwise it must be included.

---

```
UNIT CELL value value value angle angle
```

---

the value from the first header encountered

```
UNIT CELL 50 62 100.3 90 90 90
```

---



the postrefined value of the unit cell constants is not used nor is it output in the output file. You must get this information from the log file if you are interested in it.

## VERTICAL AXIS

Same as **ORIENTATION AXIS 1**

## WRITE BADDIES

Writes `*.xrej` files so that reflections from the reject file may be displayed by `display reject`. Not fully implemented. Will terminate the program and prevent scaling and postrefinement. Note that the command `@reject` must precede the input of the `.x` files.

## **WRITE REJECTION FILE**

Creates a file with 's to be rejected. UNIX file name: `reject`, VMS file name: `REJECT.DAT`. Reject file is created if it does not exist. If the file exists it is overwritten.

You can specify the threshold for what you consider to be a **rejectable probability**. The default is 0.9, which is fairly safe, but you may want to decrease this to, say, 0.5 on later rounds of rejection.

---

**WRITE REJECTION FILE** *value*

does not write the rejection file, but if it does, the default value is 0.9

**WRITE REJECTION FILE** 0.5

---

# HKL Q & A

This chapter presents a selection of the questions sent to the [server](#) over the past year or two. They were selected to illustrate points which may not be covered in the main portion of the [XDS](#) and [XPREP](#) manuals. Answers are by Zbyszek Otwinowski and Wlodek Minor. Spelling and grammar have been cleaned up only to a limited extent. The general topic headings are listed here:

2 $\theta$

Absorption correction factors

Auto Editing

Autoindexing

Box Definition

CCP4 Scaling

Computer problems and questions

Confusion & other

Consequences of Profile fitting errors in Denzo

Corrections

Output

Problems

Display

Example of Bad Detector Pathology

Ignore

Mosaicity

Negative Intensities

Overload

Refinement

Rhombohedral Lattices

Twinned Crystals

2θ

Q: θ CASSETTE rotx θ

The 2θ swing is described by CASSETTE rotx in , with the opposite sign to the R-axis definition for 2θ. DISTANCE is defined in as the distance between crystal and a point there beam intersects detector; it is = R-axis distance / cos(2θ). X BEAM and Y BEAM are the actual position of the beam on the detector during exposure at the current 2θ (not at 2θ =0). X BEAM and Y BEAM change with 2θ, mostly Y BEAM. If the detector is aligned perfectly X BEAM does not change. The Y BEAM, in extreme cases, may be outside the detector. The correct value may be negative or more than 200 in such case. The definition in is that the cassette (detector) center of rotation is the point where the beam intersects detector surface. R-axis (2θ) rotation definition makes the crystal at the center of detector rotations. All options in , including autoindexing, work with any detector orientation. Only the resolution display in is incorrect for detector with 2θ non-zero.

Absorption Correction Factors

Q:

The air absorbs equally in all directions! The program calculates air path and corrects for the absorption in the air. It is a very small correction.

AIR ABSORPTION LENGTH

lambda	air absorption length
1.7425	590 mm
1.5418	854 mm (Denzo default is 860)
1.5000	928 mm
1.0000	3145 mm

Auto editing

Q:

It is done automatically, however one can adjust REJECT fraction parameter to improve the accuracy of the rejections. Default value is 0.75, increasing it to 0.8 - 0.85 may help. Look at the display to see how many reflections turn red.

Q

You can omit whole frames from scaling, or use RESOLUTION LIMITS to avoid ice. In the extreme case you may reprocess the same image twice with resolution ranges below and above the ice rings

## Autoindexing

**Q**

**BOX SPOT**

The size of the integration spot affects autoindexing procedure in two ways:

- 1) the size affect the default value of **LONGEST VECTOR** if the **LONGEST VECTOR** command is absent from `auto.dat` file .
- 2) The spot size also affects the signal-to-noise ratio. Only spots above **WEAK LEVEL** (default value 5) are used in autoindexing. You can see these spots in window.

If the spot size is very wrong you may not get enough spots for autoindexing.

**Q**

**CRYSTAL**

**CRYSTAL**

It is normal for autoindexing with different images to result in indexings related by a lattice symmetry operation. Such a rotation **never** changes the assignment of F(+) and F(-). No rotation can. Because you can get different solution related by lattice symmetry, you must only do autoindexing **once** per crystal.

**Q:**

The problem was really quite different. The oscillation range was too large for the combination of resolution and unit cell. Lunes were overlapping and autoindexing failed, by the way this applies to all autoindexing algorithms. When you rotated the crystal you were no longer collecting data down 164Å cell and lune overlap was less severe. The right solution is to collect small oscillation sectors down long unit cell axes. If you already have collected large oscillation sectors down the long unit cell restricting autoindexing to using only low resolution data may help. In this case you may try **RESOLUTION LIMITS 40 6**.

**Q:**

Sorry, it is impossible. It may be possible in the future, but it will require substantial change to the program code. You can make clean distinction of monoclinic from orthorhombic during scaling.



---

*Q* $\varphi \quad \varphi \quad \varphi$ 

There is conversion for the angles for some space groups, however it is much simpler to do autoindexing in . If you still have a problem it is most likely due to misindexing. reverses the definition of x and y axis relative to R-axis software. If you enter **X BEAM** and **Y BEAM** swapped, may misindex.

---

*Q*

This is why **always** outputs 14 best fitting Bravais lattices when autoindexing. Often protein crystal are polymorphic and unit cell is unexpected one.

---

*Q**CRYSTAL* $\varphi$ 

You should autoindex only one image per crystal, to have consistent description of crystal orientation, and you should omit the `peaks search file peaks.file` command on subsequent images.

---

*Q:**CRYSTAL*

You can use option **PRINT zones** in . Or, you can redo the autoindexing with the desired values of **SPINDLE AXIS** and **VERTICAL AXIS**. Also can do the requested calculation.

---

*Q:**RESOLUTION LIMITS*

Specify **RESOLUTION LIMITS** 20 2.1 before autoindexing. This will remove all spots found by peak search outside 20 - 2.1 Å range.

---

*Q:*

type...`peak search file peaks.file GO FIX cell` [It you did **FIT cell** in autoindexing] **UNIT CELL** 98 99 115 90 90 90 and the program will use the **UNIT CELL** value

you supplied. This will work only if the autoindexing unit cell is close enough (say within 1-2%) of the unit cell specified by your command.

*Q:*

**Whoa, fella.** The problem is with the wrong value of **X BEAM**, **Y BEAM** and resulting misindexing of the image. If all reflections have wrong index, at most random crystal orientations, low resolution reflections will not match predicted pattern. With right beam position mosaicity 0.3 did generate all low and high resolution reflections. Program is OK.

*Q:*

Autoindexing of 9-deg rotations is tricky and often fails. You were lucky.

## Box definition

*Q:*

Yes.

*Q:*

No.

*Q:*

If you change box parameters you may change parity (even or odd number) of pixels in the box. Spot ellipse has the same parity as box, so there has to be some adjustment to its shape.

*Q:*

### **INST\_HAMBURGD**

Yes you are correct. The printout definition of axes are arbitrary and are different from display direction definition.

*Q:*

Your background area was about 5 times the spot area. You had a large safety margin. You may get by with lower ratio at the cost of data quality.

*Integr box*

*SPOT*

*BACKGROUND*

*SPOT*

*GO*

## CCP4 Scaling

*Q:*

Use `_york1` instead. `will read also concatenated` `_york1` format (but not `_ip`). `_york` is another creation by Eleanor Dodson that hopefully died by now. The error is "ERROR: Unable to open 1 for reading" comes from the fact that `_york` and `_york1` **do not require** staring batch number, as it is coded in the header of `_york` formats. The "1" is interpreted as file name with hkl's instead and cannot open it if (unless you called output from `"1"`).

**YORK OUTPUT FILE**

---

**Q:**

Ask Eleanor Dodson, author of Rotaprep or Phil Evans, author of SCALA.

## **Computer problems and questions**

---

```
$ run ccd
%DCL-W-ACTIMAGE, error activating image LIBOTS
-CLI-E-IMGNAME, image file DKA300:[SYS0.SYSCOMMON.] [SYSLIB]LIBOTS.EXE
-SYSTEM-F-SHRIDMISMAT, ident mismatch with shareable image
Thinking that this error was due to fact that I have an old version I got the
LIBOTS.* from the v 1.5. I still have the same error.
You need Open(vms)AXP v 6.1 to run . Upgrade your operating system.
```

---

**Q:**

speed is about the same. Postrefinement may take a lot of extra time, and very old versions had no postrefinement, however postrefinement is optional now so it can be as fast as the old one. Subjective variation in the speed of the program may be strongly affected by other programs on the same computer, in particular if the operating system starts to swap programs.

---

**Q:**

<DENZO > KILLED DUE TO INSUFFICIENT MEMORY/SWAP

Message killed is from operating system running out of swap (not memory!) space. Swap space is shared by all programs, so it is dependent on how many processes are active, you can run out of swap space in unexpected moments. This has nothing to do which data set you process. Rebooting the computer killed some other programs and increased the amount of swap space for , however rebooting computers can make life difficult if your system is not well managed. Find a better system manager.

---

**Q:**

Both SUN's and DECstation 5000 are OK, but you run the risk of problems with the programs specific to these machines that program authors may not be aware of.

---

**Q:**

No hints. You will not take advantage of VGX graphics over standard entry level graphics.

---

**Q:**

Increased allocation of memory for storage of reflections.

---

**Q:**

SGI creates hopeless number of almost compatible releases of operating system. SGI are generally supported, however older versions (4.0.5 or so) may produce problems.

---

**Q:**

and are now educated enough about little-endian and big-endian computers, so both programs can cope with byte-swapping problems automatically.

---

```
alias denzo '/usr/users/wladek/HKL/denzo '
alias ccd    '/usr/users/wladek/HKL/INST_CCD' etc, etc.
```

```
$ KODAK := $ZBYSZEK:[WLADEK.DISTR.VMS] CORNELL.EXE
$ FUJI  := $ZBYSZEK:[WLADEK.DISTR.VMS] FUJI.EXE
$ RX    := $ZBYSZEK:[WLADEK.DISTR.VMS] RX.EXE
$ RXS   := $ZBYSZEK:[WLADEK.DISTR.VMS] RXS.EXE
$ MAR18 := $ZBYSZEK:[WLADEK.DISTR.VMS] HAMBURGN.EXE
$ MAR30 := $ZBYSZEK:[WLADEK.DISTR.VMS] HAMBURGD.EXE
$ EMBL  := $ZBYSZEK:[WLADEK.DISTR.VMS] HAMBURG.EXE
$ DENZO := $ZBYSZEK:[WLADEK.DISTR.VMS] DENZO.EXE
```

---

**Q:**

```
ENDFILE: END OF FILE -1
APPARENT STATE: UNIT 5 NAMED
LAST FORMAT: (132a1)
LATELY READING SEQUENTIAL FORMATTED INTERNAL IO
```

**FILE**

Hamlin Archive files are binary and have to be processed with on the same type of computer. If you did reduce Hamlin data on VAX you have to run on VAX or Alpha-VMS. SGI version of does not understand VAX Hamlin binary files. Sorry.

*not*

*Q:*

Why use VAX if you have SGI and HP?

*Q:*

*Q:*

You probably try to run                      compiled on OSF or other brand of computer.

*Q:*

Yes, however future versions (2.0 and higher) will not run on Alpha/VMS.

*Q:*

MARD TGABND\_001.IMAGE

```
%LICENSE-F-NOAUTH, DEC DW-MOTIF USE IS NOT AUTHORIZED ON THIS NODE
-LICENSE-F-NOLICENSE, NO LICENSE IS ACTIVE FOR THIS SOFTWARE PRODUCT
-LICENSE-I-SYSMGR, PLEASE SEE YOUR SYSTEM MANAGER
%TRACE-F-TRACEBACK, SYMBOLIC STACK DUMP FOLLOWS
```

Install properly your window manager.

## Consequences of Profile fitting errors in Denzo

*Q:*

- 1 - count of observations deleted manually
- 2 - count of observations deleted due to zero sigma or profile test
- 3 - count of non-complete profiles (e.g. overloaded) observations
- 4 - count of observations deleted due to sigma cutoff
- 5 - count of observations deleted below low resolution limit,
- 6 - count of observations deleted above high resolution limit,
- 7 - count of partial observations
- 8 - count of fully recorded observations used in scaling

this line seems truncated:

```

      \ /
IP fitted, no o  2 1.0000  0.00  0 3520  0  0  0  0 886 1559
IP fitted, no o  3 0.9419 -0.27  0 3581  0  0  0  0 924 1503
IP fitted, no o  4 0.9741  0.30  0 3632  0  0  0  0 881 1551
IP fitted, no o  5 0.9717  0.67  0 3626  0  0  0  0 841 1500
IP fitted, no o  6 0.9535  0.79  0 3736  0  0  0  0 817 1450
IP fitted, no o  7 0.9241  0.78  0 3916  0  0  0  0 821 1364
```

...stuff deleted...

Shell limit	Lower Angstrom	Upper	Average I	Average error	stat.	Norm. Chi**2	Linear R-fac	Square R-fac
	99.00	2.77	42346.7	1428.2	286.0	12.222	0.081	0.105
	2.77	2.20	15502.0	625.4	312.4	12.300	0.113	0.143
	2.20	1.92	9276.0	495.0	345.2	10.242	0.137	0.176
	1.92	1.75	5322.0	418.9	357.7	5.829	0.142	0.189
	1.75	1.62	4095.8	430.3	392.5	3.866	0.138	0.189
	1.62	1.53	3472.9	483.1	416.1	2.413	0.138	0.191
	1.53	1.45	3163.7	514.0	456.7	2.185	0.134	0.190
	1.45	1.39	2780.1	500.9	453.3	1.522	0.126	0.207
	1.39	1.33	2624.7	523.3	480.5	1.848	0.158	0.215
	1.33	1.29	6305.7	882.4	733.4	6.906	0.241	0.363
All reflections			17538.7	756.0	346.3	10.403	0.096	0.113

Shell		I/Sigma in resolution shells:								
Lower limit	Upper limit	No.	of reflections	with I / Sigma	less than					
		0	1	2	3	5	10	20	>20	total
99.00	2.77	0	0	0	0	5	143	1372	4655	6027
2.77	2.20	0	0	0	0	9	466	2152	3283	5435
2.20	1.92	0	0	0	0	36	804	2870	1756	4626
1.92	1.75	0	0	0	0	93	1201	2770	494	3264
1.75	1.62	0	0	0	0	106	1053	1817	80	1897
1.62	1.53	0	0	0	0	160	859	1079	1	1080
1.53	1.45	0	0	0	0	118	481	515	0	515
1.45	1.39	0	0	0	0	73	204	211	0	211
1.39	1.33	0	0	0	0	61	111	112	0	112
1.33	1.29	0	0	0	0	89	602	665	0	665
All hkl		0	0	0	0	750	5924	13563	10269	23832

$\sigma$

Your problem comes from **PROFILE FITTING RADIUS** not being set. You have to rerun \_\_\_\_\_. **WEAK LEVEL** in \_\_\_\_\_ has nothing to do with \_\_\_\_\_ keyword **SIGMA CUTOFF**

Regular (film output file, `*_ip`) imaging plate output format (i.e. the format you would you to read in in FORTRAN) is:

Flag	0 - full	1 - partial
Intensity (F**2) by profile fitting		
$\sigma$ of intensity (F**2)		
$\chi^2$ of profile fitting		
Intensity (F**2) by profile summation		
Cosine of incidence angle at detector		
Predicted pixel position of spot centroid (slow, fast) directions		
Lorentz, polarization, obliquity combined factor		
Strength of averaged profile in arbitrary units		

```
format (3i4,i2,2f8.0,f7.0,f6.0,f6.0,2f7.0,f6.0,f8.0)
```

Flag      0 - full    1 - partial  
Intensity (F\*\*2) by profile fitting  
 $\sigma$  of intensity (F\*\*2)  
 $\chi^2$  of profile fitting  
Intensity (F\*\*2) by profile summation  
Cosine of incidence angle at detector  
Predicted pixel position of spot centroid (slow, fast) directions  
Lorentz, polarization, obliquity combined factor  
Strength of averaged profile in arbitrary units  
Partiality (fraction of the intensity of the reflection extected in this image)  
Diffraction angle (degrees)

The last may be very useful when preparing input to program SCALA

*Q*:  
It is now the same.

*Q*:  
  
  
  
No, only 1398/174715, about 0.6%

*Q*  
  
 $\sigma$   
  
 $\sigma$   

2	-16	-28	1	134.8	191.2	1.37	25.3	0.995	1371.5	1284.9	0.191	3252.4
0	-12	-32	1	118141	5399	24.77	-10031	0.997	1372.8	1196.3	0.191	3094.7
-8	6	-50	1	9095.4	9536.0	13.85	393.1	0.996	1372.2	805.2	0.192	5856.9

Yes, this an intended part of the program to deal with overflows.

*Q*  
  
 $\varphi$   
  
 $\varphi$   
  
 $\varphi$   
  
 $\varphi$   
  
**york** output file has this information, see above.

*Q*  
No, it is not possible. Current rotaprepa can skip this information.

*Q*  
  
 $\sigma$   
  
Nothing or almost nothing, e.g. resolution limit too high for MAR scanner.

*Q*  
  
Use current version of rotaprep (from ccp4) with `york output file`  
  
*Q*  
  
Check with Eleanor Dodson or Phil Evans.

---

*Q*

Yes.

---

*Q*

No.

---

*Q*

Yes.

---

**Display**

---

*Q*

Boxes are centered on pixel boundaries, circles are centered on precise values.

---

*Q*

Both of them.

---

**Example of Bad Detector Pathology**

---

*Q*

I would check following:

- a) Beam line
- b) Shutter
- c) Detector
- d) Crystal size, radiation sensitivity, mosaicity etc.
- e) Consistency of indexing if data from more than one crystals are used in some space groups
- f) Crystal isomorphism

*C*

---

*Q:* $\sigma$ **ERROR SCALE FACTOR****SPOT**

Reprocessing with  $\sigma$  rarely improves  $\chi^2$  dramatically , unless you correct a major mistake in the processing.



Q:

**ERROR SCALE FACTOR****ERROR SCALE FACTOR**

At the moment it looks like:

Shell limit	Lower Angstrom	Upper Angstrom	Average I	Average error	Norm. stat.	Linear Chi**2	Square R-fac	R-fac
100.00	4.31	205113.5	17423.5	13959.9	4.838	0.087	0.155	
4.31	3.42	140289.3	9859.9	7050.6	5.088	0.077	0.127	
3.42	2.99	59864.0	4943.1	3835.6	5.593	0.092	0.174	
2.99	2.71	31249.4	3211.6	2690.9	4.465	0.101	0.180	
2.71	2.52	21540.4	2658.0	2326.9	3.501	0.113	0.172	
2.52	2.37	16115.9	2358.6	2132.3	2.615	0.120	0.156	
2.37	2.25	13138.6	2214.7	2044.7	2.435	0.132	0.182	
2.25	2.15	10403.3	2107.3	1985.4	2.597	0.163	0.228	
2.15	2.07	7984.0	2084.2	2001.9	1.883	0.191	0.241	
2.07	2.00	6503.1	2127.2	2059.1	1.790	0.226	0.290	
All reflections			52388.8	4959.8	4047.5	3.657	0.092	0.149

with **ERROR SCALE FACTOR** 3.0 and **ESTIMATED ERROR** 0.06.

Things are pretty bad.

I'd try with with e.g. **ESTIMATED ERROR** 0.10 (usually I use 0.06) and **REJECTION PROBABILITY** 0.01. A value for **ERROR SCALE FACTOR**? Usually about 1.5 seems quite good, but I guess in this case I should experiment a bit - but if I need to have **ERROR SCALE FACTOR** greater than say 2.0 then I know that it still isn't right?

I tried scaling again with **ERROR SCALE FACTOR** 1.5 **ESTIMATED ERROR** 0.10 **REJECTION PROBABILITY** 0.1. This time the scaling statistics look like...

Shell limit	Lower Angstrom	Upper Angstrom	Average I	Average error	Norm. stat.	Linear Chi**2	Square R-fac	R-fac
100.00	4.31	207375.1	17638.1	6884.6	4.242	0.088	0.136	
4.31	3.42	141622.7	11636.3	3527.6	4.565	0.080	0.115	
3.42	2.99	60834.7	5307.6	1921.0	4.949	0.091	0.145	
2.99	2.71	31748.7	3013.8	1348.5	4.555	0.099	0.146	
2.71	2.52	21800.0	2247.5	1161.2	4.118	0.109	0.146	
2.52	2.37	16230.9	1822.3	1063.3	3.710	0.117	0.141	
2.37	2.25	13217.2	1602.3	1020.6	3.642	0.128	0.160	
2.25	2.15	10430.2	1422.9	988.6	4.094	0.157	0.206	
2.15	2.07	8010.3	1310.4	1000.5	3.451	0.184	0.220	
2.07	2.00	6584.4	1296.2	1033.9	3.515	0.220	0.268	
All reflections			52826.6	4803.3	2009.7	4.142	0.093	0.131

numbers Any suggestions as to how to make this a bit better? At least now all the  $\chi^2$  are similar - I know that I can compensate for these by increasing **ERROR SCALE FACTOR**, but is there a better way? What exactly is the scanner doing to give such funny data? Elspeth thought that there was a problem with the nonlinearity of the detector above a certain value, which is why I tried using **OVERLOAD VALUE** in Denzo to try and get rid of rubbish data!

I've fiddled around with all my data, and it doesn't scale well no matter what I do. It's a shame, because my data is 96% complete to 2.1Å. I think though because it doesn't scale I will not use the data, and instead collect it all again. It seems like a waste, but I think it is safer to start off with good data with good scaling etc. before making maps and drawing conclusions from their interpretation. Do you think I should give up with the data?

No other choice. It seems you got screwed up.

## Ignore

**Q:**

**IGNORE**

Use: **IGNORE** circle 96. 98. 8 (ignore circle of radius 8mm centered at 96. 98).

```
IGNORE quadrilateral    199.4 81.2
                        108.8 98.4
                        109.5 92.9
                        199.4 74.8
```

These are the four corners of quadrilateral to be ignored. Note that order of the 4 corners matter, they have to be input in counterclockwise order.

**C:**

**IGNORE**

**Q**

**IGNORE**

Do not use **IGNORE** bar. Use **IGNORE** quadrilateral instead.

## Mosaicity

**Q**

**FIT** batch mosaicity 1 to 60 mosaicity will be refined separately in each batch of data. or **FIT** crystal mosaicity 1 to 10 11 to 20 21 to 30 31 to 40 41 to 50 51 to 60 mosaicity will be refined separately for 6 groups of batches, with separate values for first, second, third, etc.. groups of 10 batches. These groupings may be different for mosaicity and for other unit cell parameters.

**Q**

If you have 60 frames and you want to process first 10 with mosaicity 0.2 degree next 10 with mosaicity of 0.25 degree, and so on, up to the last 10 with mosaicity of 0.45, you may do that in the following way: In the command file type:

```
MOSAI CI TY 0.2 ... SECTOR 1 to 10 ... GO    [to execute autoindexing]
@refine.dat  [refine first 10 images] MOSAI CI TY 0.25 SECTOR 11 to 20
@refine.dat  [refine next 20 images] MOSAI CI TY 0.3  SECTOR 21 to 30
@refine.dat  [refine next 20 images] MOSAI CI TY 0.35 SECTOR 31 to 40
@refine.dat  [refine next 20 images] MOSAI CI TY 0.4  SECTOR 41 to 50
@refine.dat  [refine next 20 images] MOSAI CI TY 0.45 SECTOR 51 to 60
@refine.dat  [refine next 20 images] END OF JOB
```

**Q**

All reflections that completely cross Ewald sphere are included.

---

*Q*

You can only refine isotropic mosaicity. It is better to have larger mosaicity in        and refined by       , as        can reliably observed (partial) reflections as non-diffracting, but not the opposite. If mosaicity is decreased some partial reflections will be relabeled as fully recorded.

---

*Q*

#### **CROSSFIRE**

**CROSSFIRE** has to do with focusing of X-rays, not reflection width in spindle angle. Both        and        use isotropic mosaicity only. Anisotropic option is not yet implemented.

### **Negative Intensivities**

---

*Q*

Output from        is  $I, \text{sig}(I)$ , where  $I$  refers to Lorentz, polarization etc. corrected diffraction intensity. Sometimes this  $I$  is referred to as  $F^*2$ . Even the best (averaged) estimate of intensity may be negative - see discussion of **SIGMA CUTOFF** keyword in       .

### **Other**

---

*Q*

One should be able to put many different `.ar` files. I tried up to 25. Be sure that they have different batch numbers! If not use the option

```
ADD 1000
FILE XXXX1.AR
ADD 2000
FILE XXXX2.AR
```

to give different batch numbers to different `.ar` files.

---

*Q*

You cannot refine mosaicity in `*.ar` files. It has not been implemented. However you can refine B factors, it is done by default, so you do not need to do anything special.

---

*Q*

Some versions of ftp will create wrong file format on VMS. Change to UCX (native VMS implementation) in such a case. Also remember to use binary ftp not ASCII ftp (sometimes a default option in ftp!)

---

*Q*

      is running out of memory and ignores some reflections. This may happen if the **BOX** is unreasonably large. Reduce the **BOX** size or use the        version of the program.

---

*Q*

Use **ScalepackBig** or **ScalepackVeryBig**

---

*Q*

Use INST\_RX instead. INST\_CORNELL is design to read in data scanned with KODAK scanner at Cornell (now decommissioned).

---

*Q:*

---

CRYSTAL6% SCALEPACK SCALEPACK: SYNTAX ERROR AT LINE 1: ^A\_^H,>q^Ce^FMH^P^B^A^K^A ^  
Yes, you are correct. This is not the right executable.

---

*Q*

Try Ctrl D (UNIX) or Ctrl Z (VMS), to terminate input into the program.

---

*Q*

Exec format error. Wrong Architecture

You try to run program compiled for different operating system or OS version. Investigate.

---

*Q*

INST and programs recognize swap bytes automatically.

---

*Q*

Sorry, we do not know. Ask SSRL staff.

---

*Q*

\*\*\* TERMINATING /usr/perutz1/wladek/DENZO /DENZO \*\*\* Received signal 10 (SIGBUS) Bus error

An error from old SUN version. Problem specific to SUN. If you get such an error contact authors of the program.

---

*Q*

It should work but it is not as good as the new one.

---

*Q:*

Yes, now it is a standard version.

---

*Q*

**YSCALE**

Yes, however it is a default value.

---

*Q*

**FILM**

I hope that you did not process Daresbury MAR with film rotation -90.0 (or 90.0). The data would be disastrously bad. Maybe you confused MAR with R-axis at Daresbury? R-axis with horizontal spindle (3 such machines exist) requires film rotation -90.0 or 90.0 Polarization number should be positive for horizontal spindle at Daresbury regardless of value of film

rotation. If you specified **POLARISATION** daresbury you are safe both with MAR and R-axis.

Q

The program is running out of memory and ignores some reflections. This may happen if the profile fitting radius is unreasonably large. Reduce the **PROFILE FITTING RADIUS** size or use the version of the program.

Overload

Q

Default value is OK for R-axis and some MAR scanners. It represents the full dynamic range of possible numbers in the file. Now reads overload table from MAR scanners. On some old MAR scanners spiral to Cartesian conversion program incorrectly handled saturation. Newer version seem to be OK. Some sites may have locally modified software (EMBL-Hamburg, MRC Cambridge), check with whoever made the modification. With MAR scanners user has to know instrument he or she is using.

Q

	BACKGROUND	SPOT
<b>OVERLOAD VALUE</b>		
The <code>backgr ovfl</code> flag may be set a an overloaded pixel, but also by a high percentage of pixels rejected from background calculation. On MAR scanners spots outside active area (corners of the image) will never have valid background and will have <code>backgr ovfl</code> condition set.		
<code>backgr ovfl</code> column is really sum of all reflections with any kind of problem with background, one of the possibilities being having one or more overflowed pixels in the background areas. Other problems are background varying too much, or too many pixels being obscured by other reflections, or too many pixels not having a valid measurements. The last affects mostly spots outside active area of the detector. Adjusting <b>RESOLUTION LIMITS</b> can prevent prediction of spots there. <b>OVERLOAD VALUE</b> has only effect on one of the factors involved.		

Q

VALUE	ERROR	OVERLOAD
Neither <b>OVERLOAD VALUE</b> nor <b>ERROR</b> systematic have any direct impact on outlier rejection. <b>OVERLOAD VALUE</b> defines detector saturation. Systematic error in is a number that describes expected correlation between positional errors on the detector. Outlier rejection happens when symmetry related reflections do not agree in measured intensity.		

Q

Another problem with MAR scanners is sometimes a wrong scale factor applied to strong fluorescence signal compared to weak signal. Setting **OVERLOAD VALUE** to below transition point between strong and weak range of data may help, but this would eliminate a lot of reflections. Transition point is roughly at about 10000 MAR units.

Q

**SHOW OVERLOADS**  
It works now with all formats.

Refinement

Q

---

**SECTOR**

**@**

**@**

**SECTOR**

**@**

**X BEAM**

**Y BEAM**

**@**

This will work.

---

**Q:**

There are some improvements to refinement algorithm that make it more stable, however the problem is likely be somewhere else. One likely explanation that would agree perfectly with your description is misindexing of lattice by one index. The solution in such a case is to input correct (or close to correct) values of **X BEAM** and **Y BEAM**.

---

**Q**

**SECTOR**

No. It does not matter where you put the **SECTOR n to m** command as long as it is before the first **GO** statement. The do-loop starts at **START REFINEMENT** and ends at **CALCULATE GO**

---

**Q**

**CRYSTAL**

**CRYSTAL**

**CRYSTAL**

**OSCILLATION**

That you collected data starting each image at a different **OSCILLATION start**, the way everybody collects his/her data.

---

**Q**

**CRYSTAL**

**CRYSTAL**

**CRYSTAL**

**CRYSTAL**

**CRYSTAL**

What counts is only sum of the **OSCILLATION start** + **CRYSTAL rotx**. It seems that you are inconsistent in how you describe crystal orientation using different procedures.

---

**Q:**

It depends. If mosaicity from postrefinement is higher than that set in , then definitely yes. If you want to precisely characterize the detector, you may wish to rerefine and reintegrate the data with unit cell parameters fixed to postrefinement results. Detector parameters refined in will then be much more precise. Otherwise the gain from reprocessing is very small.

---

**Q**

**START REFINEMENT**

Start a loop that finishes at **CALCULATE GO**. Loop range is defined by **SECTOR 1 to 50**. This will execute the loop 50 times.

C:                       $\sigma$                       PROFILE FITTING RADIUS

PROFILE FITTING RADIUS

**Rhombohedral**

---

*Q*                      You can reindex to hexagonal setting in                      .                      Scenario 5 of the                      manual.  
You can use both primitive and hexagonal settings in                      , you can also reindex  
primitive setting (from                      ) to hexagonal in                      .

C:

**Twinned Crystals**

---

*Q*

1) \_\_\_\_\_

2) \_\_\_\_\_

3) \_\_\_\_\_

The algorithm needs to be somewhat more complicated. One has to include the possibilities of overlap between lattices. Maybe in the future.

---

*Q*                      Put detector as far back as possible to separate the two lattices. Process one at the time. Hope for best. Do not expect outstanding results.

# INDEX

;

; (semicolon) keyword, 71

@

@ (input) keyword, 59

@ (input) keyword, 129

2

2 $\theta$  angle (beam to detector axis), 28, 53, 1422 $\theta$  angle (beam to detector axis), 23, 54

2-D profiles of the spots - displaying, 20

## A

A/D test button, 19

absorption, 8, 31, 113, 144

AIR ABSORPTION LENGTH keyword, 51

correction factors, 142

ADD PARTIALS keyword, 121

ADD keyword, 121

advantages of using , 6

AIR ABSORPTION LENGTH keyword, 51

Alpha (DEC workstation), 147, 148

ANGULAR OFFSET keyword, 33, 51, 66

angular offset in *Denzo* output file, 45

anomalous signal, 8, 89, 92

ANOMALOUS keyword (flag), 112, 121

change of anomalous signal sign, 102

NO ANOMALOUS keyword, 129

NO MERGE original index keyword, 130

OUTPUT ANOMALOUS keyword, 132

Patterson difference, 113

SCALE ANOMALOUS flag, 92, 113, 136

scenario, 112

archive file format, 124, 127, 147

BACKGROUND CORRECTION keyword, 122

FIXED WINDOW keyword, 126

Archive file format, 139

ARCHIVE file format

NO PROFILE TEST keyword, 130

ORIGINAL WINDOW keyword, 132

PROFILE TEST keyword, 133

UNIT CELL is included in the header, 139

auto editing Q&amp;A, 142

auto.dat command file, 13, 29, 157

example, 26

autoindexing, 5, 19, 23, 24, 25, 102, 142, 153

, 143

BOX and SPOT commands, 29

outputs 14 best fitting lattices, 144

parameters, 29

don't keep re-running autoindexing!, 81

failure, 28

FIT command, 32

general description, 24

ice rings, 39

LONGEST VECTOR keyword, 60

misindexing, 30, 157

oscillation range, 29

PEAK SEARCH FILE peaks.file command, 64

peaks used in, 25

rhombohedral space groups autoindexing, 69

twinned crystal, 31

## B

b factor

FIX B keyword, 126

B factor, 88, 89, 91, 93, 95, 96, 97, 116, 121, 130, 133, 134, 136, 154

B RESTRAIN keyword, 122

DEFAULT B FACTOR keyword, 122

FIT B keyword, 126

INITIAL B FACTOR keyword, 128

background, 6, 9, 21, 22, 31, 35, 37, 39, 40, 63, 72, 158

BACKGROUND CORRECTION keyword, 91, 122

BACKGROUND keyword, 25, 36, 42, 51

, 43

background overflow, 156

background size, 145

background subtraction and fluctuation, 137

BOX keyword, 52

high intrinsic background, 38

non-random signal in background, 135

REJECT fraction *Denzo* command, 68

BAD VALUE keyword, 52

batch (data set) in , 124

batch processing, 15, 16, 41, 81, 156

batchDenzo.com command file, 41

film pack parameters, 75

keywords required for..., 76

, 89, 94



**SECTOR** keyword, 69  
 batchDenzo.com command file, 41  
 Bayesian reasoning, 10, 90  
   rejection of outliers, 135  
 beam, 9  
   angular spread, 9, 29, 36, 53  
   beam search routine, 81  
   beam stop shadow, 58  
**CROSSFIRE** keyword, 53  
   direct beam, 23, 28, 30, 53, 81  
   efficient use, 7  
   findbeam.dat search routine, 42  
   focus, 9, 53  
   intensity fluctuations, 9, 30  
**LAMBDA** keyword, 72  
**NO BEAM** keyword, 62, 71  
   synchrotrons, 61  
   thinned beam stop, 28, 81  
**USE BEAM** keyword, 42, 62, 71  
**USE NO BEAM** keyword, 42  
**WAVELENGTH** keyword, 25, 72  
   wavelength modifier to **FI T** keyword, 125  
   x beam & y beam parameters, 78  
**X BEAM** and **Y BEAM** positions, 28  
**X BEAM** keyword, 25, 28, 53, 73, 81  
   x beam log statement, 45  
**Y BEAM** keyword, 25, 28, 53, 73, 81  
   y beam log statement, 45  
 Bijvoet differences, 48, 74, 136  
 box (integration box), 20, 44, 49, 51, 63, 151  
 Box (integration box), 145  
**BOX** keyword, 14, 25, 29, 42, 52, 58, 143, 154  
 Bravais lattice, 24, 27, 69, 117, 118, 144  
   **centered monoclinic**, 119  
   **centered orthorhombic**, 119  
   **centered tetragonal**, 119  
   **cubic**, 27, 74, 114, 118  
   **hexagonal**, 27, 74, 118  
   **monoclinic**, 27  
   **orthorhombic**, 27  
   **primitive monoclinic**, 119  
   **primitive orthorhombic**, 119  
   **primitive tetragonal**, 119  
   **primitive triclinic**, 119  
   **rhombic**, 27, 118  
   **tetragonal**, 27, 74, 114  
   **triclinic**, 27  
   trigonal, 74, 114

---

## C

, 158  
**CALCULATE** keyword, 41, 52, 54, 66, 70, 157  
 calculate log statement, 42  
**CASSETTE** keyword, 25, 28, 33, 42  
**CASSETTE** keyword, 52  
 CCD detectors, 6, 9, 69, 74, 115, 145  
   , 146

**CELL** keyword, 53  
 cell reduction, 24  
 centered space group, 78  
 chymotrypsin, 8  
 circle, 59  
**Close** button, 20  
**Color** button, 19  
 command file, 13, 46  
   auto.dat, 26  
   batchDenzo.com, 41  
   experiment.dat  
     example, 26  
   findbeam.dat, 42  
   refine.dat, 41  
   scaletwosca.com, 101  
   site.dat  
     example, 25  
 Computer problems and questions, 146  
 computer recommendations, 77  
 Control input in , 89  
 conventions  
   detector orientation, 47  
   Gravity/beam orientation system, 48  
   input to , 46  
   International Tables, 117  
   rotation in , 79  
   Spindle/beam orientation system, 48  
**COUPLING** keyword, 53  
**CROSSFIRE**, 33, 53, 154  
**CROSSFIRE** keyword, 53  
**CRYSTAL** keyword, 54, 79  
**CRYSTAL** rotz, roty and rotx keyword  
   rotation conventions, 80  
 crystal symmetry  
   Bravais lattice, 24, 27, 69, 117, 118, 144  
   pseudo symmetry, 24  
 crystallography as an iterative process, 10  
 crystals  
   cracked crystal, 29  
   frozen, 39  
   intrinsic parameters, 6  
   orientation, 23, 26  
   parameters, 32  
   satellite crystals, 9, 25, 29, 135  
   size, 36  
   slippage, 8  
   twinned crystals, 29, 31, 146  
 Crystals  
   twinned crystals, 158  
 cutoff, 68

---

## D

DEC Alpha, 147, 148  
**DEFAULT B FACTOR** keyword, 122  
**DEFAULT SCALE** keyword, 122, 129  
 density of the spots, 66  
 Denzo keyword modifier  
   absorption

of the film, 56  
   ip format, 123, 127, 128, 133, 139  
   \_york1 format, 121, 124, 127, 128, 133, 139, 146  
 detectors, 8, 9  
   alignment, 9, 55  
   calibration, 8  
   CCD, 6, 9, 69, 74, 115, 145  
   data file template  
     **RAW DATA FILE** keyword, 67  
   DIP (MacScience), 47, 49, 52, 55, 63, 81  
   distortion, 9, 22  
   dynamic range, 19  
   Fuji, 47  
   Hamlin (SDMS), 122, 147  
     archive file format, 122, 124, 126, 127, 147  
     **FIXED WINDOW** keyword, 126  
   Image Plate, 5, 9  
   Kodak, 47  
   manual IP, 9, 28, 42, 81  
   Mar, 9, 47, 49, 63, 81  
   MAR, 37  
   multi-wire proportional counters, 9  
   non-planarity of film or IP, 9  
   orientation, 23  
   orientation conventions, 47  
   parameters, 6, 25, 28, 34, 78  
     **FORMAT** keyword, 58  
   precession method, 6  
   problems, 8, 9, 19, 96, 113, 135, 151  
   quality, 90  
   R-axis, 9, 21, 37, 42, 47, 49, 53, 55, 63, 81, 91, 127, 142,  
     144, 148, 153, 156  
   saturation, 9, 39, 156  
     **OVERLOAD VALUE** keyword, 63  
   shutter, 8, 96, 151  
   single pixel spikes, 9  
   spindle, 96  
   spindle motor backlash, 8  
   spiral, 6, 9, 69  
   unusual instruments, 42  
   vibration, 17  
   Weissenberg method, 6, 53, 56  
**Diff Vec**, 21  
 diffuse or satellite scattering, 35  
 diffuse scattering, 68  
**Dim** button, 20  
 DIP (MacScience)  $\Rightarrow$  detectors, 55  
 Display, 151  
**DI STANCE** keyword, 28, 33, 53, 54  
 distortion, 24, 26, 27, 30, 144  
**DO NOT EXTEND PARTIALS** keyword, 123  
**DO NOT REJECT OUTLIERS** keyword, 123

---

## E

eigenvalue filter, 34, 89, 91, 95, 97, 134  
 Eigenvalue filter  
   **EIGENVALUE FILTER** keyword, 54  
**END FIT** keyword, 57

**END FIT** keyword, 123  
**END FIX** keyword, 57  
**END OF JOB** keyword, 42, 54, 153  
**END OF PACK** keyword, 54  
 error  
   **ERROR** keyword, 54  
   error model, 10, 34, 40, 54, 90, 96, 113  
   error scale factor, 88, 96, 123  
   **ERROR SCALE FACTOR** keyword, 95, 97, 123  
   estimated error, 35, 64, 88, 95  
   **ESTIMATED ERROR** keyword, 95, 97, 98, 100, 123,  
     131, 152  
   expected error, 40, 55, 90, 135  
   in the fitting, 40  
   overload, 55  
   partiality, 55  
   positional, 55  
   profile fitting, 148  
   systematic error, 55, 95, 123  
   y/x factor, 55  
    $\chi^2$ , weighted measure, 54  
   **ERROR SCALE FACTOR** , 151  
   **estimate of I**, 137  
   Ewald sphere, 33, 153  
   experiment.dat command file  
     example, 26  
   experimental parameters, 13  
   experimental procedure, 7  
**EXTEND PARTIALS** keyword, 123

---

## F

**F\*\*2**, 149  
   , 143  
 factors which lower  $R_{\text{merge}}$ , 90  
 Fast Fourier Transform, 24  
 Fcalc, 138  
**FIDUCIALS** keyword, 56  
**FILE** keyword, 124  
 film  
   optical density, 55  
**FILM** keyword, 56, 58, 73  
**FILM ROTATION** keyword, 25  
 filter, 61  
 final statistics, 7  
 findbeam.dat command file, 42  
 fit algorithm, 33  
**FIT B** keyword, 126  
 fit cell, 69  
 fit command  
   first round, 32  
**FIT** keyword, 57  
   **FIT cell** modifier description, 57  
**FIT** keyword, 124  
**FIX B** keyword, 126  
**FIX** keyword, 57  
**FIXED WINDOW** keyword, 126  
**Floor Down** button, 19  
**Floor Up** button, 19

**FORMAT** keyword, 58, 62, 63  
**FORMAT** keyword, 127  
 Fourier map, 8  
**FRAME WIDTH** keyword, 127  
**Frame** button, 20  
 Fuji scanners, 81  
**Full Scale** button, 19

---

## G

**G0** keyword, 58  
**G0** statement, 32  
**Go** button, 19  
 goniometer, 23, 80  
 graphite, 61  
 Gravity/beam system, 48  
 Green Box, 21  
 guard, 63

---

## H

Hamlin (SDMS)  $\Rightarrow$  detectors, 126, 127  
 hand of the structure, 48  
**Heavy Atom Search**, 98  
**Help** button, 20  
 histogram in log file, 35  
 history of , 4  
**HKL MATRIX** keyword, 127  
**HKL SCALE** keyword, 128  
**HKL SHIFT** keyword, 128  
 hklpredictions, 18, 22, 73  
 reference , 4

---

## I

**I+ and I-**, 112  
**I BOX** keyword, 58  
 ice rings, 25, 29, 39, 58, 135, 142  
 , 142  
 Ignore, 153  
**IGNORE** keyword, 58  
**IGNORE OVERLOADS** keyword, 128  
 imaging plate, 59  
**IMAGING PLATE** keyword, 56, 58  
 include no partials, 130  
**INCLUDE OVERLOADS** keyword, 128  
 include partials, 130  
 indexing, 8  
 manual indexing, 78  
 infinite loop, 155  
**INITIAL B FACTOR** keyword, 122, 128  
**INITIAL SCALE** keyword, 129  
**INPUT** keyword, 59  
 input file format, 46  
**INPUT** keyword, 129  
**Int. Box** button in , 20

intensity, 20  
 displayed in , 21  
 maximum pixel intensity in , 21  
**INTENSITY AVERAGING WINDOW** keyword, 59  
 interactive processing by  
 keywords required for..., 76  
 interactive session, 41  
 International Tables, 117

---

## K

keyword  
 format conventions, 25  
 keywords  
 defaults, 46  
 Keywords with no Defaults, 76

---

## L

**LAMBDA** keyword, 59, 72  
 lattice  
 superlattice, 31  
**LATTICE** keyword, 60  
 Launching the program, 17  
**LEFT MARGIN** keyword, 60  
 linear and square  $R_{\text{merge}}$ , 96  
 Linux operating system, 77  
**LIST** keyword, 34, 60  
 local scaling, 130  
 log file, 15, 65  
**LONGEST VECTOR** keyword, 29, 60, 143  
 Lorentz polarization factor, 6  
 lune, 29  
 lysozyme, 8

---

## M

MacScience, 51  
 MacScience (DIP)  $\Rightarrow$  detectors, 55  
 MAD, 130  
 MAD structure determination method, 10  
 madness, 127  
 magnification, 20  
 major axis, 78  
 major zone, 78  
 MAR, 37  
 memory, 17, 77, 146, 147, 154, 156  
**MERGE** keyword, 129  
 Minor Wladek, wladek@iwonka.med.virginia.edu, 4  
 MIPS R8000, 77  
 misindexing, 30, 144, 157  
**MONOCHROMATOR** keyword, 61  
 mosaicity, 8, 13  
 mosaicity modifier to **FIT** keyword, 125  
 very low, 29  
 Mosaicity, 153

**MOSAICITY** keyword, 61  
 , 43  
**MOSAICITY** keyword, 129  
 mouse buttons, 19, 24  
   left, 20, 151, 158  
   middle, 18, 20  
   right, 19, 20  
**Multiple Native data sets**, 100  
 multiple passes of data collection, 115  
 myexperiment.dat command file, 13

---

## N

negative intensities, 137  
 Negative I's, 154  
**NO ANOMALOUS** keyword, 129  
**NO BEAM** keyword, 62, 71  
**NO MERGE** keyword, 130  
**NO PROFILE TEST** keyword, 130  
 none, 63  
 non-isomorphism, 90, 99  
 non-vertical spindle, 42  
**NUMBER OF ITERATIONS** keyword, 130  
**NUMBER OF RECORDS** keyword, 62  
**NUMBER OF ZONES** keyword, 131  
 numerical parity, 21

---

## O

Off-line Image Plate, 81  
**OMIT** keyword, 62  
 operating systems  
   Linux, 77  
   OSF, 77  
   Unix, 77  
   VMS, 77, 147, 148  
 OPERATING SYSTEMS  
   VMS, 22, 140, 147, 154  
 operative philosophy, 77  
**ORIENTATION AXIS 1** keyword, 70  
**ORIENTATION AXIS 1** keyword, 131  
**ORIENTATION AXIS 2** keyword, 72  
**ORIENTATION AXIS 2** keyword, 131  
 original index, 130  
**ORIGINAL WINDOW** keyword, 132  
**OSCILLATION** keyword, 62  
 oscillation image, 25  
 oscillation range, 8, 28, 29, 143  
   low range & still images, 29  
 OSF, 77  
 Otwinowski Zbyszek zbyszek@mix.swmed.edu, 4  
 outliers, 8, 135  
 outliers rejection, 8, 135  
**OUTPUT ANOMALOUS** keyword, 132  
**OUTPUT FILE** keyword, 74  
*output file format*  
   Denzo\_ip, 149

  \_york1, 149  
**OUTPUT FILE** keyword, 132  
   , 43  
 overflows, 150  
**OVERLAP** keyword, 63  
 overlaps, 8, 29  
 overload, 55, 156  
**OVERLOAD VALUE** keyword, 63  
 overloaded reflections, 92, 96

---

## P

  , 147  
 parser pitfall, 46  
 partiality, 55  
 partiality refinement, 34  
 partially recorded reflections, 6, 81, 92  
 partials, 72  
 Patterson, 113  
 Peak picking, 24  
**PEAK SEARCH FILE** peaks.file command, 64  
**Peak search** button, 19  
 peaks.file, 19, 22, 25  
 phase information, 89  
 Photon Factory, 53  
 Photon Factory - synchrotrons, 56  
 plate rotation, 64  
 polarization, 114, 155  
**POLARIZATION** keyword, 61  
**POLARIZATION** keyword, 132  
 polymorphic crystals, 144  
**POSTREFINE** keyword, 97, 132  
 postrefinement, 28, 35  
 Postrefinement, 124  
 postscript file., 19  
 potential derivative, 98  
 precession method, 6  
**predicted decrease**, 40  
 predictions, 14  
**PRINT** keyword, 133  
 print zones, 144  
 print/end print Denzo keyword, 64  
**PRINT/END PRINT** keyword modifier  
   chi square, 64  
   fitted, 64  
   hkl, 64  
   memory, 64  
   profiles, 64  
   shifts, 64  
   solver, 64  
   statistics, 64  
   zones, 65  
**Profile Fit R** button, 18  
 profile fitting, 14, 149  
   profile fitting radius, 18  
 Profile fitting, 148  
 profile fitting radius, 14, 66, 149  
**PROFILE TEST** keyword, 133  
**PROFILES FITTED** keyword, 133

**PROFILES SUMMED** keyword, 134  
prompt, 22  
pseudo symmetry, 24  
pseudo zone, 79

---

## Q

quadrilateral, 58, 153

**QUIT** keyword, 66, 71

---

## R

**RADIAL OFFSET** keyword, 33, 51, 66  
radiation damage, 96  
radius of convergence, 32  
RAM, 17  
**RAW DATA FILE** keyword, 67  
R-axis  $\Rightarrow$  detectors, 37, 55  
reciprocal space, 22, 29, 31, 126  
recommended hardware, 77, 82  
**RECORD LENGTH** keyword, 67  
**RECSQ** keyword, 134  
redundancy, 90  
reference batch, 89  
**REFERENCE BATCH** keyword, 95, 97, 98, 100, 103, 134  
**REFINE NO PARTIALITY** keyword, 67  
**REFINE PARTIALITY** keyword, 67  
refine.dat command file, 13  
refine.dat command file, 41  
refinement, 9, 14  
refinement Q&A, 156  
strategy, 6  
refineone.dat, 33  
reflection conditions, 117  
reindexing in Scalepack, 102  
reject, 68  
**REJECT** keyword, 40, 68  
**REJECT HKL** keyword, 134  
**REJECT OUTLIERS** keyword, 135  
reject.dat, 135  
rejection criteria, 68, 91  
**REJECTION PROBABILITY** keyword, 135  
rejects, 89  
remeasured, 44  
reprocess, 35  
**RES** keyword, 68  
resolution  
displayed in , 21  
, 144  
resolution limit, 8, 91  
**RESOLUTION LIMITS** keyword, 29, 68, 71, 143  
**RESOLUTION** keyword, 136  
resolution shells, 91  
**RESOLUTION STEP** keyword, 136  
**Reverse** button, 19  
Rhombohedral, 158

R-linear, 135  
 $R_{\text{merge}}$ , unweighted error measure, 8, 10, 89  
, 146  
R-squared, 135

---

## S

saturated image pixels, 19  
, 146  
**SCALE ANOMALOUS** keyword, 92, 136  
**SCALE RESTRAIN** keyword, 137  
SCALEPACK, 42  
scaletwosca.com command file, 101  
scattering, 68  
Scenarios, 94  
screen output., 26  
screening through your frames, 19  
 $\Rightarrow$  , 127  
, 157  
**SECTOR** keyword, 69, 70  
**SECTOR** keyword, 137  
**SECTOR WIDTH** keyword, 137  
, 46  
, 147  
**Show overflow** button, 19  
shutter, 96, 151  
shutter malfunction, 8  
sigma cutoff, 91, 137  
**SIGMA CUTOFF** keyword, 91, 137  
simulation as a teaching tool, 31  
simulation of diffraction pattern, 8  
single measurements, 91  
site.dat command file, 13  
example, 25  
**SKEW** keyword, 33, 69  
slope, 68  
small molecules, 5, 10, 29, 52, 89  
, 150  
, 150  
, 150  
**SPACE GROUP** keyword, 69  
as replacement to **LATTICE** keyword, 60  
Space Group Identification, 117  
**SPACE GROUP** keyword, 138  
space groups, 8  
spindle axis, 78  
**SPI NDL E AXI S** keyword, 54, 70, 78  
**SPI NDL E AXI S** keyword, 131, 139  
spindle motor backlash, 8  
Spindle/bam orinetation system, 48  
**SPOT**, 36, 63  
distortion, 55  
elliptical, 36  
radius, 36  
spot radius, 60  
**SPOT** keyword, 29, 70, 143  
spot profile, 9  
spot separation, 8

spot shape, 35  
 Spot Shape, 37  
 spots  
   ellipsoidal, 36  
 spread, 8  
 spurious peaks, 29  
   , 155  
   , 157  
**START REFINEMENT** keyword, 70  
**starting point**, 19  
 statistic error measures  
   problems with no good statistical criteria, 10  
 statistic error measures  
    $\chi^2$ , weighted, 10  
    $R_{\text{merge}}$ , unweighted, 10  
 statistical error, 95  
 still images, 29  
**STOP** keyword, 71  
 structure solving, 10  
 successfully indexed, 78  
 sulfur, 8  
 summary, 34  
 Sun, 77  
 superlattice, 29, 31  
**SWAP BYTES** Denzo keyword, 71  
 swap space, 17  
 Swap space, 146  
 synchrotrons, 9, 28, 42  
   Brookhaven, 81  
   CHESS, 81, 115  
   , 155  
   Hamburg, 156  
   Photon Factory, 53, 56, 81

---

## T

Text strings, 46  
 thin frames, 137  
**Thin or few frames**, 116  
**TITLE** keyword, 71  
**TOP MARGIN** keyword, 71  
 total error, 95  
 triangle, 59  
 TRUNCATE, 138

---

## U

unit cell, 8  
   cell reduction, 24  
**UNIT CELL** keyword, 53, 71  
**UNIT CELL** keyword, 139  
 Unix, 77  
   , 154  
**Update Pred.** button, 19  
**USE BEAM** keyword, 42, 62, 71  
**USE FIDUCIALS** keyword, 72  
**USE NO BEAM** keyword, 42

**USE NO FIDUCIALS** keyword, 72  
**USE PARTIALS NO POSITION** , 72  
**USE PARTIALS POSITION** keyword, 72

---

## V

vertical axis, 78  
**VERTICAL AXIS** keyword, 54, 72, 78  
**VERTICAL AXIS** keyword, 131, 139  
 visualisation of the data, 10  
 VMS OPERATING SYSTEM, 22, 77, 140, 147, 148, 154

---

## W

wavelength, 33  
   Cu  $K_{\alpha}$  (1.5418), 59  
   **LAMBDA** keyword, 59  
**WAVELENGTH** keyword, 72  
**WEAK LEVEL** keyword, 29, 73  
 weak reflections, 91  
   , 43  
 weaker pixels, 20  
 Weissenberg method, 6, 53, 143  
 workstations  
   Alpha, 147  
   DEC Alpha, 13, 147, 148  
   SGI, 77, 147  
   SGI Indigo, 13  
   SUN, 77, 147  
**WRITE BADDIES** keyword, 139  
**WRITE NO PREDICTIONS** keyword, 73  
**WRITE PREDICTIONS** , 73  
**WRITE REJECTION FILE** keyword, 135, 140  
**Write/Print** button, 19

---

## X

**X BEAM** keyword, 25, 28, 53, 73, 81, 142  
 x beam log statement, 45  
**XDisplayF**  
   *general description*, 17  
 xengen urf file format, 127  
 X-plor, 131  
 xtal###.x files, 15

---

## Y

**Y BEAM** keyword, 25, 28, 53, 73, 81  
 y beam log statement, 45  
**Y SCALE** keyword, 33, 48, 49, 74  
 york output file, 74, 146  
**YORK OUTPUT FILE** keyword, 74

<b>Z</b>	$\chi^2$ , weighted error measure, 30 $\chi^2$ , weighted error measure, 95
<b>Zoom close</b> , 21	<input type="checkbox"/>
<b>Zoom In</b> button in , 20	$\lambda$ , wavelength, 33
<b>Zoom Out</b> button in , 20	
<b>Zoom Window</b> button, 19	<input type="checkbox"/>
<b>a*</b> , <b>b*</b> , <b>c*</b> , reciprocal unit cell vectors, 126	$\sigma$ , standard variation of intensity, 14, 37, 88, 90, 104, 105, 106, 136
$\alpha^*$ , $\beta^*$ and $\gamma^*$ , reciprocal unit cell angles, 126	<b>F**2</b> , 149
<input type="checkbox"/>	output file with $\sigma_{I+}$ , $\sigma_{I-}$ , 112
$\chi^2$ , weighted error measure, 10, 15, 40, 54, 64, 113, 114, 117, 136, 149, 151, 152	<b>SIGMA CUTOFF</b> keyword, 137
chi square - modifier to keyword <b>PRINT</b> , 64	