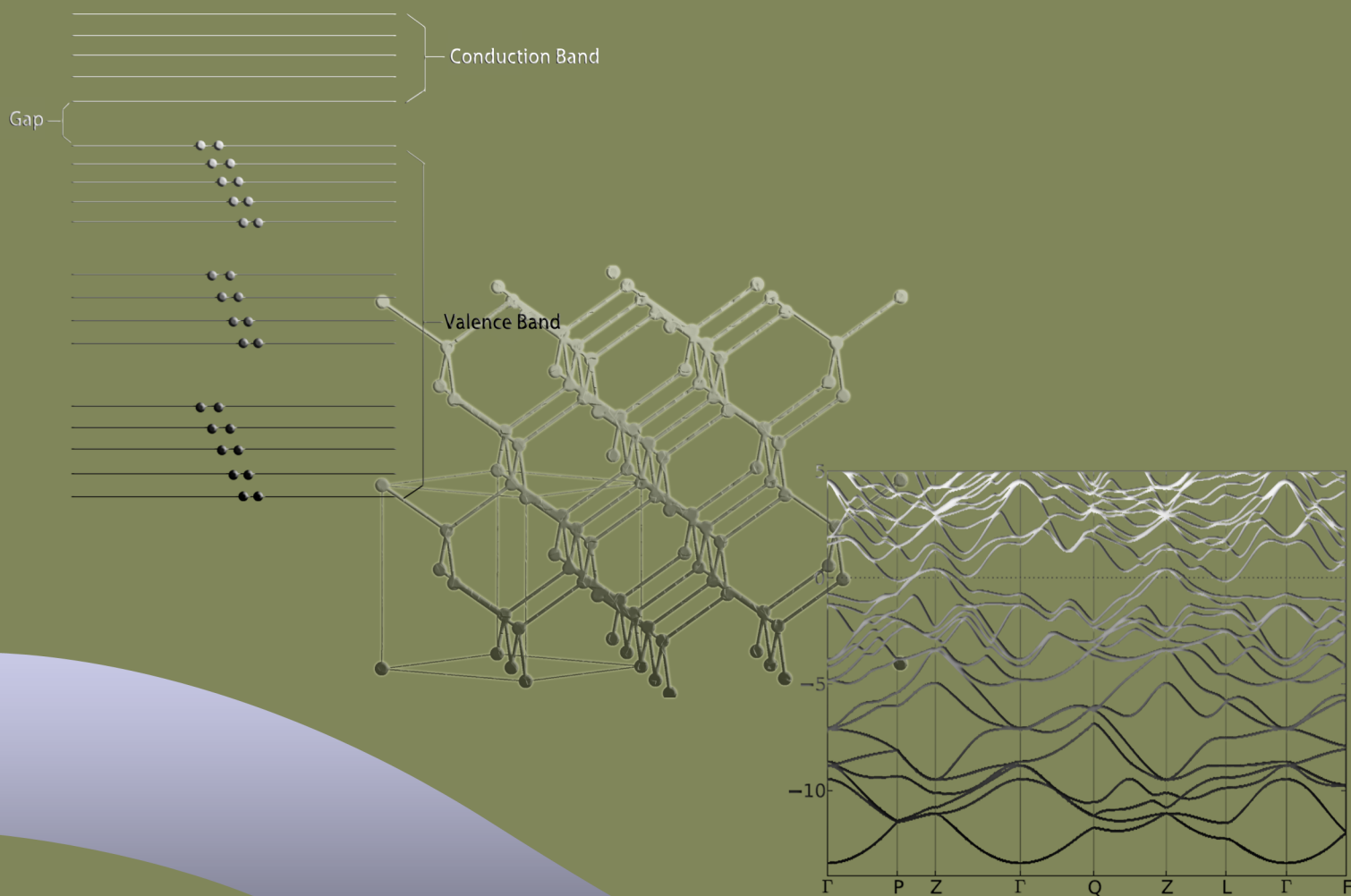


# Queen Mary, University of London

## School of Physics and Astronomy

### XAFS数据处理相关问题



## Background Subtraction

在XAFS出具处理过程中，对于得到的 $\mu(E)$ 吸收谱，在进行E-k转换之前，需要首先进行背景扣除（background subtraction）及归一化（normalization），就是从得到的吸收谱中，扣除孤立原子产生的吸收谱，只保留由于周围原子作用而产生的波动部分，同时对实验过程中其他因素导致的吸收谱变化进行归一化处理，首先给出背景扣除及归一化的公式：

$$\chi(E) = \frac{\mu - \mu_0}{\Delta\mu_0} \quad (0-1)$$

需要指出的是，这里之所以选择 $\Delta\mu_0$ 而不是 $\mu_0$ 作为分母，是考虑到了使用绝对值作为分母时，有时候会出现吸收系数为负的情况（由 $\mu = \ln \frac{I_0}{I_t}$  — 透射模式吸收系数的计算公式— 可知，吸收系数为负的情况下，透射X-ray强度大于入射强度，这主要是由荧光过强导致的），此时如果使用绝对值作为分母则无法统一衡量边前— pre-edge与边后— post-edge的吸收谱变化规律，而使用边前边后的差值 $\Delta\mu_0$ 作为分母则解决了这一问题，在Athena中，具体的实现方法参见图1：给定吸收边前与吸收边后背景吸收的拟合曲线（具体的实现算法此处不做介绍，

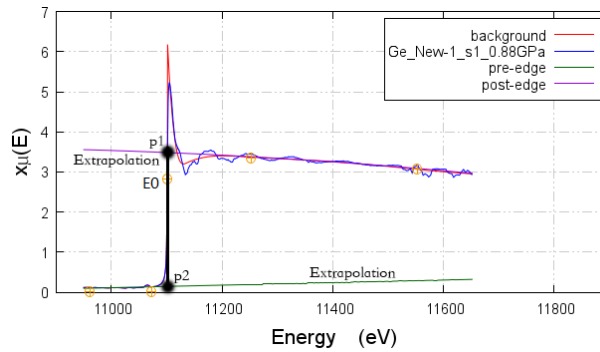


Figure 1 XAFS背景扣除

详见Athena Document相关部分），再给定吸收边 $E_0$ 的位置，将边前与边后背景曲线分别外延至吸收边 $E_0$ 的位置分别得到 $\mu_{p2}$ 与 $\mu_{p1}$ ，则 $\Delta\mu_0$ 由式： $\Delta\mu_0 = \mu_{p1} - \mu_{p2}$  给出。

## A problem about E-k transformation

When using EXAFS analysis, E-space spectrum is transformed to k-space spectrum following:

$$k = \sqrt{\frac{2m(E - E_0)}{\hbar^2}} \quad (0-2)$$

Hence by default, the following analysis in k-space (and further transformation to R-space through Fourier transform) does not consider the pre-edge spectrum characters. However, this is just what EXAFS is about because the energy corresponding to EXAFS area of the absorption spectrum is quite high, which makes the absorption spectrum independent of energy level features just around the final state (this is, in turn, just what pre-edge spectrum is about).



### Selection of k-range for Fourier Transform

When doing the Fourier transform to get spectrum in R-space from k-space, the selection of k-range does make obvious difference to the result. As can be seen from Fig. 2., when the width of window function is too small, the wiggles is more obvious which could make it dominate the Fourier transform result, thus influence the determination of real absorption spectrum in R-space. Moreover, when the k-range for Fourier transform is too small, there could exist more

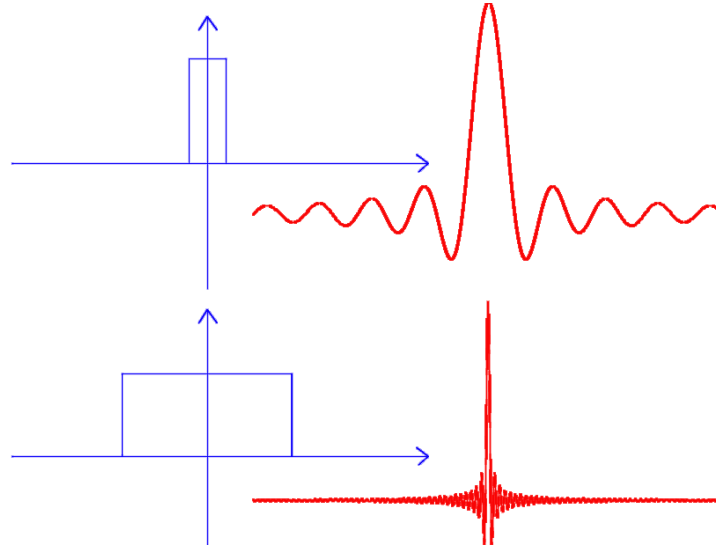


Figure 2. The influence of different window function width.

uncertainties within the k-space spectrum, which means there could exit more possibilities of frequency combination which could account for the k-space spectrum. Actually, it is the same with the understanding above, basically. However, if the k-range is set too large, there are still some problems: more noise (especially the noise at the end of the spectrum which could be dominant) becomes dominant. So the best way to determine the k-max is to first use largest k-weight value (to make the signal at the end of spectrum more obvious) while changing the k-max value at the same time until there is no obvious artificial wiggles within the Fourier transform pattern. For the selection of k-min value, it should be noticed that near the absorption edge, the approximated background has a jump due to the sharp change of data in that area. If k-min is set to near 0, the Fourier transform then takes the vibration in near-edge range into consideration. However, when doing the background subtraction, the jump of approximated background is then taken as part of the vibration of signal, which is obviously not correct information. This is why the [tutorial](#) says 'to select k-min to a value where background does not vibrate much', because otherwise the wiggle from approximated background brings false information.

$R_{bkg}$  parameter



When subtracting the background from signal, there is a quite important parameter in Athena –  $R_{bkg}$ , which refers to the largest angular frequency (distance in R-space) for background. In another word, in k-space, there is a wavelength ( $k_{not-min}$ ), and any wavelength larger than  $k_{not-min}$  could possibly contribute to the background. Any vibration with shorter wavelength is then taken as real signal. That means the ' $R_{bkg}$ ' defines the largest possible frequency for background approximation. So in k-space, when sampling for spline approximation (a method to approximate the curve within individually divided area, refer to [Wikipedia](#)), the sampling interval  $2dk$  (the factor 2 here comes from the expression for EXAFS, in which the sine function contains ' $2kR$ ' term, refer to [tutorial](#)) should follow the relationship as required by [sampling theorem](#):

$$2\pi \frac{1}{2dk} > 2R_{bkg} \quad (0-3)$$

Then we have:

$$dk < \frac{\pi}{2R_{bkg}} \quad (0-4)$$

Hence the number of  $k_{nots}$  should be:

$$k_{nots} \geq \frac{\Delta k}{dk} = \frac{2\Delta k R_{bkg}}{\pi} \quad (0-5)$$

where  $\Delta k$  is the range of selected area for Fourier transform in k-space. If the ' $R_{bkg}$ ' value is set too low, more noise will be taken as real signal. If too high, part of the real data can then be taken as noise, or to say, if ' $R_{bkg}$ ' is set too high, the background has more freedom to go around possible values which will assemble the vibration mode of real data. So it requires more experience than other parameters setting, and quite often people use half of the nearest neighbour distance as the best choice for ' $R_{bkg}$ ' value.

### The function for correlated Debye Model

$$\sigma^2 = \frac{\hbar}{2\mu} \int \frac{2\omega^2}{(\frac{k_B\Theta_D}{\hbar})^3} \left[ 1 - \frac{\sin(\frac{\hbar\omega R_j k_D}{k_B\Theta_D})}{\frac{\hbar\omega R_j k_D}{k_B\Theta_D}} \right] \coth\left(\frac{\hbar\omega}{2k_B T}\right) \frac{d\omega}{\omega} \quad (0-6)$$

In equation (0-6),  $\omega$  refers to the vibration frequency within the lattice.  $\Theta_D$  is the Debye Temperature of specific material, and in FEFFIT fitting,  $\Theta_D$  can be chosen as one of the fitted parameters.  $R_j$  is the half length of scattering path (in FEFFIT fitting).  $k_D = (\frac{6\pi^2 N}{V})^{\frac{1}{3}}$  is the Debye wave vector.

### Relation between heat capacity and Debye Temperature

$$c_V = \frac{12\pi^4}{5} n k_B \left( \frac{T}{\Theta_D} \right)^3 \quad (0-7)$$

Here,  $n = N/V$  is the number of atoms per unit volume.



### The problem of Muffin-Tin approximation

The use of constant interstitial potential, i.e. the muffin-tin approximation, introduces problem especially when the kinetic energy of photo electron is close to the constant interstitial potential used. So for the calculation of XANES spectrum using *fmdnes* package, parameter named 'V0imp' should be changed for the purpose of calibration compared to experimental reference spectrum. The 'V0imp' parameter here concerns the reference of photoelectron kinetic energy, which is just the muffin-tin interstitial potential.

### Phase Correction in Athena

After background subtraction and the following Fourier transform, the experimental EXAFS spectra turns into R-space which can give us local bonding information. However, the obtained local bonding signal may not exactly correspond to theoretical value, there will be always a shift compared to corresponding theoretical value. In Athena program, there is a checkbox with which we can determine whether to do the phase correction to let us obtain the 'correct' spectra in R-space. So the question comes that how the program knows what the right value should be before we do the phase shift, scattering amplitude and other theoretical calculation (This is what FEFF does, but Athena is not a program doing such calculation)? The answer is the Athena program embeds the phase shift result (and all other theoretical result needed to build up theoretical EXAFS spectra) from FEFF8 calculation as text document. When box for phase correction is checked, the Athena program then reads the phase shift results to build up a semi-theoretical spectra and then obtain a 'correct' pattern in R-space. The experimental result is then shifted to that semi-correct position in R-space. Here the word 'semi' is used because only the phase shift theoretical value of absorption atom is used (for all atoms) to calculate theoretical spectra. The calculated result is obviously not the real correct ab initio spectra! And this is the reason why many theorists do not agree the author of Athena to add such 'useless' function in the program. (refer to the webpage link about [phase correction](#).)

### The extraction of coordination numbers in Artemis

In EXAFS equation, we have two parameters which can reduce the absorption coefficient (the amplitude). One is the coordination numbers and the other one is passive reduction parameter  $S_0^2$ . However, when we got the experimental spectra and put them into any programs, we then can never tell the reduction of absorption amplitude comes from the reduction of coordination numbers or just from  $S_0^2$ . Technically we call it the correlation between coordination number  $N$  and passive reduction parameter  $S_0^2$ .

Then how do we extract the coordination number, e.g. in Artemis? There is online tutorial by [Bruce Ravel](#), etc. telling us how to do it and here we summarize briefly in this part. Basically, we should firstly have the experimental EXAFS spectrum for standard reference (for whatever material we have, e.g. Ge, Si, Au, etc.). Then we fit the EXAFS spectra for our sample and

the reference in Artemis, respectively. For the fitting, we could fix the coordination numbers for our sample as exactly the same value with that of reference. And we can compare the fitted  $S_0^2$  value for the reference ( $S_{0Ref}^2$ ) with that for our sample ( $S_{0Sam}^2$ ) to extract coordination numbers ( $N_{Sam}$ ) for our sample following:

$$N_{Sam} = N_{Ref} \frac{S_{0Sam}^2}{S_{0Ref}^2} \quad (0-8)$$

in which the  $N_{Ref}$  refers to the theoretical coordination numbers of corresponding shell. Actually there are quite a few ways to extract the coordination numbers, for example we can introduce math expression for  $S_0^2$  parameter to include the coordination numbers in it, and then we can directly obtain the coordination numbers for our sample through the fitting result. For details about this method, refer to the online tutorial from [Scott Calvin](#). It should be mentioned that no matter which method we choose for the extraction of coordination numbers, the basic idea is exactly the same, and one can feel this basic idea from equation-(0-8).

### The restraint fitting in Artemis

For some cases of EXAFS data fitting using Artemis, when we release all parameters to let Artemis fit them to give us results, some unexpected or non-physics fitted value may pop up. This should be due to the way that Ifeffit uses to determine the best-fit values. It is essentially the sum of the difference of squares (chi-square) between the data and the theory evaluation results. And the best-fit values are those minimize chi-square. However, when we have many parameters released at the same time for Artemis to fit, the strong correlation (basically, it is the influence of changing one parameter value on another parameter value) may not allow simultaneous minimization for different parameters. For example, sometimes when we release the passive amplitude factor  $S_0^2$  to fit, we may get very low value even for standard crystalline sample for which the  $S_0^2$  value should be around 1. And also sometimes, we could even obtain negative value for  $S_0^2$  or  $\sigma^2$  (Debye-Waller factor), which is obviously non-physics value. In such cases as discussed above or sometimes we have an expectation (either from the literature or our experience) before the fitting, we may need to 'restrict' the fitted parameters value to let it to go following the way that we want them to. Of course, before we decide to do the 'restricted' fitting, we should be definitely confident in our expectation for parameters value.

There are two ways to do the 'constraint' or 'restraint'. The first one is through using the 'def' parameter type when we assign property (e.g. 'guess', 'set', 'def', 'restrict', etc.) for our parameters. Let's say we put 'b' in our 'amp' input box to account for the passive amplitude factor  $S_0^2$ , and this time will define our parameters following:

$$\begin{aligned} \text{guess } a &= 0.9 \\ \text{def } b &= \min(\max(0.7, a), 1.0) \end{aligned}$$



Here the expression for  $b$  actually means that we manually constraint the value for  $b$  to be between 0.7 and 1.0 while we guess parameter  $a$ . In another word, if the guessed value for  $a$  is indeed between 0.7 and 1.0, then we accept the fitted  $a$  value as our  $b$ , if it is not, then we take 0.7 for parameter  $b$  (i.e. our  $S_0^2$ ). And it should be noticed that there are lots of ways to define relationships between parameters, for example if we have two different kinds of first-shell atoms (atom X and Y) around the absorption atom and we want to evaluate the weight of contribution from atoms X and Y. The direct way to do this is to float an  $S_0^2$  variable and then use a weighting parameter to adjust the relative amounts of atoms X and Y:

```
guess  amp = 0.9
guess  frac = 0.5
def    amp_x = amp * frac
def    amp_y = amp * (1 - frac)
```

Another way is a bit clever, and in this case we put 'amp\_guess' in the input box for 'amp' (again, the  $S_0^2$ ) and define some other parameters following:

```
guess          amp_guess = 0.85
set            amp_target = 0.85
set            scale = 100
restrain      amp = abs(amp_guess - amp_target) * scale
```

Here we introduce a 'target' parameter  $amp\_target$  together with a 'restrain' parameter  $amp$  and a 'set' parameter  $scale$ . The 'restrain' parameter  $amp$  adds another component to the final 'sum' of chi-square to be minimized to find the best-fit. Then it equivalently to artificially restrict the 'guess' parameter  $amp\_guess$  should change around the pre-set  $amp\_target$ . If the value of  $amp\_guess$  during the fitting tends to go away from  $amp\_target$ , the program will give a 'punishment' for the current value, thus continue to look for best-fit until the total chi-square is minimized. Therefore in this case of restraint fitting, our result for  $amp\_guess$  should be around the pre-set  $amp\_target$  value.

### The Determination of Independent Fitting number using Nyquist-Shannon Theorem

As is shown above when we talk about  $R_{bkg}$  parameter, the number of  $k_{nots}$  does depend on  $R_{bkg}$  value following Nyquist-Shannon sampling theorem. Furthermore, we also have the most number of parameters that we can fit in Artemis is also determined based on Nyquist-Shannon theorem. And here we have two important values — R-range and k-range for us to put into a formula similar with (0-5). First of all, we give the formula as following:

$$N_{param.max} = \frac{2\Delta k \Delta R}{\pi} \quad (0-9)$$





### The correlation between Debye-Waller factor and $s_0^2$

Considering the traditional XAFS expression for single scattering problem:

$$\chi(k) = \sum_j \frac{N_j e^{-2k^2 \delta_j^2} e^{-2R_j/\lambda(k)} f_j(k)}{k R_j^2} \sin[2k R_j + \delta_j(k)] \quad (0-10)$$

It is straightforward to understand the correlation between Debye-Waller factor and amplitude factor  $S_0^2$ , since both of them effectively reduce the amplitude of the absorption signal! Thus during the fitting, the change of any of them will definitely influence the other, which is just what the 'correlation' refers to, in this case. Thus when we fix one of them then release the other for fitting, the fitting error of the other parameter will definitely decrease dramatically.

### About $\sigma^2$ in Artemis fitting

Although the  $\sigma^2$  obtained from Artemis fitting is directly related to the vibration of atoms in the cluster, it is actually different from the Debye-Waller factor from X-ray diffraction. For EXAFS,  $\sigma^2$  is the variance in the absorber-scatterer (scattering path) distance, and the actual Debye-Waller factor from X-ray diffraction is the variance in the position of an atom relative to its lattice point in the crystal. If the motion of absorber and scatter atoms is independent from each other, i.e. uncorrelated vibration, then the  $\sigma^2$  in EXAFS can be thought of twice as big as the Debye-Waller factor in X-ray diffraction, roughly. However, for correlated vibration of atoms in crystal, e.g. the correlation between the absorber and the first-shell neighbours, the EXAFS  $\sigma^2$  is much smaller than Debye-Waller factor in X-ray diffraction. Specially, if the vibration of absorber and 1st shell neighbours is completely correlated, then even the Debye-Waller factor is non-zero for each atom, the EXAFS  $\sigma^2$  is zero since there is no relative motion between absorber and scatterer, thus no vibration in the scattering path length.

Detailed discussion about this topic can be found in the following link: [Click Me](#) or [Backup Link](#).

### About background removal in Athena

Actually, the program Athena for EXAFS signal background removal and Artemis for fitting the background removed EXAFS data are both part of the IFEFFIT program. However IFEFFIT can only do the fitting based on the path information file from FEFF calculation, the Artemis software actually embeds the FEFF6 version to do the multiple-scattering calculation. To do the EXAFS fitting, it is necessary to make some preparations, the first one of which is the normalization to exclude the influence of sample thickness, measuring facility difference, etc. And the second step is to remove the background absorption signal, which means the absorption signal corresponding to a bare absorption atom without any surrounding atoms. The 'Spline' algorithm is just what is used in Athena (actually, in IFEFFIT, and the algorithm used in IFEFFIT is called AUTOBK). Simply, the 'Spline' is to divide the whole range of data





(which will be specified before spline process) to many pieces and build up the background fitting polynomial within each separate piece of data. Then all of the polynomial within each single piece of data range will be connected smoothly. Although the exact way of doing spline background fitting in IFEFFIT is unknown, it should be something like this: first pick up the representation points in the data (either randomly or following some rules), then use any given polynomial to do interpolation between each selected two points. The next step is of course to connect all neighbouring polynomial. However, in IFEFFIT, this is not over since we still don't know what the background is, actually. In the Athena documentation, it was given the reason why we need to do the background subtraction, which can be found following the link (P45): [Click Me](#). Basically, it is to remove the influence of the background signal on the real EXAFS oscillation signal. Thus IFEFFIT (actually, the AUTOBK algorithm) provides a parameter named 'Rbkg', which defines the value below which there is no real EXAFS oscillation signal (thus, the background). So what AUTOBK algorithm does is to vary the piece-wise spline to fit the EXAFS signal (after Fourier transform to R-space) at the low frequency range (specified by 'rbkg', below which is what we mean by 'low frequency').

### About negative absorption coefficient

In some cases, the detector setup were such that the spectrum crosses zero-axis, which means we may then have negative absorption coefficient. One particular example is gold. However, this does not influence the common analysis for the EXAFS data. If imaging the obtained absorption coefficient based on the transmission mode is actually right after the logarithmic operation. Therefore the value of the absorption coefficient itself (no matter positive or negative) still reflects the variation of the absorption capability of the material to specific X-ray energy. It should be mentioned the way to deal with the normalization of the absorption spectra for the above mentioned case where we have negative absorption coefficient does require some care. The most important care is that actually the way to extract the experimental  $\chi(k)$  spectra should be adjusted from its original form:

$$\chi(k) = \frac{\mu(E) - \mu_0(E)}{\mu_0(E)} \quad (0-11)$$

to the form that is generally used:

$$\chi(k) = \frac{\mu(E) - \mu_0(E)}{\Delta\mu(E)} \quad (0-12)$$

where the  $\Delta\mu(E)$  in the denominator is the edge step. (0-12) avoids the problem caused by the negative absorption coefficient since it is the edge step instead of the real absorption coefficient at the absorption edge that is used for the extraction of  $\chi(k)$ . About the detailed discussion mentioned above, refer to the following link: [Click Me](#) or [Click Me](#).



### Importing chi-k data in Athena

To import chi-k data in Athena, the data type ' $\chi(k)$ ' can be selected in the dropdown menu. Although it is the 'chi-k' data that is going to be imported, Athena does not adjust the data column name accordingly and still the 'Energy', 'Numerator', 'Denominator' is given as the option. If the 'chi-k' data to be imported contain the energy column, then it is straightforward to select the data column when importing. After importing the data, the plot will be given in ' $\chi(k)$ -Energy' form, which means the x-axis will be given in energy but not ' $k$ '! Since the ' $\chi(k)$ ' data type was selected at the very beginning, the absorption edge box, together with all the background subtraction boxes, is greyed out. And since no absorption edge value is provided in this case, the real k-space spectra actually cannot be obtained! Also the following Fourier transform cannot be done as well. To overcome this problem, a possible solution is to select the data type as ' $\mu(E)$ ' at the very beginning step when importing the data. Although the data is not really in energy space, Athena can still give a reasonable background, which is basically just a flat line, to the 'fake' energy space spectra. Then everything else is similar to the normal case where the real energy space data is provided.

