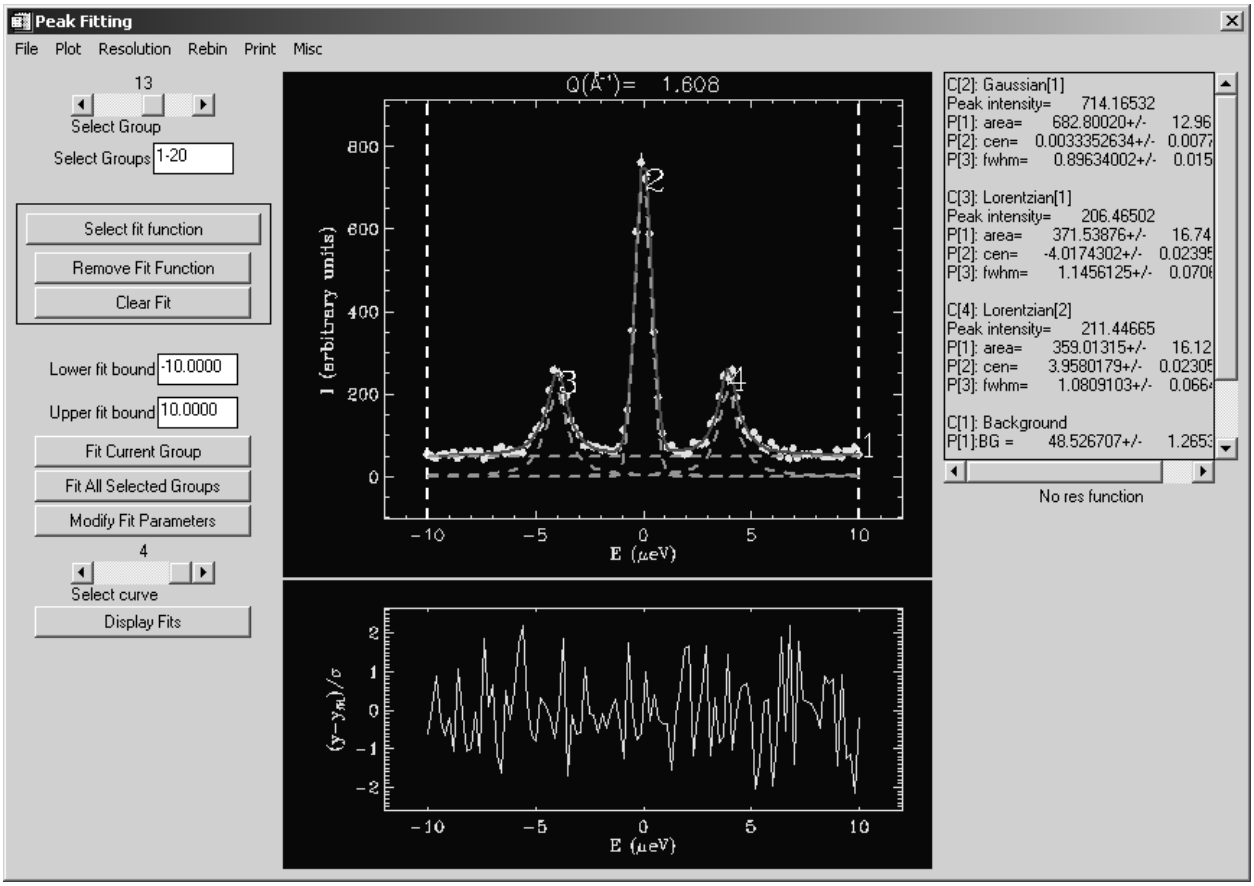
Peak Analysis: PAN
A general-purpose curve-fitting program in DAVE

A user's guide



Robert Dimeo and Zema Chowdhuri

March 1, 2002

1.1 Getting started with PAN

This documentation assumes that you have created a .DAVE file. This is explained in the other documentation related to data reduction for the inelastic neutron scattering spectrometers.

To start PAN select *Peak Analysis* from the *Data Analysis* menu in DAVE as shown in figure 1. PAN requires that the input data be in a particular format, denoted by a .DAVE extension in the filename. If your data has not been converted, you need to do so before you can proceed. Use the *Data Reduction* menu to choose the appropriate utility for the data that you have. Read the converted file into PAN by using the *Load Dave* command under the *File* menu (figure 2).

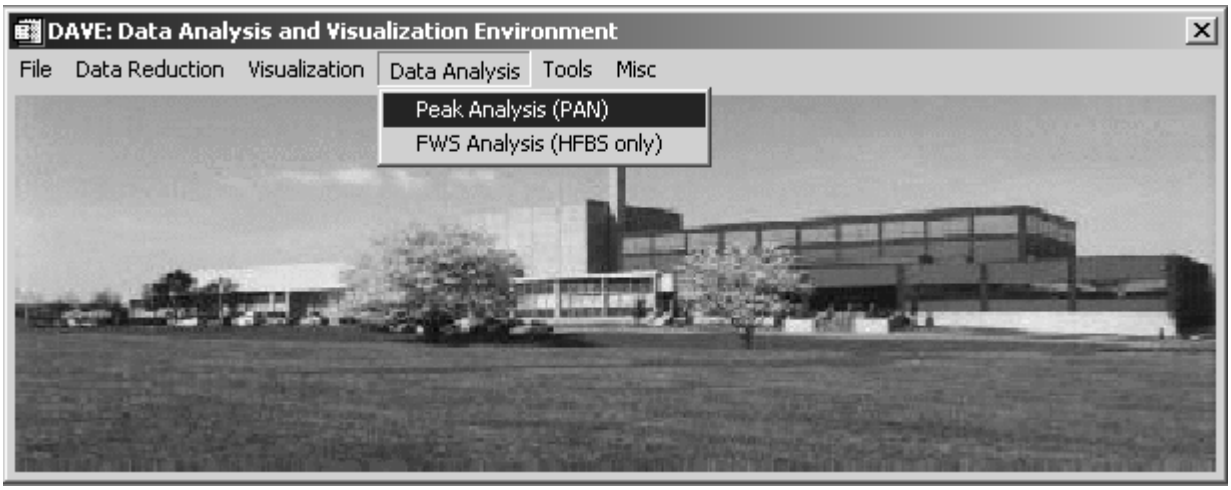


Figure 1 Selecting PAN from the DAVE application.

The data that results from the data reduction frequently is not regularly spaced in energy and, if you are going to use an instrumental resolution function in the fitting, it is necessary to rebin the data to produce a regular grid for the analysis. This can be done either using the *Rebin DAVE data file(s)* in the DAVE→TOOLS menu before launching PAN or you can select the *Rebin data* option from the REBIN menu within PAN. A dialog will launch with entries for the bin width and the energy range for the spectrum. The actual size of the bin is not important (as long as it is not less than the original binwidth).

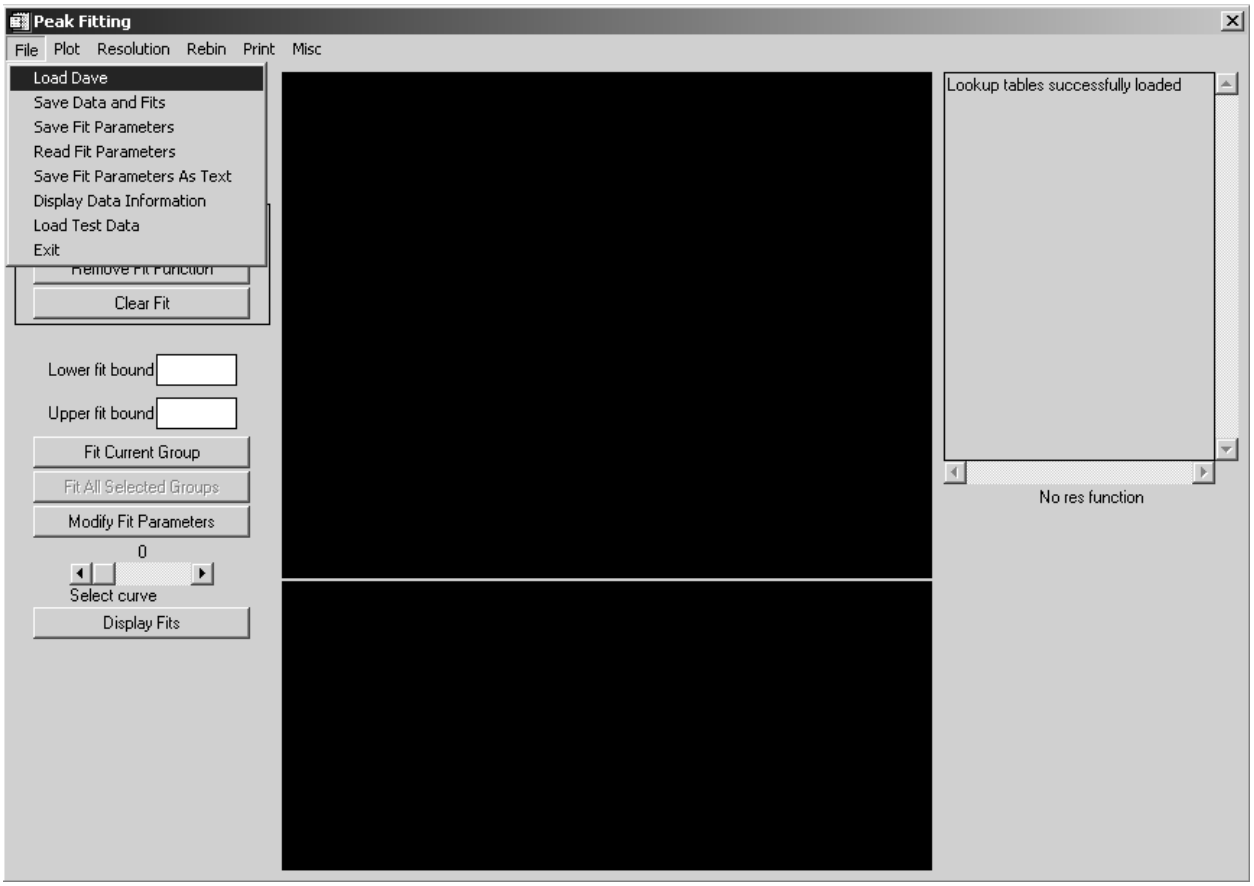


Figure 2 Selecting Load Dave from PAN.

1.2 Loading the resolution function

Load the resolution function using the *Load Res Function* under the Resolution menu. After you select the file, you will be prompted to limit the energy range of the data. The default is one-half the full range over which the data was collected.

Note that you must always load the data file first, then the resolution file. Both files must be in the DAVE format and have the same detector grouping.

1.3 Fit functions available in PAN

The functions available in PAN are:

Pseudo δ-function
Gaussian
Lorentzian
Lognormal
Voigt

Damped harmonic oscillator
Sloping background
Smooth step function
Methyl tunnel EISF
Three-site jump EISF
Two-site jump EISF
Kohlrausch function (Fourier transform of stretched exponential)

1.3.1 Pseudo-δ Function

Convolution in PAN is performed numerically. Unfortunately numerical convolution will not give a correct integrated intensity if there are few points defining one of the functions in the convolution product. This is encountered frequently in the analysis of quasielastic scattering where the overall lineshape is given by the sum of a Dirac delta function and a Lorentzian:

$$S(E) = I_o\left[A\delta(E) + (1-A)\frac{\Gamma}{\pi\left(E^2+\Gamma^2\right)}\right] \tag{1.1}$$

where I_o is the total integrated intensity of the lineshape, A is the fraction of the elastic scattering, and 2Γ is the full-width at half-maximum for the Lorentzian component. On a neutron spectrometer, one measures the convolution of this function with the instrumental resolution function. In terms of the resolution function, R(E), and the scattering function, S(E), this is expressed as a convolution product and integral,

$$\begin{aligned} I(E) &= R(E)\otimes S(E) \\ &= \int d\tilde{E}\, R\left(E-\tilde{E}\right)S\left(\tilde{E}\right). \end{aligned} \tag{1.2}$$

Since A provides valuable information on the geometry of motion, it is critical to be able to extract it reliably. In order to extract the integrated intensity of the elastic scattering component one must be able to represent the delta function somehow.

In order to circumvent this limitation of numerical convolution one can invoke a *pseudo-δ function* in which two approximations are made. First, a boxcar

approximation of the instrumental resolution function, R(E), is made. We only have knowledge of the resolution function sampled at discrete points, E_i. Thus we only really know $R_i = R(E_i)$. So our first approximation entails redefining the resolution function as a constant function of energy as

$$\widetilde{R}(E)=\sum_{i=1}^{N} R_i\left[\Theta\left(E-\left(E_i-\frac{\Delta}{2}\right)\right)-\Theta\left(E-\left(E_i+\frac{\Delta}{2}\right)\right)\right], \qquad (1.3)$$

where $\Theta(E)$ is the Heaviside step function and Δ is the interval between points in the resolution function. We will refer to $\widetilde{R}(E)$ as the *boxcar approximation* of the resolution function, R(E), given the resolution function data, $R(E_i)$.

The second approximation is to represent the delta function as a narrow gaussian curve. In the limit of small standard deviation, σ, the gaussian approaches a delta function:

$$\widetilde{\delta}(E;\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{E}{\sigma}\right)^2} \qquad \lim_{\sigma\to 0}\widetilde{\delta}(E;\sigma)=\delta(E) \qquad (1.4a,b)$$

The convolution between the pseudo delta function, $\widetilde{\delta}(E)$, and the boxcar approximation to the resolution function, $\widetilde{R}(E)$, can be performed analytically:

$$\widetilde{\delta}(E)\otimes\widetilde{R}(E)=\frac{1}{2}\sum_{i=1}^{N} R_i\left[\operatorname{erf}\left(\frac{E-E_i+\frac{\Delta}{2}}{\sqrt{2\sigma^2}}\right)-\operatorname{erf}\left(\frac{E-E_i-\frac{\Delta}{2}}{\sqrt{2\sigma^2}}\right)\right], \qquad (1.5)$$

where

erf(E) is the standard error function,

$$\operatorname{erf}(E)=\frac{2}{\sqrt{\pi}}\int_0^E dx\, e^{-x^2}.$$

Note that this requires a constant interval between each data point in the resolution function.

Description of PAN Fit Functions

Gaussian

$$f(E;A,E_0,W)=\frac{A}{\sqrt{2\pi(W/2.354)^2}}e^{-\frac{1}{2}\left(\frac{E-E_0}{(W/2.354)}\right)^2}$$

A: integrated area
E_o: mean of distribution
W: full-width at half-maximum

Lorentzian

$$f(E;A,E_0,W)=\frac{W}{2\pi}\frac{A}{(E-E_0)^2+(W/2)^2}$$

A: integrated area
E_o: mean of distribution
W: full-width at half-maximum

Lognormal

Distribution function for a random variable whose natural log is *normally* distributed. Note that this distribution uses absolute values so that it is symmetric about E = 0.

$$f(E;A,E_0,\sigma)=\frac{A}{2|E|\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{\ln(|E|-|E_0|)}{\sigma}\right)^2}$$

A: integrated area
E_o: mean of underlying gaussian distribution
σ: standard deviation of the underlying gaussian distribution

You can use these parameters to get the mode, E_{mode}, and standard deviation, SD, of the lognormal distribution via:

$$E_{mode}=\exp\left(E_0+\frac{\sigma^2}{2}\right)$$

$$SD=\sqrt{\exp\left(2E_0+\sigma^2\right)\left(\exp\left(\sigma^2\right)-1\right)}$$

The mode and standard deviation are actually displayed whenever you choose to display the fit parameters.

Voigt

The voigt function implemented here is a convolution of a lorentzian and a gaussian, both centered at E_0, with identical full-widths at half maximum

A: integrated area
E_o: center of distribution
W: full-width at half maximum of the lorentzian and gaussian components

---

Damped Harmonic Oscillator

$$f(E;A,E_0,W) = N(E)\frac{AW}{2\pi}\left[\frac{1}{(E-E_0)^2+(W/2)^2} - \frac{1}{(E+E_0)^2+(W/2)^2}\right]$$

A: amplitude of function (NOT AREA!)
E_o: center of each Lorentzian component
W: full-width at half maximum of the lorentzian component
N(E): Bose occupation factor defined as

$$N(E;T) = \left(1-\exp\left(-\frac{E}{kT}\right)\right)^{-1}$$

T: Temperature in Kelvin

Smooth Step Function

This function is sometimes more useful than a sloping background if the measured background has a smooth transition from one flat level to another.

$$f(E;A,E_0,W) = \frac{A}{2}\left(1+\mathrm{erf}\left(\frac{E-E_0}{\sqrt{2(W/2.354)^2}}\right)\right)$$

A: Amplitude of transition
E_0: location of midpoint of transition
W: full-width at half maximum of the transition

Sloping Background

$$f(E;m,b) = mE + b$$

m: slope
b: y-intercept

Methyl Tunnel EISF

For a methyl group undergoing reorientation via tunneling the neutron scattering lineshape is given by

$$S(Q,E) = f(Q;r)\delta(E) + \frac{(1-f(Q;r))}{2}[\delta(E-E_t) + \delta(E+E_t)]$$

where E_t is the tunneling energy and f(Q;r) is the elastic incoherent structure factor (EISF) given by

$$\boxed{f(Q;r) = \frac{5 + 4j_0(Qr\sqrt{3})}{9}}.$$

r: radius of the methyl group (C-H distance)

Three Site Jump EISF

For a methyl group undergoing jump diffusion among three equivalent sites the neutron scattering lineshape is given by

$$S(Q,E) = f(Q;r)\delta(E) + (1-f(Q;r))L(E;2\Gamma)$$

where $L(E;\Gamma)$ is a lorentzian lineshape with a FWHM of 2Γ and f(Q;r) is the elastic incoherent structure factor (EISF) given by

$$\boxed{f(Q;r) = \frac{1 + 2j_0(Qr\sqrt{3})}{3}}.$$

r: radius of the methyl group (C-H distance)

Two Site Jump EISF

For a methyl group undergoing jump diffusion among two equivalent sites the neutron scattering lineshape is given by

$$S(Q,E) = f(Q;r)\delta(E) + (1 - f(Q;r))L(E;2\Gamma)$$

where L(E;Γ) is a lorentzian lineshape with a FWHM of 2Γ and f(Q;r) is the elastic incoherent structure factor (EISF) given by

$$\boxed{f(Q;r) = \frac{1 + j_0(Qr)}{2}}.$$

r: jump distance

Kohlrausch Function

$$\boxed{K(E;\beta,\tau) = \frac{A}{\pi}\int_0^\infty dt \cos\omega t \exp\left(-\left(\frac{t}{\tau}\right)^\beta\right)}$$

A: integrated area
β: stretching exponent
τ: relaxation time

Actually the fit routine uses energy, E = ħω for the independent variable, and the average relaxation time, $\bar{\tau}$, for display and fitting purposes where it is defined as

$$\bar{\tau} = \frac{\tau}{\beta}\Gamma\left(\frac{1}{\beta}\right).$$

2.1 Analyzing data with PAN: An example

We can analyze some simulated neutron scattering data to show how to use PAN in some detail. In particular, we will load in the test data set, fit an appropriate model to the data, and extract the Q-dependence of some important parameters.

First launch PAN if it is not already running. Next, select File→Load test data. In a few moments a data set that contains a large central peak corrupted by noise should appear. Also, the Select group text field should have been updated to read "1-20" indicating that there are now 20 detector groups (or Q-values). You can move the slider control labeled *select group* and the spectrum displayed in the window will change. As you move to larger detector numbers, the central peak will be accompanied by two satellite peaks. This data simulates a tunneling spectrum taken on a backscattering instrument. Move the slider control so that you are viewing group (detector) 16. Click the right mouse button so that you autoscale. If you wish, you can hold the left mouse button down and drag a rubberband-type zoom box across the plot to enlarge a particular portion of the data window. A right mouse click will always result in zooming out all the way.

We will fit this group first and then fit the remaining groups automatically. The data looks like it has 4 components. The central peak looks approximately Gaussian, the two satellite peaks are approximately Lorentzian, and there is an underlying offset (a background level). Choose *sloping background* from the drop-menu called *Select fit function*. Move the cursor into the plot window and hold down the left mouse button. A dashed green line should appear across the extents of the data plot. As you move the cursor up and down in the data window while holding down the left mouse button, the green line should follow the cursor. This action is used to specify the offset of the overall model function. Once you are satisfied with the level, release the left mouse button. Now hold down the left mouse button again and move the cursor around the data window. You should see the slope of the line change and pivot about the midpoint of the x-axis. Once you are satisfied with the slope, release the left mouse button. For this function, the slope should be quite small.

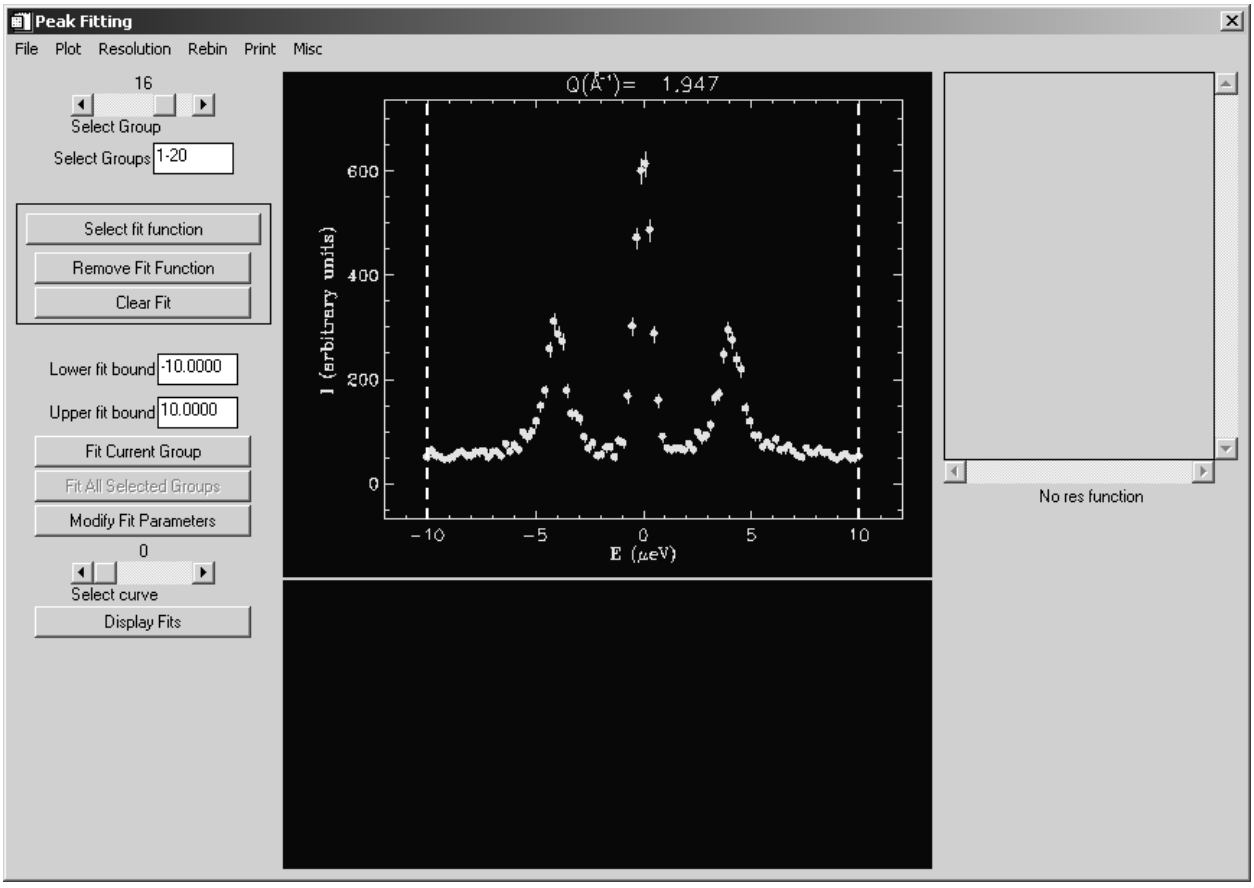


Figure 3 Test data in group 16.

Next we will add a Gaussian curve to model the central peak. Choose *Gaussian* from the *Select fit function* menu. Move the cursor over the plot window and hold down the left mouse button. A Gaussian should follow your cursor movements both in location and in height. Change the vertical position to change the Gaussian peak height and change the horizontal position to change the Gaussian peak center. Once you are satisfied that the peak has the correct amplitude as that of the central data peak and is located at the correct position, release the left mouse button. Next hold down the left mouse button and move the cursor horizontally, thus changing the width of the curve. Once you are satisfied with the width release the left mouse button.

Next add a Lorentzian to fit the satellite peak to the left of the central peak. The method to change the amplitude, center, and width of the Lorentzian is exactly the same as for the Gaussian. Finally, add another Lorentzian to model the satellite peak to the right of the central peak.

You have just specified the initial guesses for the fit using the mouse and now it is time to fit the model to the data. Simply press the button named *Fit current group*. In a moment you should see the result of a least-squares fit of the model to the data. Your final answer may be different but mine is shown below in figure 4.

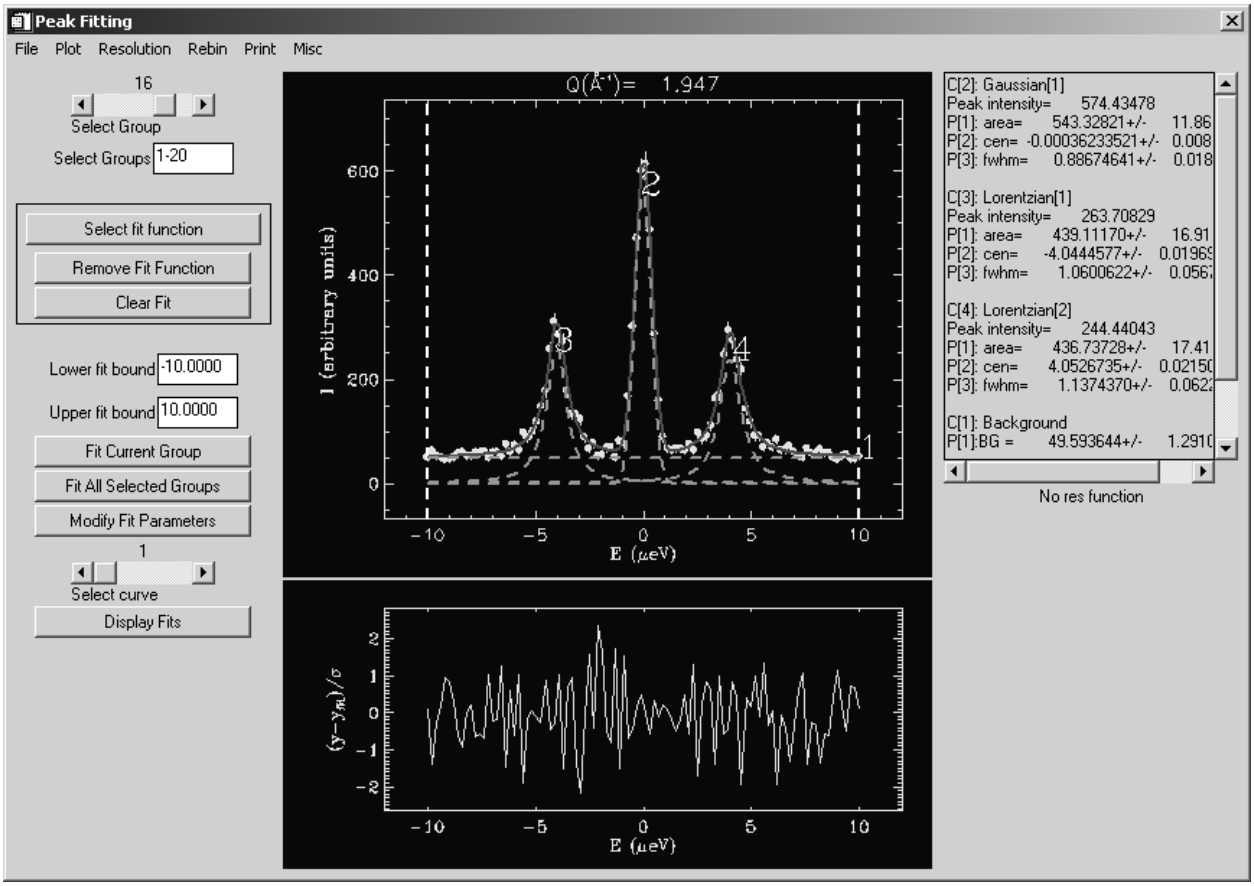


Figure 4 Fit of model to data in group 16.

The model parameters with uncertainties are found in the text panel displayed to the right of the data window. Note that each curve is labeled as in the window with a number. For instance, curve 2 above is the Gaussian, listed in the text panel as C[2]. Curve 4, C[4], is the Lorentzian located to the right of the central peak. Note also that the curves are labeled in the data window. The parameters for each curve component are labeled in a similar manner. For instance, the peak center for the Gaussian is listed in the text panel as P[2] of C[2]. This gives us a means to identify any of the parameters from any of the curves present in the fit.

The next part of the analysis we want to do is to fit all of the detectors. However we need to be careful at this stage. Scroll the *Select group* slider control down to group 1 and notice how small the satellite peaks have become. There is a possibility that the least-squares routine will have trouble fitting this same model to this data. Therefore we would like to impose a constraint on the area and width of the satellite peaks. Select the left Lorentzian by moving the *Select curve* slider control to 3 (recall that this is the curve labeled C[3]). Next press the button labeled *Modify Fit Parameters*. A dialog like the one shown in figure 5 below will appear. You should see only a single box checked (lower bound for the full-width at half-maximum, FWHM). Click the box for the low constraint on the area. This will prevent the area of the Lorentzian from going below zero. Also click the high constraint for the FWHM and type 3.0 in the text box. Press the button labeled *Exit* to get back to the Peak Analysis main window. What you have done is impose a lower limit on the area of the Lorentzian and an upper bound of 3.0 on the FWHM of the Lorentzian.

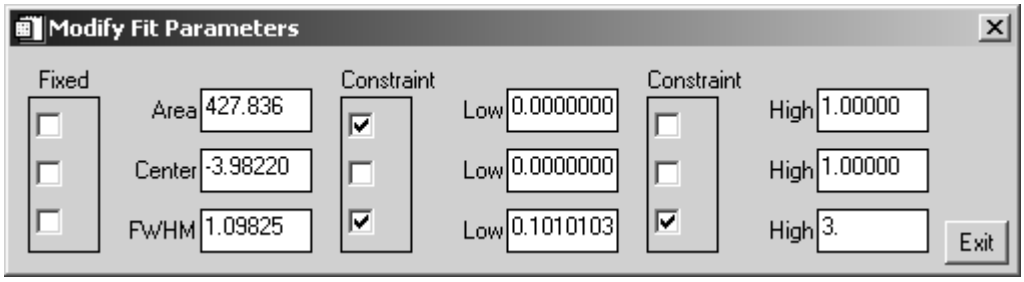


Figure 5 Dialog to modify fit parameters, impose constraints, and fix parameters.

Repeat the same process for the other Lorentzian, C[4]. When you have finished this, press the button labeled *Fit all selected groups*. This button tells PAN to fit all of the detectors listed in the text box labeled *Select groups* (e.g. 1-20) with the same model as the initial fit, in our case it is group 16. You should see fits being performed for each group starting with group 1 and finishing with group 20. Once it is finished you can change the *Select group* slider to inspect each fit. Whenever a new group is displayed with its fit, its fit parameters and uncertainties are also displayed in the text panel to the right of the data window for you to examine.

Now we would like to look at the dependence of the peak position of the Lorentzian to the right of the central peak on the group number (i.e. the Q-dependence of the peak position). First we should note which parameter and curve this corresponds

to in terms of our notation discussed above. The right Lorentzian is C[4] and the center of the peak is specified by P[2]. Next, select the menu option from the application menu bar *Misc→Enter curve and parameter to plot*. A dialog will pop up that asks you for the curve and parameter to plot as shown in figure 6. The default is parameter 2 of curve 1, denoted C[1].P[2]. We actually want to replace this with C[4].P[2]. Type this in and press the button labeled *done*. PAN will then extract this parameter from the fits in detectors 1-20 and pop up the results in a new plot window as shown in figure 7. You can save these parameters and errors in a three-column formatted ASCII file, print to a postscript file, zoom into the window using the same mouse operations as in the main PAN application, and/or exit the plot utility.

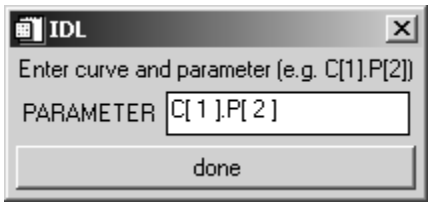


Figure 6 The dialog for plotting a particular parameter.

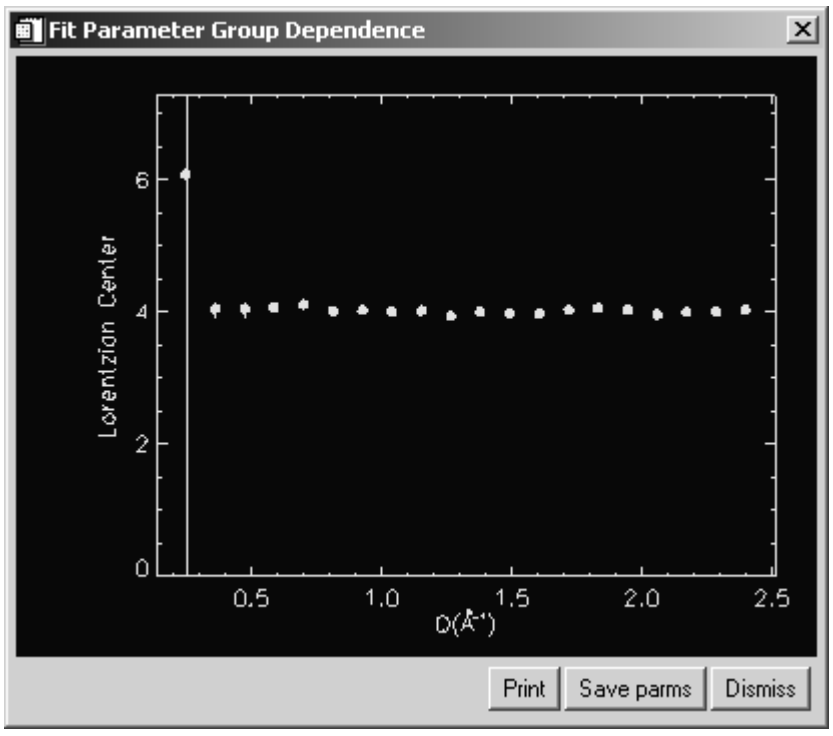


Figure 7 Group dependence of the center of the right Lorentzian.

The final part of the analysis is to plot the elastic-incoherent structure factor (EISF) which is defined as the ratio of the integrated elastic intensity to the total integrated intensity. This is straightforward to calculate and PAN has a function built-in to do it. Choose *Misc→Plot EISF* from the main menu bar. You will get a dialog box like the one shown in figure 8 below.

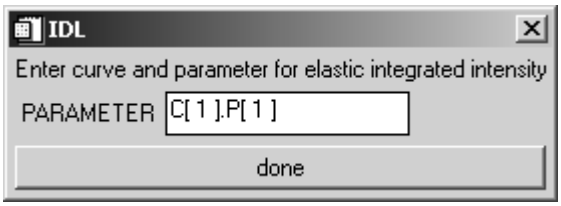


Figure 8 Dialog for entering the parameter for the elastic integrated intensity.

For the elastic intensity we want the area of the Gaussian component, C[2].P[1]. The next dialog box that pops up is shown in figure 9.

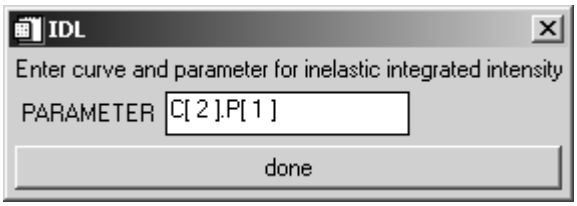


Figure 9 Dialog for entering the parameter for the inelastic integrated intensity.

This should be the area of the left Lorentzian, C[3].P[1]. Next a dialog asks if you wish to add another inelastic integrated intensity. The answer is yes since we have two inelastic peaks. Press *yes* and enter C[4].P[1] in a dialog box identical to the one shown in figure 9. When you are asked to add another inelastic feature, press *no*. The final question you are asked is if you want to subtract an inelastic feature from the denominator…press *no*. Finally you should see the EISF plotted in a plot utility similar to that shown in figure 10. This has all of the same functionality as that shown in figure 7.

You may choose to examine the Q-dependence of any of the fit parameters by simply specifying the curve and parameter using the notation discussed above and using the menu bar selection *Misc→Enter curve and parameter to plot*.

Note also that the fit parameters can be saved to a binary file to be restored in a future session by selecting *File→Save fit parameters*. Furthermore you can save

the currently-displayed data and fit in ascii form by selecting *File→Save data and fits*.

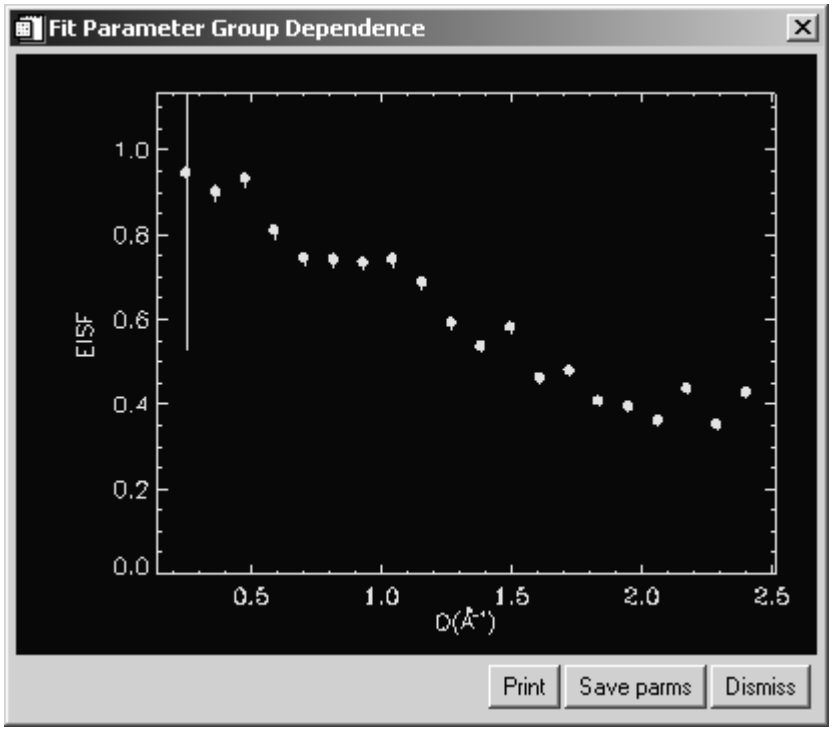


Figure 10 Plot of the EISF based on fits to all of the detector groups.