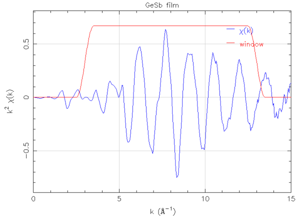
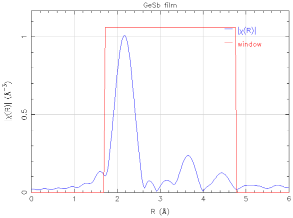
Information Content of EXAFS (I)

Sometimes, we have *beautiful* data. This is the merge of 5 scans on a 50 nm film of GeSb on silica, at the Ge edge and measured in fluorescence at NSLS X23a2.





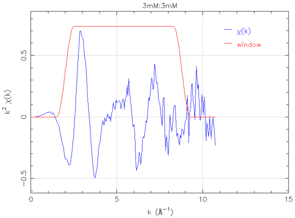
Here, I show a Fourier tranform window of [3 : 13] and I suggest a fitting range of [1.7 : 4.7]. Applying the Nyquist criterion:

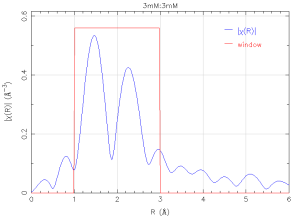
$$N_{idp} \approx \frac{2\Delta k \Delta R}{\pi} \approx 19$$

This gives us an upper bound of the information content of that portion of the EXAFS spectrum.

Information Content of EXAFS (II)

Sometimes, we have less-than-beautiful data. This is the merge of 42 scans on a solution containing 3 mM of Hg bound to a synthetic DNA complex, measured in fluorescence at APS 20BM.





Here, I show a Fourier tranform window of [2 : 8.8] and I suggest a fitting range of [1 : 3]. Applying the Nyquist criterion:

$$N_{idp} \approx \frac{2\Delta k \Delta R}{\pi} \approx 8$$

This talk discusses strategies for dealing with severely limited information content.

What is This Nyquist Criterion?

Given that we apply Fourier analysis to $\chi(k)$, we can treat EXAFS as a signal processing problem. If

- The signal is ideally packed **and**
- The error in the fitting parameters is normally distributed **and**
- We understand and can enumerate all sources of error **and**
- We know the theoretical lineshape of our data **then**

$$N_{idp} \approx \frac{2\Delta k \Delta R}{\pi}$$

where, for EXAFS, Δk is the range of Fourier transform and ΔR is the range in R over which the fit is evaluated.

Unfortunately ...

None of those conditions really get met in EXAFS. N_{idp} is, at best, an upper bond of the actual information content of the EXAFS signal.

Statistical Parameters: Definitions

IFEFFIT uses a Levenberg-Marquardt non-linear least-squares minimization, a standard χ^2 fitting metric, and a simple definition of an R-factor:

$$\chi^2 = \frac{N_{idp}}{\epsilon N_{data}} \sum_{i=min}^{max} \left[\text{Re}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 + \text{Im}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 \right] \quad (1)$$

$$\chi^2_\nu = \frac{\chi^2}{\nu} \quad (2)$$

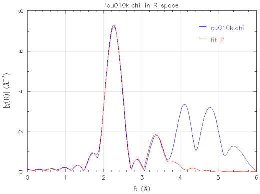
$$\nu = N_{idp} - N_{var} \quad (3)$$

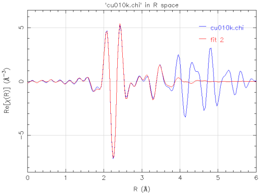
$$\epsilon = \text{measurement uncertainty}$$

$$\mathcal{R} = \frac{\sum_{i=min}^{max} \left[\text{Re}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 + \text{Im}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 \right]}{\sum_{i=min}^{max} \left[\text{Re}\left(\chi_d(r_i)\right)^2 + \text{Im}\left(\chi_d(r_i)\right)^2 \right]} \quad (4)$$

An Obviously Good Fit

Here is a fit to the first two shells of copper metal at 10 K





This is an unambigously good fit:

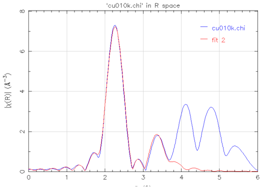
$\mathcal{R}$	0.0025
N_{idp}	16
ν	12
S_0^2	0.95(3)
E_0	5.98(36) eV
a	3.6072(26) Å
Θ_D	505(16) K

Yet $\chi^2 = 32.03$!

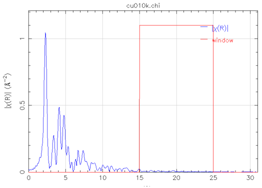
What's goin' on here?

Why is χ^2 for an obviously good fit so much larger than 1?

Statistical Parameters: Fit Evaluation



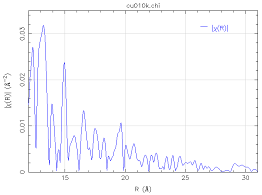
The determination of measurement uncertainty is, perhaps, a bit hokey in IFEFFIT. It is the average of the signal between 15 Å and 25 Å in the Fourier transform – a range that probably does not include much signal above the noise.



Is that signal between 15 Å and 25 Å in copper metal? Perhaps....

In any case, this method ignores the following:

- Approximations and errors in theory
- Sample inhomogeneity
- Detector nonlinearity
- Gremlins ;-)



Statistical Parameters: Interpretation I

OK then ... what is the implication of ϵ never being evaluated correctly by IFEFFIT?

1. χ^2_ν is always somewhere between big and enormous.
2. χ^2_ν is impossible to interpret for a *single* fit.
3. χ^2_ν **can** be used to compare different fits. A fit is improved if χ^2_ν is significantly smaller.
4. Error bars are taken from the diagonal of the covarience matrix. If χ^2_ν is way too big, the error bars will be way too small. The error bars reported by IFEFFIT have been scaled by $\sqrt{\chi^2_\nu}$.
5. Thus the error bars reported by IFEFFIT are of the "correct" size if we assume that the fit is a "good fit".

Statistical Parameters: Interpretation II

How do we know if a fit is "good"?

- The current fit is an improvement over the previous fit if χ^2_ν is sufficiently smaller.
- You should be suspicious of a fit for which N_{var} is close to N_{idp}, i.e. a fit for which ν is small.
- All variable parameters should have values that are physically defensible and error bars that make sense.
- The results should be consistent with other things you know about the sample.
- The R-factor should be small and the fit should closely overplot the data. (That was redundant. ☺)

Interpreting Error Bars

The interpretation of an error bar depends on the meaning of the parameter.

- A fitted σ^2 value of, say, 0.00567 ± 0.00654 is troubling. That result means suggests that σ^2 is quite ill-determined for that path and not even positive definite. Yikes!
- On the other hand, a fitted E_0 value of, say, 0.12 ± 0.34 is just fine. E_0 can be positive or negative. A fitted value consistent with 0 suggests you chose E_0 wisely back in ATHENA.

Outside Knowledge

Because the information content of the XAS measurement is so limited, we are forced to incoporate knowledge from other measurements into our data analysis and its interpretation.

- Other XAS measurements — for instance, the “chemical transferability” of S_0^2
- Diffraction tells us structure, coordination number, bond lengths, etc.
- Things like NMR, UV/Vis, and IR can tell us about the ligation environment of the absorber
- Common sense:
  - $R_{NN} \ngeq 0.5\,\text{Å}$, $R_{NN} \ngeq 4.0\,\text{Å}$
  - $\sigma^2 \nless 0\,\text{Å}^2$
- ... and anything else your (physical ‖ chemical ‖ biological ‖ whatever) intuition tells you