

## XRDUG Seminar III

# *Computer Algorithms Used for XRD Data Smoothing, Background Corrections, and Generating Peak Files*

Edward Laitila  
3/1/2009

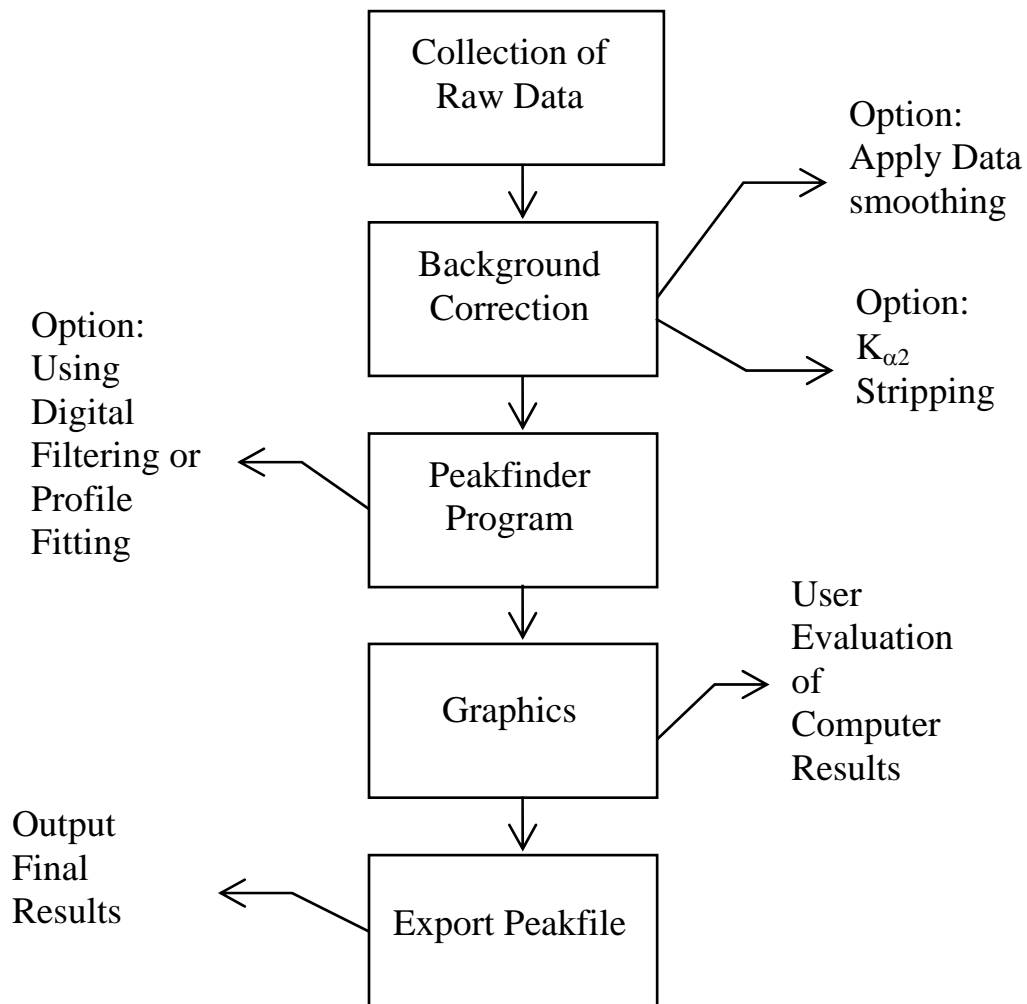
## XRDUG Seminar III

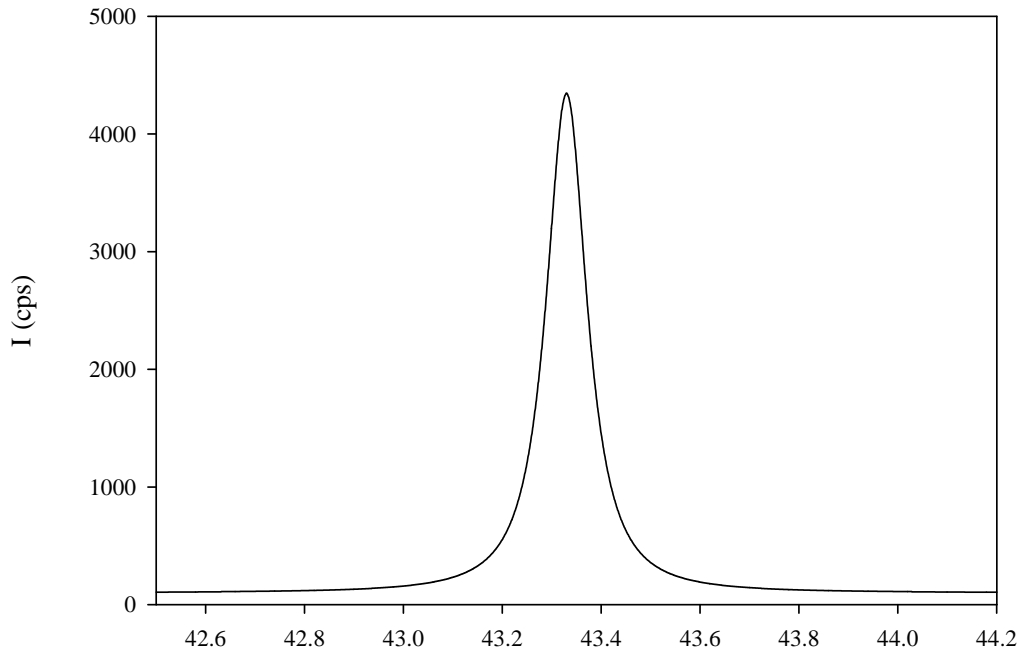
### Computer Algorithms Used for XRD Data Smoothing, Background Correction, and Generating Peak Files:

#### Some Features of Interest in X-ray Diffraction Data:

- Position of diffraction peaks, in particular the d-spacing
- Net intensities of the peaks
- Integrated intensities (areas) of the peaks
- FWHM (full width at half maximum)

#### Procedure for Digital Signal Conditioning of X-ray Diffraction Data for DMSNT Software





**Figure 1 Diffraction Peak Parameters of Interest** – in general three distinct areas, parabolic top, linear inflexion points, asymptotic tails.

### Data Smoothing Methods:

#### 1. “Boxcar” Smooth

- Specify the number of points (N) to use in the smooth, must be an odd number.
- Average data over symmetric interval around data point of interest:

$$\# \text{ points to left} = (N-1) / 2$$

$$\# \text{ points to right} = (N-1) / 2$$

N = number of points averaged in the smooth

- Apply to all data points collected in the diffraction spectrum.
- Example: A 3 point “boxcar” smooth on 7 diffraction data points:

Original Data	10	77	32	35	87	28	36	
1 <sup>st</sup> smooth (on 10)	-	77	32	35	87	28	36	Can't be done
2 <sup>nd</sup> smooth (on 77)	-	40	32	35	87	28	36	
3 <sup>rd</sup> smooth (on 32)	-	40	36	35	87	28	36	
4 <sup>th</sup> smooth (on 35)	-	40	36	52	87	28	36	
5 <sup>th</sup> smooth (on 87)	-	40	36	52	56	28	36	
6 <sup>th</sup> smooth (on 28)	-	40	36	52	56	40	36	
7 <sup>th</sup> smooth (on 36)	-	40	36	52	56	40	-	Can't be done

- 7<sup>th</sup> smooth shows the end result of “boxcar” smooth.
- Lose end data points of the range smoothed:

$$\# \text{ data points lost} = (N-1) / 2$$

- If number of points smoothed (N) is large this process can distort and /or shift diffraction peaks.
- Number of points needed depends on counting statistics and step size, typically 7 points displays dramatic results with minimal distortion of peaks.
- Available in Background Correction in DMSNT software.

## 2. Fast Fourier Transform (FFT) Noise Filter

- Can describe any function as a series of cosine and sine functions.

In General:

$$f(x) = \sum_n A_n \cos(2\pi nx) + \sum_n B_n \sin(2\pi nx)$$

- Describe diffraction data as a Fourier series in the frequency domain.
- Recognize that the statistical fluctuations (noise) in the data can be represented by high frequency terms of the Fourier series, whereas the diffraction peaks are typically low frequency in nature.

- Pass data through a “low pass” frequency filter to eliminate “noise” by removing frequencies above some value.
- Reconstruct function or data with high frequency component removed to produce smoothed data.
- Filter size can be adjusted by the user or automatically chosen by computer.
- Resolution value in software roughly correlates to size of feature removed and related to the frequency cutoff.
- Available in Background Correction in DMSNT software.

### **Background Correction Methods:**

To determine net peak and net integrated intensities, need to subtract the background. In general any function that properly describes the background can be used.

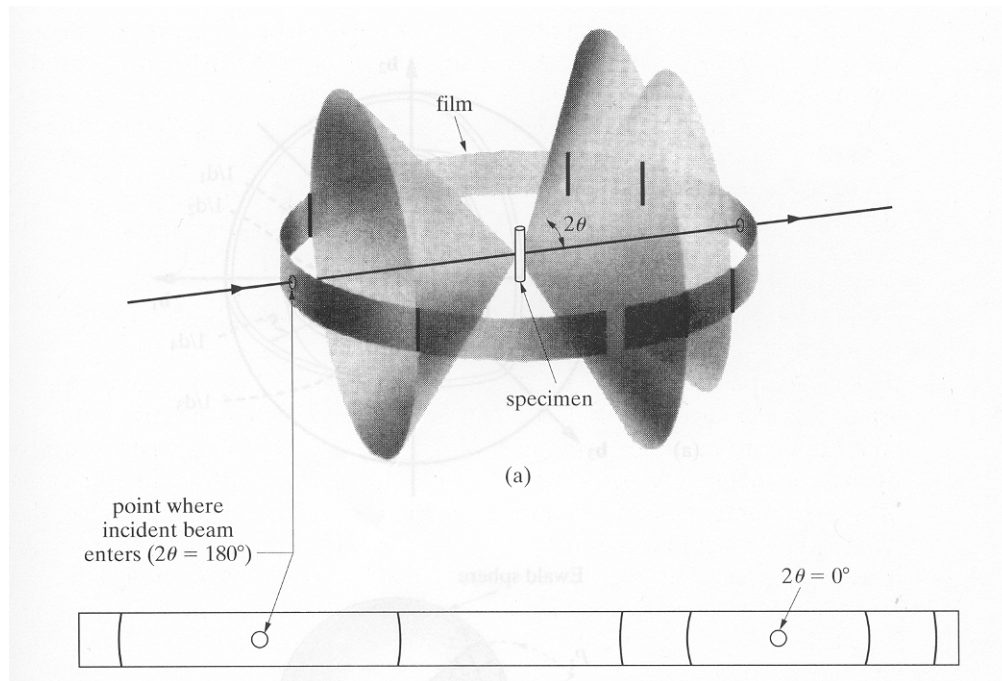
#### 1. Linear background correction:

- Simplest correction method fit line to end points of the data range.
- Use only on small ranges of data, typically one peak, diffraction background spectrum typically is not linear, but over small ranges this is a good approximation.

#### 2. Parabolic background correction:

- One factor that has a large influence on the intensity of a diffraction spectrum is the Lorentz-Polarization Factor (LPF).

$$I \propto \frac{(1 + \cos^2 2\alpha \cos^2 2\theta)}{\sin^2 \theta \cos \theta} \quad \text{LPF}$$



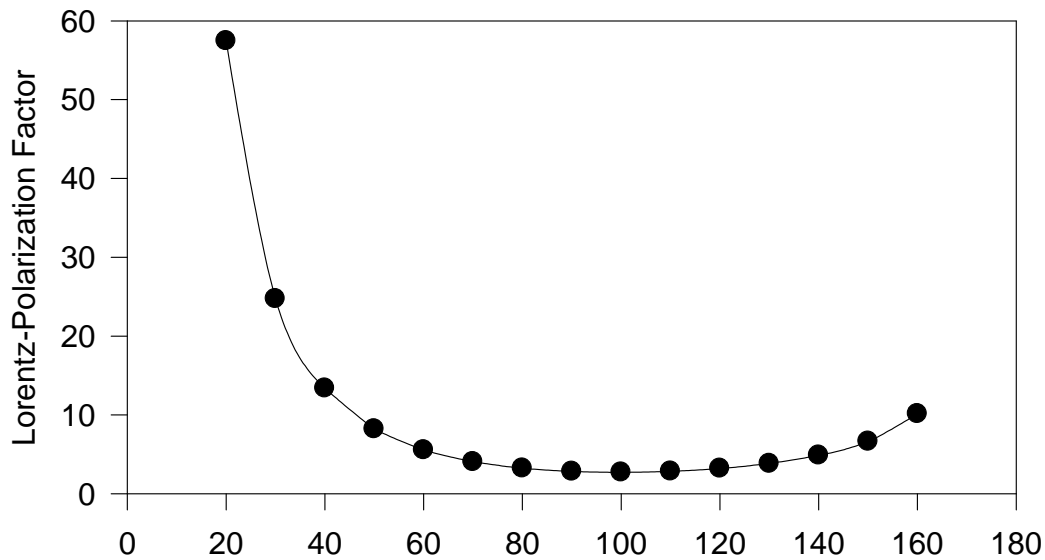
**Figure 2 Debye-Scherrer Diffraction Showing Need for Lorentz Correction**

- LPF is parabolic in nature (See Figure 3).

General Parabolic Equation:

$$y = ax^2 + bx + c$$

- Can describe the whole background spectrum with a parabolic fit of the background.
- Typically must define a minimum of 3 widely-separated points that are in the background in order to fit the equations to the data. Use a least-squares fit and minimize the error between the data and the parabolic function.



**Figure 3 Lorentz-Polarization Factor with a Graphite Monochromator**

- Ideally, this is a good way to correct whole background spectrum; but other factors cause deviations from a parabolic nature and influence the intensity of the background spectrum, such as:
  - Amorphous scattering from the sample holder, or a powder binding material.
  - Intensity loss due to the incident beam being larger than the sample at small  $2\theta$  angles.
  - Metallurgical effects e.g.: presence of an amorphous phase, clustering or short range order of solute atoms, density gradients in packed powder samples, etc..

### 3. 3<sup>rd</sup> Order Polynomial Background Correction:

- General form of the equation:

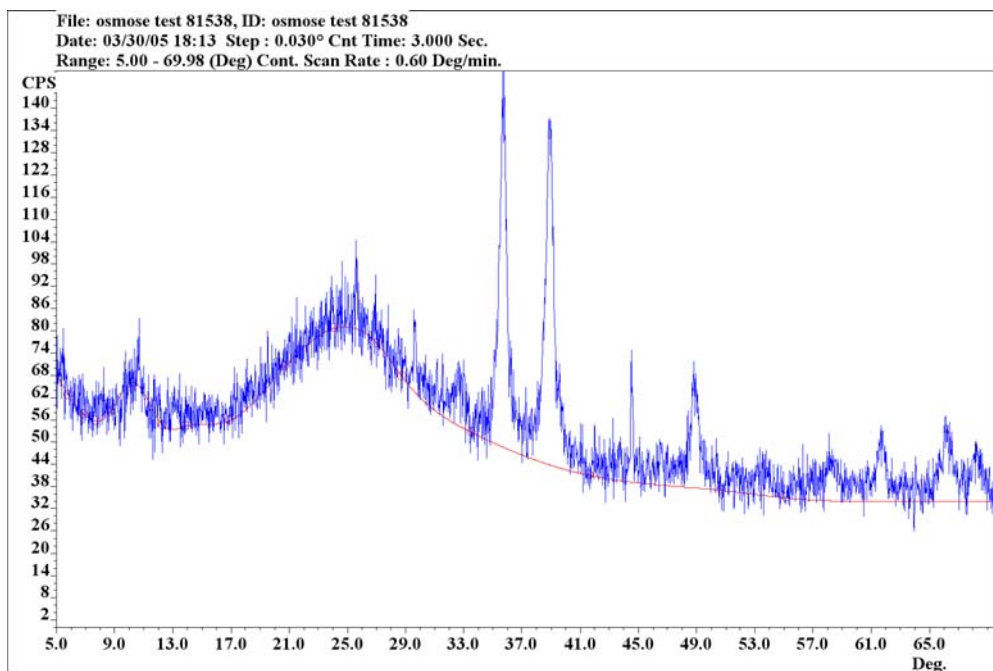
$$y = ax^3 + bx^2 + cx + d$$

- Can adjust to aberrations in background that affect parabolic fit.

- Typically must define a minimum of 4 widely-separated points, that are in the background, in order to fit the equations to the data. Use a least-squares fit and minimize the error between the data and the 3<sup>rd</sup> order function.

#### 4. Cubic Spline Background Correction:

- Assumes nothing about the shape of the background.
- Excellent fit to a nonlinear background.
- A method of cubic polynomial interpolation between intervals that are spliced together to fit the whole pattern.
- Most versatile type of background fit (available in the software).
- Available in DMSNT software, data points used must be selected; however, initially the end points are inserted and cannot be removed but position can be adjusted.



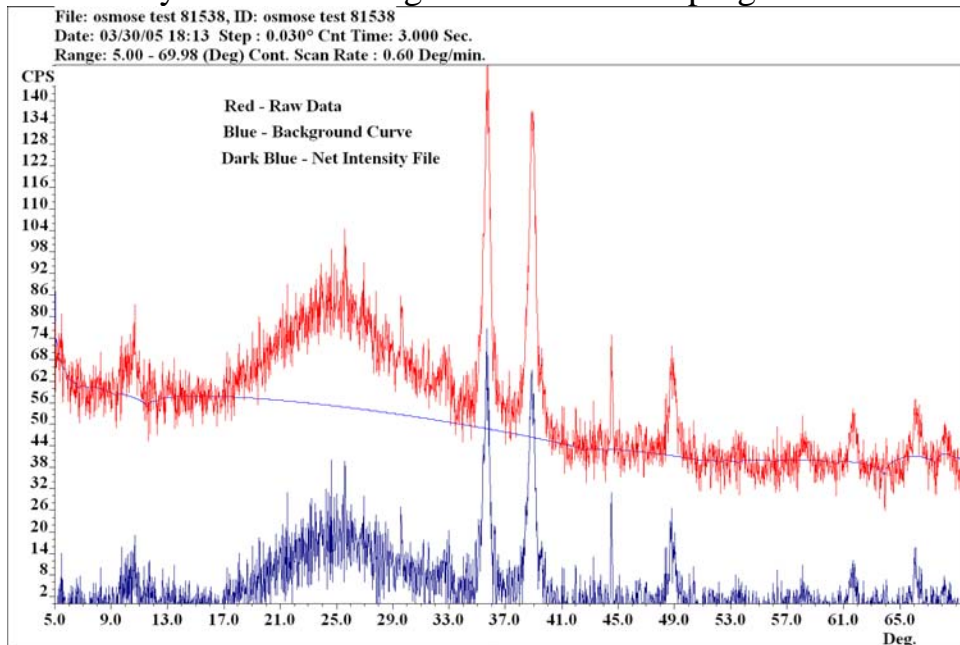
**Figure 4 Data Point Selection in Cubic Spline Background Correction**



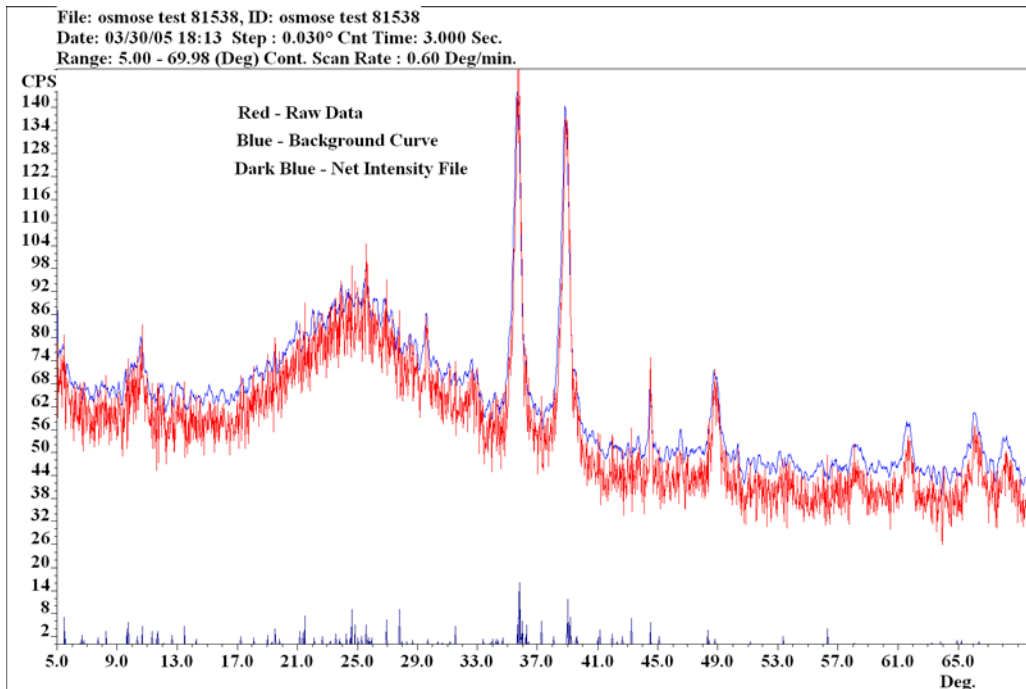
- When selecting points try to keep on the low side of the average background, typically at the bottom of the inner noise band.

#### 5. Box Car Background Correction:

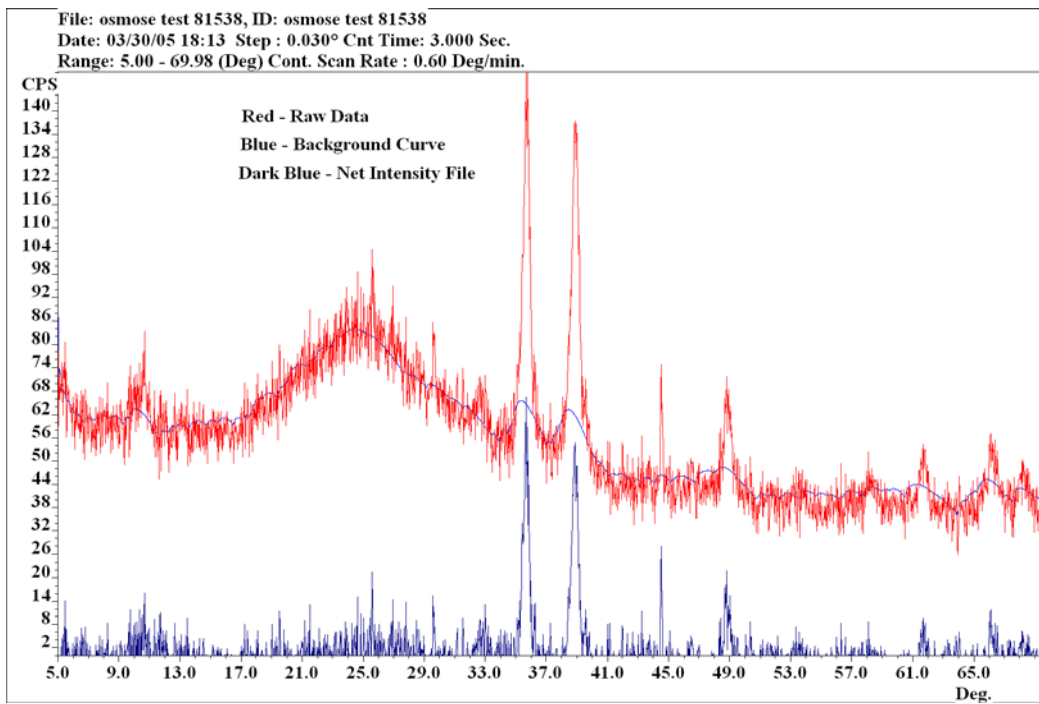
- Applies a “boxcar” smooth to data point intervals specified by the user, called **Filter Width** (number of points averaged to determine correction data points).
- **Filter Width** must be in the range of 0.2 to 10.
- The higher the number the “flatter” the background fit will be, essentially more data points are used to describe the background (Figure 5).
- Correction can follow a portion of the peak if **Filter Width** and/or step size is small (Figure 6).
- Fits amorphous type artifacts in the background better with lower values of **Filter Width** (Figure 7).
- Used by DMSNT Background Correction program.



**Figure 5 Box Car Correction Filter Width 10**



**Figure 6 Box Car Background Correction Filter Width 0.2**



**Figure 7 Box Car Background Correction Filter Width 1.5**

6. DMSNT Background Removal, Smoothing, and Correction Program (see figures 1-5 in Screen Shots):

- Background Icon in DMSNT does all three, background correction, smoothing, intensity corrections.
- When using the Cubic Spline option make sure to zoom in on the background, maximum intensity in the range of 100 cps before picking background points.
- Creates a Net Intensity file that contains the background corrected raw data (and other optional operations) that is used in the Peak Finder program to determine peak parameters.
- Options for Data Smoothing have been discussed.

Correction Program:

Option for  $K\alpha_2$  Stripping from Diffraction Peaks:

- To eliminate the identification of  $K\alpha_2$  peaks.
- Do not use if profile fitting the data manually.
- Uses the “Rachinger Method” of  $K\alpha_2$  stripping to determine the  $I_{K\alpha_1}$  and  $I_{K\alpha_2}$  components of each diffraction peak.

Makes use of the known relationships between the  $K\alpha_1$  and  $K\alpha_2$  peaks:

$$I_{K\alpha_1} = 2 I_{K\alpha_2}$$

$$\Delta\lambda = \lambda_{K\alpha_2} - \lambda_{K\alpha_1}$$

From Bragg Equation:

$$\Delta 2\theta = 2 \tan \theta \frac{\Delta\lambda}{\lambda}$$

Choose intervals over which the correction will be applied to the data:

$$n \text{ intervals} = \Delta 2\theta / m \quad \text{where } m \text{ is an integer} \\ \text{typically } 1, 2, \text{ or } 3$$

Apply the following equation:

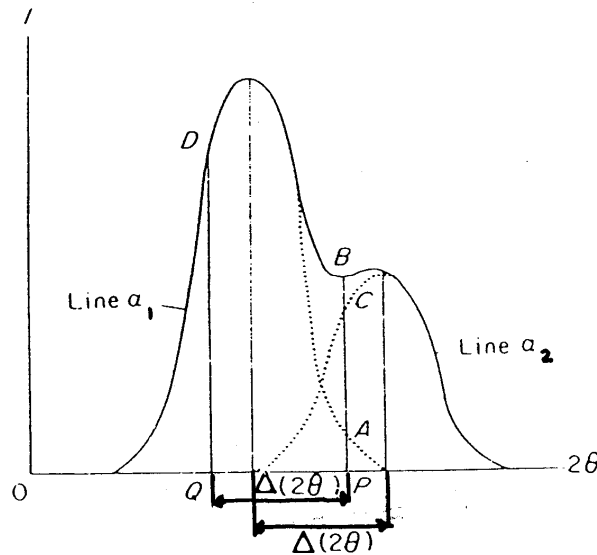
$$I(2\theta) = I_{K\alpha_1}(2\theta) + (1/2) I_{K\alpha_1}[2\theta + \Delta(2\theta)]$$

$I$  = intensity of experimental profile

$I_{K\alpha_1}$  = intensity due to  $K\alpha_1$  only

Sum over the interval  $n$

- Crystal monochromators typically influence the intensity ratio between  $K\alpha_1$  and  $K\alpha_2$  which can leave artifacts on the high angle side of the peak after stripping.



**Figure 8 Rachinger Method  $K\alpha_2$  Stripping**

Can also apply intensity correction for Lorentz Polarization Factor and constant beam area conversion:

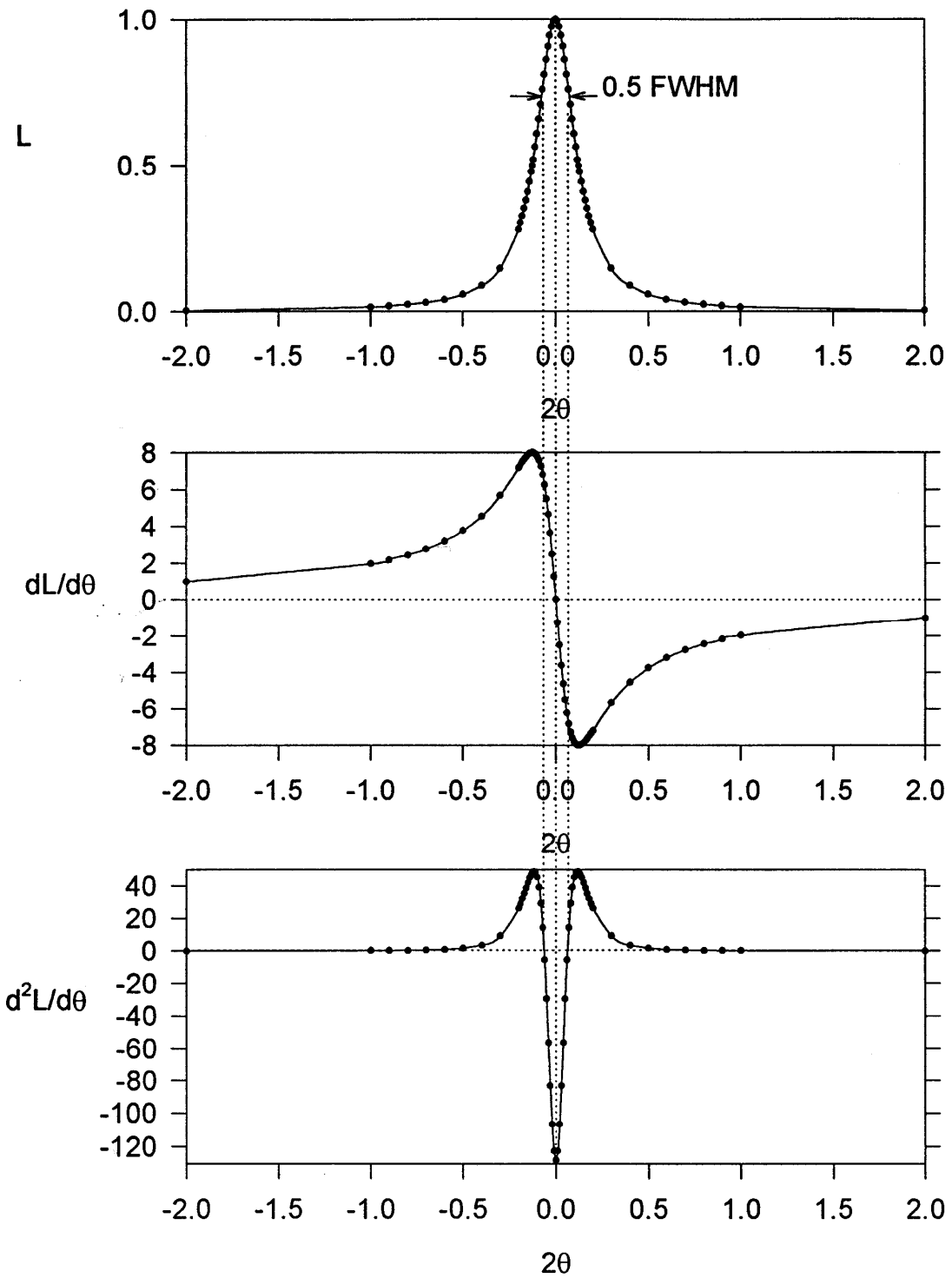
- Many quantitative calculations require a Lorentz Polarization Factor correction so this can be accomplished while creating the net intensity file.
- Constant beam area conversion for converting variable slit data to constant beam slit data for comparison. This does present quantitative problems.

### **Computer Algorithms for Peak Determination Routines:**

These methods create d (d-spacing) and I (intensity) files, or “peak files” which can contain the following information:  $2\theta$  and d-spacing of peak position, peak intensity (I), integrated intensity (area), FWHM, and relative intensities (%).

#### 1. Derivative Method:

- Typically use the 2<sup>nd</sup> order derivative method, more sensitive to small changes in the profile and inflection points or shoulders, other orders can also be used.
- Determine the peak position from most negative value of the 2<sup>nd</sup> derivative function of the peak.
- Can learn about the FWHM by the roots of the 2<sup>nd</sup> derivative function (where it crosses zero).
- Very sensitive to statistical noise in the data, usually requires a smoothing operation which can distort the peak data.
- Excellent at finding peak positions from peaks that overlap.
- Peak intensity is determined from raw data using the determined peak position.
- Attempts have been made to calculate the area by the negative region of the 2<sup>nd</sup> derivative function, but results are suspect.



**Figure 9 Derivative Method Peak Position**

## 2. Trend Oriented Peak Search:

- Determines the start of a peak by looking for an increasing trend in the slope of the average background, usually specified by the user, and determines the end of the peak by similar trends in the negative slope.
- Statistical noise can cause the algorithm to identify false peaks.
- More reliable results if data at top of the peak is reduced to 3 average data points.
- Peak position determined by the 3-point Parabola Method which fits a parabola to the “peak” of the diffraction profile.

### 3-Point Parabola Method:

Peaks can be represented by mathematical profiles, two common functions are Gaussian and Cauchy functions. Expanding both functions as a power series, showing only the first two terms, we have the general case:

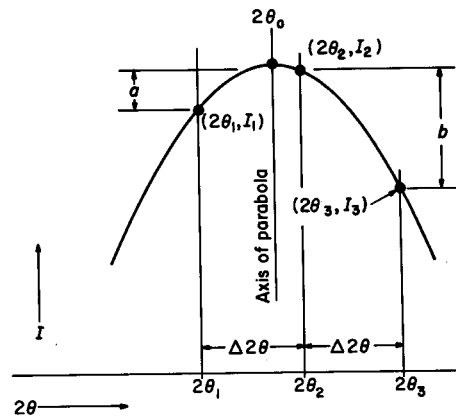
$$I(2\theta_i) = I_0 - (I_0/a^2)(2\theta_i - 2\theta_{\text{peak}})^2 + \dots\dots\dots$$

Ignoring higher order terms we have the equation of a parabola of the general form:

$$y = ax^2 + bx + c$$

If  $2\theta_1$ ,  $2\theta_2$ , and  $2\theta_3$  are 3 points that describe the peak and are separated by the angular interval  $\Delta 2\theta$  then the vertex (peak position) is given by:

$$2\theta_0 = 2\theta_1 + \frac{\Delta 2\theta}{2} \left[ \frac{3a + b}{a + b} \right]$$



**Figure 10 Figure Showing 3-point Parabola Method**

- Peak intensity is determined from the raw data point corresponding to the vertex of the parabola fit.
- Area of the peak is calculated by summing the net intensities for each data point (step size or chopper increment) over the region described by the trend search.
- Area values are sensitive to different choices of step size or chopper increment.
- FWHM determined by dividing the integrated intensity by the net peak intensity therefore this is an integral breadth not a true FWHM.
- Program can automatically separate  $K\alpha_2$  component only if they are resolved, based on the relationships between  $K\alpha_1$  and  $K\alpha_2$ .

### 3. DMSNT Peak Finder Routines (see figure 6 in Screen Shots):

#### a) Peak Finder using Digital Filtering:

- Locates the starting intensity of a peak above background by comparing the intensity of the raw data points with a defined value calculated by a user defined parameter know as the **Ripple Multiplier**.



- The **Ripple Multiplier** is a user adjustable value that is multiplied times the ESD (estimated standard deviation) of the background and added to the average background to produce the value used to locate the start of a peak.
- ESD's are produced for each data point in a raw data file based on the intensity of the data point and the preset count time used. This is determined by the square root of the total counts collected.
- Another user defined criterion is the **ESD Multiplier**. This is used to define the minimum intensity that a peak can have to be accepted. The **ESD Multiplier** times the ESD of the average background added to the average background gives the minimum intensity in cps that a peak must have to be accepted.
- Options available to correct for peak position errors using internal and/or external correction methods.
- Finds peak position by utilizing the “top 15 percent Parabolic Fit” which uses the top 15 percent of the data points, in terms of intensity, of the peak to determine the position

Top 15 Percent Parabolic Fit:

In general:

$$I_j = a + b\delta_j + c\delta_j^2$$

$\delta_j$  = increment between data points

Maximum intensity found by:

$$\frac{\partial(I_j)}{\partial(\delta_j)} = 0$$

a, b, c can be determined from a least-squares fit to the data by minimizing:

$$S = \sum_{j=-n}^n (a + b\delta_j + c\delta_j^2 - I_j)^2$$

the apex of the parabola is defined by:

$$2\theta_{\text{peak}} = 2\theta_0 - b/2c$$

$2\theta_0$  is the working origin initially chosen by the most intense data point.

This is found to be a more accurate method than the “3 Point Parabola Method” (provided more than 3 points).

Parabola must be a satisfactory fit to be acceptable, step size could have a large effect on the fit criteria.

- The FWHM is estimated from the parabola equation.
- The area is estimated by the following equation:

$$\text{Area} = (I_{\text{peak}} * \text{FWHM})/2$$

- The only reliable quantitative data obtained by this program is the peak position; all other parameters should be used with caution.

b) Peak Finder using Pearson VII Profile (see figures 7-9 in Screen Shots):

- This is time consuming especially if there are a large number of peaks.
- Program has difficulties with very broad peaks.
- **Background** tab is not available.
- **Peakfinder** tab - input peak finding information:

- **Use Existing Peakfile** – use data from **Peak Finder using Digital Filtering**.
  - Number of points for **Fourier smooth**.
  - Peak Searching Info: **Threshold** - pick minimum peak size, or **# of Peaks** – enter number of peaks to find.
  - Select general breadth of peaks.
- **Profile Fitting** tab select the type of profile, only two to choose from, and whether weighting of intensity is employed (higher intensity given more weight in the least squares minimization).
  - Uses values of 2 for exponent in the Pearson VII function and finds the best FWHM fit for all peaks, i.e. restricts the function used. Note in help menu reads, “Does not calculate area best to do in profile fit algorithm”; however, the area appears in the peak file using this method.

4. DMSNT Peak File Output – See Screen Shots Figures 10 and 11.

#### **Summary Comments on Peak Searching Programs:**

- Always check the results obtained graphically to see if they are reasonable.
- I recommend that results other than the peak position be used with caution, they should be reliable with results in a given scan but comparisons with other scans could possibly be suspect.
- Quantitative results for areas, etc., can be more accurately determined by profile fitting the diffraction peaks with known mathematical profiles. This procedure typically requires a lot more time to determine the results.
- Peak search routines work well for phase identification analysis.
- It is important to experiment with all of the variables available to determine which work best for your particular data.