

---

## FEATURES

---

### Integration of fibre diffraction patterns: FTOREC and LSQINT

Richard Denny

#### Introduction

The aim of these two programs is to provide an automatic method for integrating intensity on fibre diffraction patterns. The programs are in fact extensively rewritten and augmented versions of programs written to analyze diffraction patterns from Pfl virus by Colin Nave, Richard Bryan and Ian Clifton, packaged to allow refinement of cell and intensity profile parameters and processing of multiple data sets.

The programs are well suited to the analysis of "simple" patterns in which the diffracted intensity of interest is either all Bragg sampled or all continuously distributed along layer lines, although it is hoped that in some cases the background subtraction methods built in to the program LSQINT will be sufficiently sensitive and well-conditioned in order to provide a method for separating continuous from Bragg sampled intensity, leaving the continuously distributed intensity intact for further analysis. The assumption of a "simple" diffraction pattern allows effective integration of intensity by prediction of the positions and profiles of sampling points (either intervals along a layer line or Bragg peaks) and then fitting this linear model to the observed pattern once the background has been subtracted or if a suitable function for the background can be included as part of the model.

#### FTOREC

The program FTOREC provides a means of transforming data from detector space to reciprocal space, but currently assumes the user has prior knowledge of the centre and rotation of the diffraction pattern and the tilt of the specimen. The backstop can be masked out to prevent this region from being fitted in LSQINT and other more general areas using a combination of the BSL functions .MSK and .ADD. By default, the pattern is transformed and quadrant folded with optional corrections for film absorption.

#### LSQINT

Fitting of the data is performed on the reciprocal space output from FTOREC. The first stage of the program is to generate intensity profiles for the sampling points defined by the input lattice parameters. There may be only one parameter defining the c-axis repeat for one-dimensionally crystalline specimens or six for fully crystalline examples. The profiles of the sampling points are described by up to five parameters for the fully crystalline case or two for the one-dimensional case. The two which are always required are the reciprocal particle length and disorientation (arcng). Bragg sampled patterns require one more parameter to define the reciprocal particle diameter and optionally another two terms giving a quadratic dependence of spot width with reciprocal space radius. At this stage, an image can be computed for comparison with the transformed

---

data consisting of equally weighted sampling points smeared out with the given intensity profile parameters. This gives the user an immediate appreciation of the closeness of the input cell and profile parameters to the true values.

Once the user has arrived in the right ball park, it is safe to continue with background subtraction and fitting. There are four background subtraction methods available in LSQINT, three of which require information generated by the profile calculation. It is probably safest, therefore, to start with the option which does not assume any knowledge of the sampling point profiles.

### *Background subtraction methods*

1. A method devised by Paul Langan is used to subtract background when no knowledge of the positions and intensity profiles of the sampling points is to be assumed. A notional aperture of user-defined dimensions is centred on a pixel in the diffraction image. The pixel values that fall within this aperture are sorted into ascending order and a user-defined interval of this sorted list are averaged to calculate the background value for the central pixel of the aperture. This process is repeated for every pixel in the image so that a complete (smoothed) background image is formed. In order to trap excessively high background, the calculated values are compared with the circularly averaged background, which is determined by iterating the procedure of averaging pixel intensities in resolution bins and then rejecting (high) outliers. The number of iterations and the definition of an outlier are again under user control.
2. A similar notional aperture can be used in conjunction with a least squares fit of background and all sampling points whose intensity profiles impinge upon the aperture to estimate the height of the background at the central pixel. In this scheme, the background across the aperture is initially assumed to be a horizontal plane and the height of this plane is fitted along with the relevant peaks. The aperture is then translated by half its width (or half its length when it comes to the end of a raster) in order to build up a grid of known background values. Cubic spline interpolation is then performed over this grid in order to complete the background image for the first cycle. The next iteration will construct a background over each aperture position, not using a plane but with the surface formed by the interpolated background. The background is then re-fitted using this surface, and the relevant peaks, scaling the height of the central pixel of the aperture in the fit. This procedure can be repeated as many times as is desired but it is important to remember that the usefulness of the fit will depend on having good estimates of the intensity profiles and that too many iterations may allow the background to accumulate a linear combination of the intensity profiles.
3. If the background is sufficiently smooth, a simple linear function can be used to fit it simultaneously with the peak fitting. The function currently employed for this

---

purpose is:

$$Background = \sum_{i=0}^3 A_i D^i + B_i \cos 2(i+1)\phi$$

where  $D = \sqrt{R^2 + Z^2}$  and  $\phi = \arctan(R/Z)$ .

4. During the intensity profile calculation, the profile of any sampling point is trimmed if the contribution of the outermost points is insignificant. This means that the significant extent of each sampling point is known and that the total area on the pattern impinged upon by profiles can be excluded from being considered as background. The remaining partial background image can then be interpolated to give a complete background image. If the interpolation gives rise to background values too far from the circularly averaged background (as defined in method 1), the values are replaced by the circularly averaged background at those points. The usefulness of this method is limited if there are large areas of the image where no background information is available.

#### *Peak fitting*

Once the background has been subtracted, the peaks defined by the input lattice and profile parameters can be matched to the image by a linear fit. This can be done either with a least squares method, or with a maximum entropy technique. Depending on the degree to which the sampling points have been smeared out, there may have been some information loss in the observed pattern compared to the ideal case of the squared molecular transform sampled by Dirac delta functions (in one or more dimensions). If this is the case, the matrix describing the linear transformation between the perfect sampling and the observed pattern will be singular and no direct inverse transformation will be available. Even if this is not exactly the case, the large number of operations performed on the matrix using the limited precision available may cause it to appear almost singular. If a least squares fit is desired, it is almost always necessary to filter out the smaller eigenvalues of the normal matrix in order to eliminate the problems of singularity and construct the pseudo-inverse with only the surviving eigenvalues (and eigenvectors). This approach is suitable when the number of peaks to be fitted is around 500 if the fit is performed in double precision and works well for Bragg sampled patterns.

Alternatively, a maximum entropy criterion can be imposed upon the solution relative to some prior distribution with the constraint that the transformed solution agrees reasonably well with the data. This approach guarantees positivity of the solution and removes the problem of singularity by selecting that solution (out of a set of solutions which agree equally well with the data) which contains the minimum information content with respect to the prior distribution. The algorithm of Skilling and Bryan utilized in LSQINT, has the added benefit that it operates only on a subspace of low dimension of parameter space in any iteration, so that diagonalization of matrices is fast and the difficulties associated with roundoff error are reduced. For large problems, this algorithm is faster and more stable than least squares and is recommended for fitting intensity which is continuous along layer lines.

---

### *Cell and profile parameter refinement*

Once a set of intensity values has been assigned to the sampling points by the fitting procedure, the cell and profile parameters may be refined. The maximum number of parameters that can be refined is eleven for a triclinic cell (five profile parameters along with the six cell parameters) but the number is usually less than this. A simple refinement algorithm, due to Nelder and Mead, which requires no derivative calculation has been employed as calculation of the profiles is fast compared to the fitting and background subtraction procedures. The user specifies the number of cycles of refinement, the initial shifts for the profile parameters and the reciprocal cell parameters and the tolerance used to define convergence. The refinement can then be used in tandem with the fitting procedure to provide an overall refinement scheme for intensities and the sampling point position and shape description.

### **Input and output**

The action of both programs is decided by keyworded input from the user. The image file format used is the same as that used by BSL and OTOKO with the minor extension that input files for LSQINT require some reciprocal space mapping information written into the second header record by FTOREC. Intensity output files from LSQINT are formatted with seven columns:

$$h, k, l, R, M, I, \sigma(I)$$

corresponding to the indices and reciprocal space radius of the sampling point, the multiplicity, intensity and standard deviation. If the multiplicity is greater than one, the last reflection in the multiplet list is assigned the total intensity value. In addition to the formatted intensity output file, it is optional to output an image file containing two frames for every frame fitted. Odd numbered frames correspond to the calculated background while even frames represent a simulation of the fitted image from the addition of the calculated background and the smeared out fitted intensity values.

### **Availability**

Executable files are available for SUN, HP and DEC platforms from /nfs/dls12/ccp13/CCP13 on any of the NCD nodes at Daresbury Laboratory. Documentation is in the doc subdirectory.

### **FIT2D**

An alternative program is available for users who have a suitable version of the PHIGS graphics system. Written by Andrew Hammersley at the E.S.R.F., FIT2D also offers the facility to correct Molecular Dynamics image plate data for spatial distortion. Work is currently underway to port the program to use SunPHIGS and to make it available to NCD users at the SRS.