Oversampling and conversion to $k$-space

Grant Bunker*

Department of Physics, Illinois Institute of Technology, Chicago, IL 60616, USA. *Correspondence e-mail: bunker@iit.edu

Methods of converting X-ray absorption fine-structure (XAFS) data to $k$-space are described.

1. Introduction

Basic overviews of X-ray absorption fine-structure (XAFS) data acquisition can be found in Stern & Heald (1983), Koningsberger & Prins (1988) and Bunker (2010). The visualization and processing of extended X-ray absorption fine-structure (EXAFS) data are aided by conversion of the absorption data from energy space, in which the data are represented as a function of photon energy $E$, to $k$-space, in which the data are represented as a function of $k$, where the wavevector magnitude $k = p/h = [2m(E - E_0)]^{1/2}/h \simeq [0.2625(E - E_0)]^{1/2}$ in eV Å units, $p$ is the momentum and $m$ is the electron mass. Representing data in this manner naturally displays the sine-like nature of EXAFS oscillations in $k$-space, and it is a necessary step for the subsequent use of Fourier transformation and filtering. Conversely, in $E$-space the data oscillate more slowly as the energy increases above the edge, which partially obscures the information content of the EXAFS.

Data are usually acquired as a function of energy. The near-edge data must be acquired on an energy grid that is sufficient to resolve X-ray absorption near-edge structure (XANES) and pre-edge features, so the sampling interval should be several times smaller than the combined core-hole broadening added in quadrature with the monochromator resolution. The following discussion refers to the extended (EXAFS) region, where it is essential to sample (measure) the EXAFS data on a sufficiently fine energy grid that the highest frequency oscillations in the data are accurately represented. Although some data-acquisition software is set up to acquire data points on a uniform grid in $k$, which decreases the density (in energy space) of sampling points as $E$ increases, other software divides the data into energy regions with different sampling densities. Still other software, for example continuous-scan XAFS (QXAFS; Frahm, 1988), may acquire data on an approximately equal density grid in $E$-space. These data must be converted into $k$-space, maintaining a sufficient sampling density in $k$ to accurately represent the signal. Appropriate averaging by rebinning helps to preserve the statistical weight of data points, as described and illustrated below. Rebinning is supported by popular EXAFS data-analysis programs (see, for example, Ravel & Newville, 2005; Newville, 2013) or it can readily be carried out by other means, as shown below.

2. Nyquist frequency

As discussed more fully elsewhere in this volume, the Nyquist sampling theorem indicates that in order to accurately
represent a band-limited signal (i.e. one whose signal content strictly lies within a definite frequency band), the highest frequency component that is present in that signal must be sampled at least twice per oscillation of that component. If it is not sampled frequently enough, the high-frequency components will alias to (appear as) lower frequency components, which misrepresent the information content of the signal. The Nyquist frequency is $\pi/(2R_{\text{grid}})$ for an EXAFS signal corresponding to a path of effective distance $R_{\text{eff}}$. For $R_{\text{eff}} = 4 \text{ Å}$ this would imply a theoretical minimum spacing between data points of $\delta k \simeq 0.4 \text{ Å}^{-1}$. In practice it is desirable to oversample the data by at least a factor of 8, giving $\delta k \simeq 0.05 \text{ Å}^{-1}$ or less. This more accurately represents the signal waveform and reduces the aliasing of high-frequency noise, which would otherwise masquerade as the desired lower frequency signal.

3. Interpolation, resampling (rebinning) and noise

It is straightforward to convert the independent variable from $E$ to $k$ simply by the mathematical transformation given above, but one still has to change from the initial energy grid to the desired $k$-grid. This involves interpolation of the data set, which can be accomplished by various means, such as cubic spline, Lagrange or Hermite interpolation. Interpolation uses data samples at values of $k$ that are not generally on the desired grid to estimate what the signal would be at points on the desired $k$-grid. This is an approximate process, and it is important to use a sufficiently fine $k$-grid spacing (i.e. 0.05 Å$^{-1}$ or less) on which to interpolate in order to adequately sample the peaks of the oscillatory waveforms, otherwise the envelope of the waveform (the amplitude of the oscillations) will be underestimated because the probability of a grid point landing on the maximum or minimum values of the waveform is smaller with fewer grid points. The interpolation does not ‘know’ the original waveform so the amplitude will normally be underestimated.

In data analysis the spectra are normally weighted with a power of $k$, typically $k^2$ or $k^3$, to make the EXAFS spectrum more constant in amplitude in order to better resolve Fourier transform peaks and to provide more consistent weighting of the data in background-subtraction and fitting procedures. The data are, however, subject to noise from various sources, such as counting statistics and electronics. When the data are weighted by $k^2$ or $k^3$, this noise is greatly enhanced at high $k$. To mitigate this, it is helpful to acquire the high-$k$ data with a longer integration time or with a higher density of data points, which are then signal-averaged.

Although interpolation provides some averaging effect, when converting data from $E$ to $k$ it is desirable to preserve the statistical weight of the data samples. For example, if the data are acquired (as in QXAFS) with an approximately uniform density (say 1 eV or less) in $E$-space, this results in a density in $k$-space that grows linearly with $k$, so that several (or many) measured points may correspond to a single point on the desired $k$-space grid, and these should be averaged to reduce noise. A simple and effective procedure is then to group together the points that map into each $k$-space grid interval (for example from $k_n - \delta k/2$ to $k_n + \delta k/2$) and then average their $k$-values and corresponding signal values to produce a single averaged point with a $k$-value that is close to the desired $k$-grid; these can then be interpolated precisely onto the desired $k$-grid with only small errors.

As a concrete example, we consider synthetic QXAFS data sampled uniformly in energy space on a grid of 0.5 eV from 15 to 1000 eV above the edge. The synthetic data are modelled as $\sin(2k R_{\text{grid}}) \exp(-2k^2 \sigma^2)/k^2$, with $R_{\text{grid}} = 4.0$ Å and $\sigma^2 = 0.005$ Å$^2$, and a normally distributed random variate of zero mean and standard deviation $5 \times 10^{-4}$ is added to simulate noise; the result is then weighted with $k^3$. This models the effect of increasing noise level at high $k$. These data are then processed in two different ways: firstly, by rebinning, as described above, into bins of width $\delta k = 0.05$ Å$^{-1}$ that are centred on the desired $k$-grid points; in Mathematica/Wolfram language, for example, this can be concisely performed with the code Mean/@GatherBy[data, l + Floor[(First[#] - kmin + dk/2)/dk] &], where the variable data is an

![Figure 1](image1.png)

**Figure 1**
Synthetic QXAFS data with noise, rebinne b before interpolation to the desired $k$-grid (data points are shown; the solid curve represents the exact data with zero noise).

![Figure 2](image2.png)

**Figure 2**
Synthetic QXAFS data with noise, interpolated to the desired $k$-grid without rebinning (data points are shown; the solid curve represents the exact data with zero noise).
array of \([k, k^3 x]\) pairs, as mapped from the original energy grid, and \(k_{\text{min}}\) is the minimum value of \(k\) on the grid; the result is then interpolated onto the desired grid (Fig. 1). The second (simplistic) way omits the rebinning step: it just takes the data mapped from energy space to \(k\)-space and then interpolates the data to the desired \(k\)-grid (Fig. 2). The latter procedure gives results that are evidently much noisier than the data set produced by rebinning and averaging before interpolation; in this case, the standard deviation of the residuals (data minus exact data) of the rebinned and averaged data is less than 1/3 of the standard deviation of the data that were simply interpolated without rebinning (Fig. 3).

In summary, resampling from irregularly sampled data to the desired \(k\)-space grid can be performed reliably provided that the original grid is oversampled at several times the Nyquist limit, and rebinning/averaging is performed to preserve the statistical weight of the acquired data points.

References