SAND2014-4547C

Global Analysis Peak Fitting for Imaging NEXAFS Data

Mark H. Van Benthem and James A. (Tony) Ohlhausen

Materials Science Research Division

Sandia National Laboratories

SESSION 8 IMAGING AND DATA ANALYSIS

2014.06.05 1700

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Overview

- **NSLS**
  - **Beamline U7A**
- **NEXAFS (*a.k.a.* XANES)**
  - **Near Edge X-ray Absorption Fine Structure**
  - **Spectroscopy**
  - **Imaging**
  - **Data Arrays**
- **Multivariate Analysis**
  - **Peak Fitting**
  - **Least Squares**
  - **Principal Component Analysis (PCA)**

Motivation

- NEXAFS gives information about bonds in various types of materials including organics
- Peak fitting can help elucidate the nature of bonding in polymers
- Typically, peak fitting is performed on single spectra
- Fitting NEXAFS images, multiple spectra, simultaneously can provide information about the areal extent of bonding in the material as well as mixed species

National Synchrotron Light Source (NSLS)

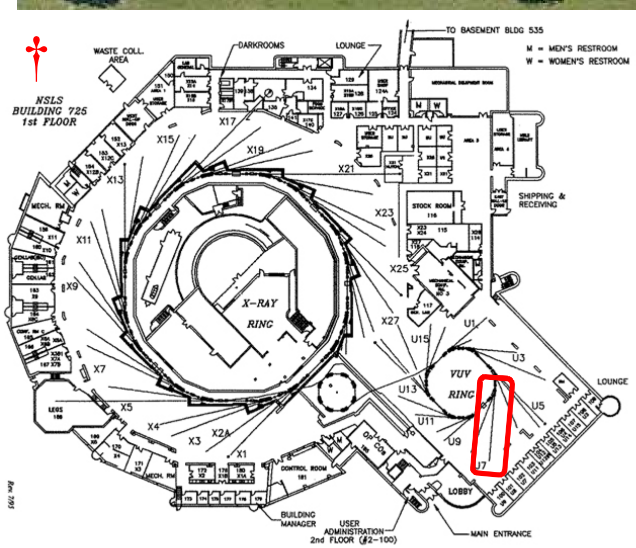
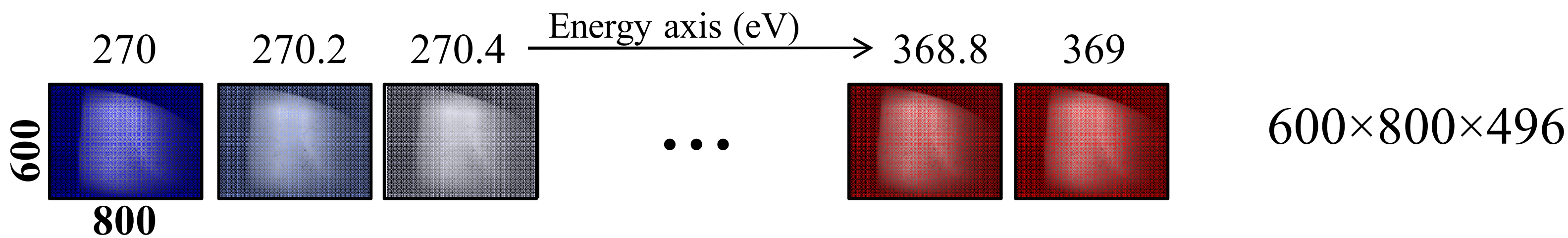


Image Sources: * www.bnl.gov and †cdac.carnegiescience.edu

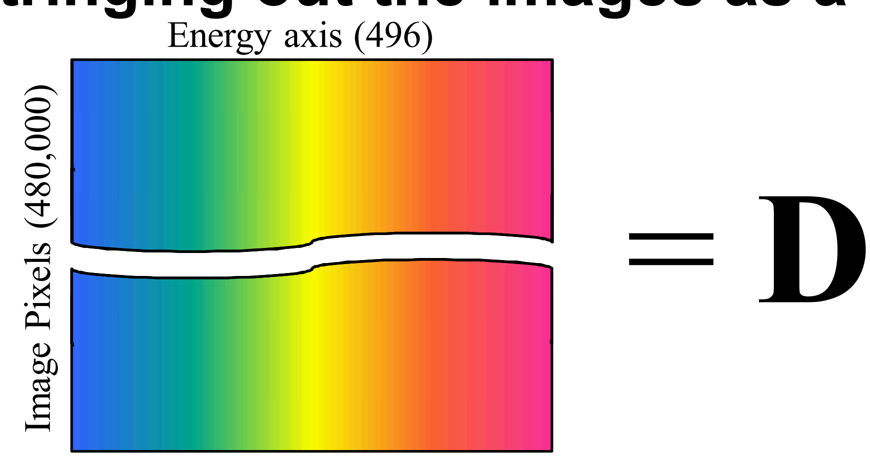
Sandia National Laboratories

NEXAFS Data Arrays

- Consider a collection of Imaging NEXAFS data



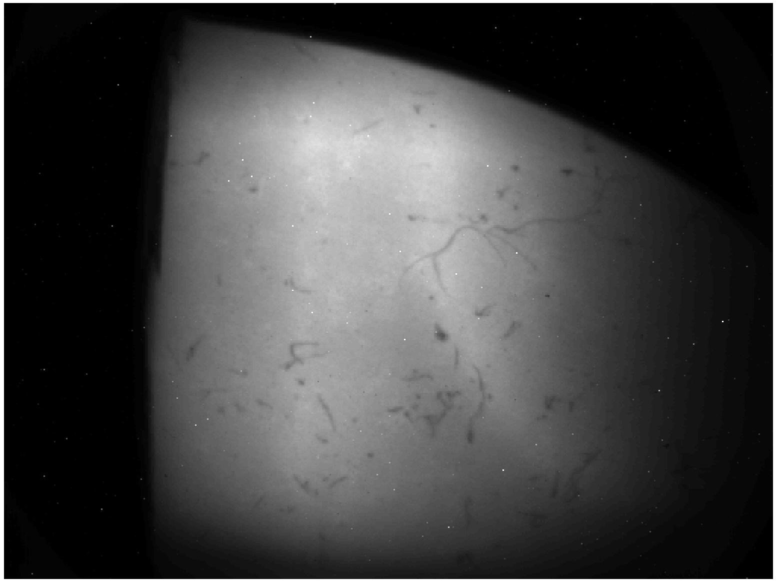
- These data can be reorganized as a matrix by stringing out the images as a vector of pixels



Sandia National Laboratories

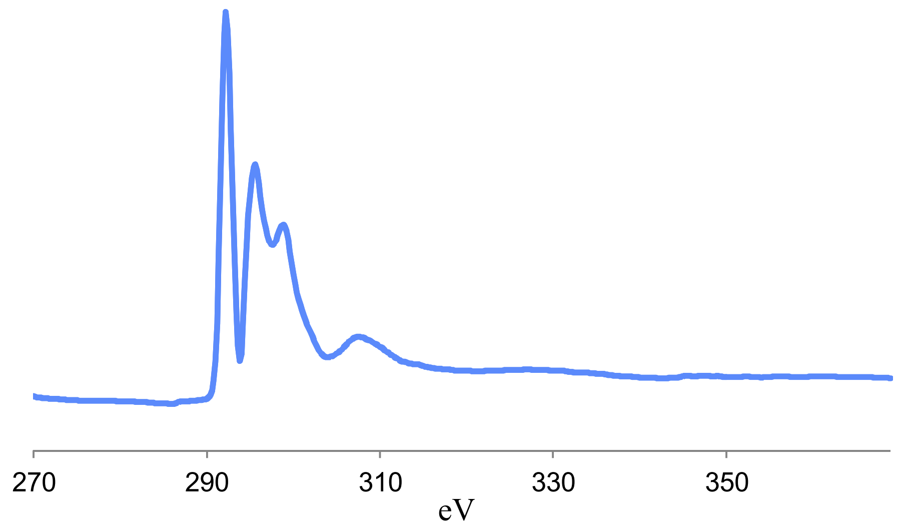
Grayscale Image

- If we collapse **D** along the spectral dimension, we obtain the monochrome image.



Energy Spectrum

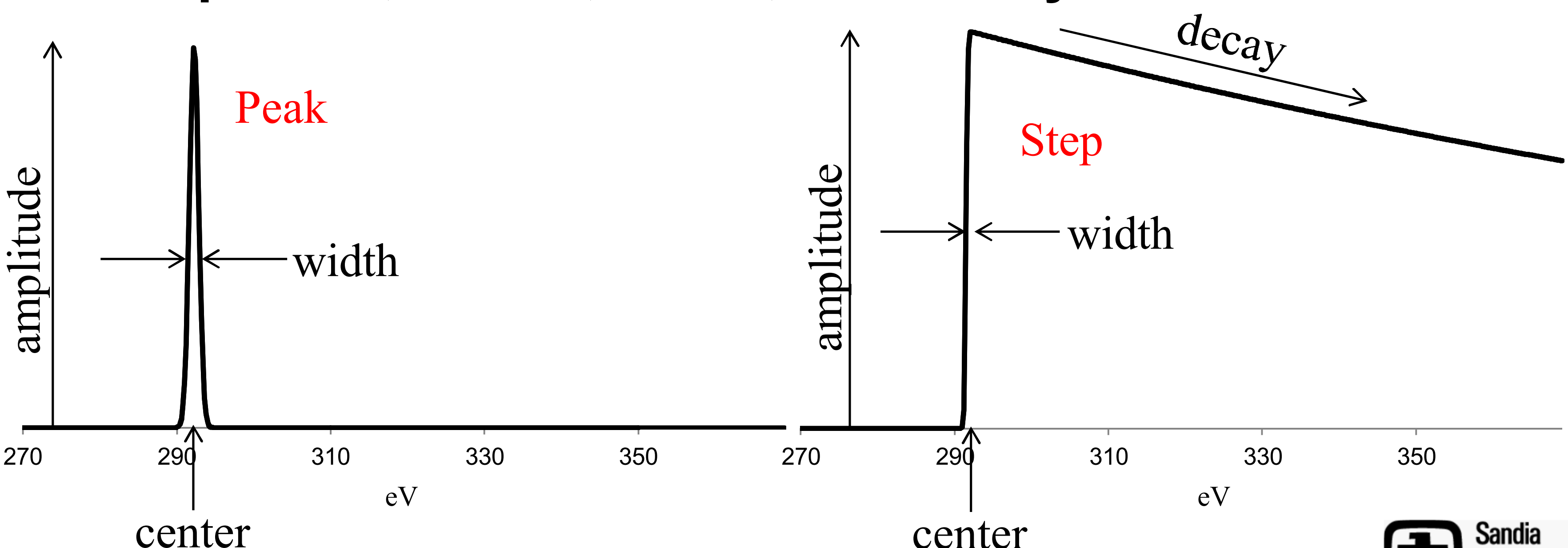
- If we collapse **D** along the image dimension we get the energy spectrum



- The typical user will fit the energy spectrum with Gaussian, Lorentzian, or Voigt curves or asymmetric variants thereof. They will also fit a step function.

Curve Fitting

- **Gaussian and Lorentzian peaks are characterized by three parameters: amplitude, center, and width**
- **The step function has four parameters: amplitude, center, width, and decay rate**



J. Stöhr, *NEXAFS Spectroscopy, Springer, Berlin (1992).*

Peak, Step and Offset Definitions

- **Gaussian:** $I_G = A e^{-\left(\frac{E-E_0}{w} c\right)^2}$; where $c = 2\sqrt{\log 4}$
- **Lorenztian:** $I_L = A \left(\frac{\left(w/_2 \right)^2}{(E-E_0)^2 + \left(w/_2 \right)^2} \right)$
- **Pseudo-Voigt:**

$$I_V = A \left[\eta \left(\frac{\left(w/_2 \right)^2}{(E-E_0)^2 + \left(w/_2 \right)^2} \right) + (1-\eta) e^{-\left(\frac{E-E_0}{w} c\right)^2} \right]$$

- **Asymmetric Peaks: Set** $w = mE + b$
  - **Both** m **and** b **are common to all shifted peaks in sample spectra**
- **Shaped Step:** $I_S = A \left[\frac{1}{2} + \frac{1}{2} \mathrm{erf} \left(\frac{E-E_0}{w} d \right) \right]$; where $d = 2\sqrt{\log 2}$
  - **Can also introduce exponential decay term into step function**
- **Offset:** $I_O = A$
- **Red indicates linear term**

J. Stöhr, *NEXAFS Spectroscopy, Springer, Berlin (1992).*

Set Up the Least Squares Problem

- **The model is $\mathbf{D} = \boldsymbol{AS}^{\mathrm{T}}$**
  - D **is the data matrix, dimensioned as number of image pixels by number of spectral channels**
  - A **is the matrix of linear coefficients, dimensioned as number of pixels by number of peaks, steps and offsets (factors)**
  - S **is the matrix of nonlinear terms, dimensioned as number of spectral channels by number of factors**
  - **Superscript** T **represents matrix transpose**
- **The least squares criterion: minimize $\left\|\mathbf{D} - \boldsymbol{AS}^{\mathrm{T}}\right\|^2$**

Knorr, F. J. and J. M. Harris, Analytical Chemistry **53(2):** 272-276, (1981).
Beechem, J. M., Numerical Computer Methods. L. Brand and M. L. Johnson, Methods in Enzymology San Diego, Academic Press. **210:** 37-54, (1992).

Sandia National Laboratories

Solving the Problem

1. Solve nonlinear terms using a nonlinear solver, like nonlinear least squares
   - Initialize with best guesses for peak or step parameters
   - Each peak or step is computed using the estimated parameters and the given energy axis
   - The offset is entered as a column of ones; it has no nonlinear term
2. Given the estimate of $\hat{S}$ from nonlinear solution, solve the linear terms using least squares
   - $\hat{\mathbf{A}} - \mathbf{D}\hat{\mathbf{S}}\left(\hat{\mathbf{S}}^{\mathrm{T}}\hat{\mathbf{S}}\right)^{-1}$ (can impose nonnegativity)
   - This is done within the nonlinear function call
3. Iterate until convergence

Sandia National Laboratories

Compression

- We can represent the data as the product of two orthogonal matrices using principal component analysis (PCA): $\mathbf{D} = \mathbf{T}\mathbf{P}^{\mathrm{T}}$
  - T is the matrix of orthogonal “scores” dimensioned as #pixels by #principal components (#PCs)
  - P is the matrix of orthonormal “loadings” dimensioned as #spectral channels by #PCs
  - Number of PCs<<min(#pixels,#channels)
- Recall the model is $\mathbf{D} = \mathbf{A}\boldsymbol{S}^{\mathrm{T}}$
  - Now we can write $\mathbf{T}\mathbf{P}^{\mathrm{T}} = \mathbf{A}\boldsymbol{S}^{\mathrm{T}}$
  - Finally, we can define $\mathbf{P}^{\mathrm{T}} = \widetilde{\mathbf{A}}\boldsymbol{S}^{\mathrm{T}}$

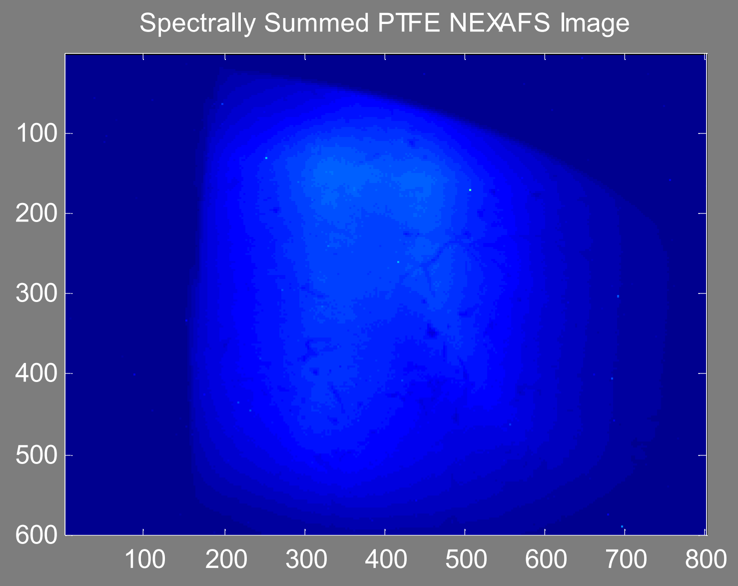
Henry, E. R. and J. Hofrichter, Numerical Computer Methods. L. Brand and M. L. Johnson, Methods in Enzymology San Diego, Academic Press. **210:** 129-192, (1992).

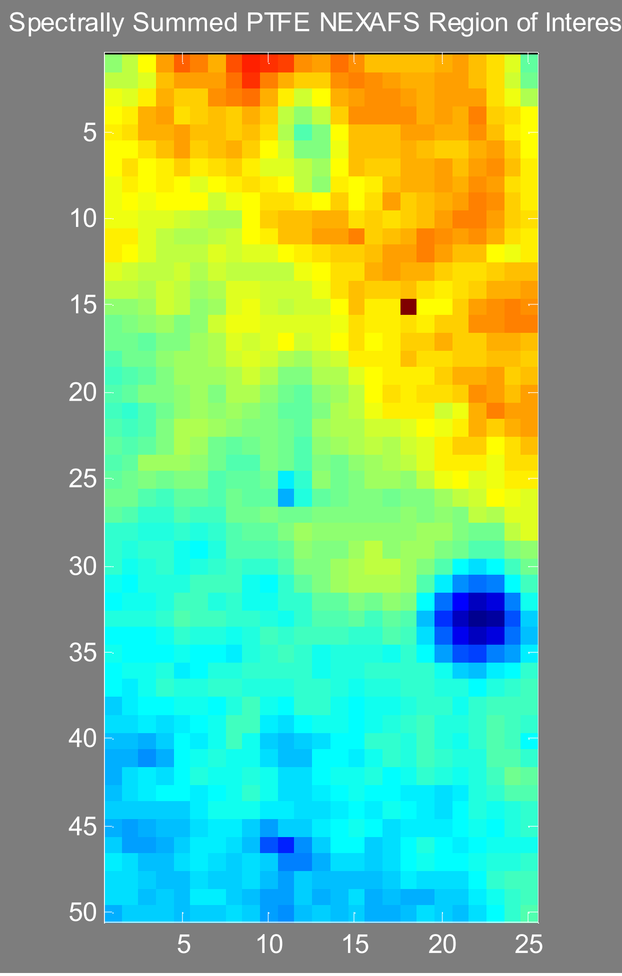
Compression Use

- **One can treat the following equation identically to the full data least squares problem**
  - **Model** $\mathbf{P}^{\mathbf{T}} = \widetilde{\mathbf{A}}\boldsymbol{S}^{\mathbf{T}}$
  - **Minimize** $\left\|\mathbf{P}^{\mathbf{T}} - \widetilde{\mathbf{A}}\boldsymbol{S}^{\mathbf{T}}\right\|^{2}$
  - **Solve nonlinear part to obtain** $\hat{\boldsymbol{S}}$
  - **Solve** $\hat{\hat{\mathbf{A}}} - \mathbf{D}\hat{\mathbf{S}}\left(\hat{\mathbf{S}}^{\mathbf{T}}\hat{\mathbf{S}}\right)^{-1}$
  - **After convergence compute** $\widehat{\mathbf{A}} = \mathbf{T}\hat{\hat{\mathbf{A}}}$
- **Nonnegativity can be imposed with only a minor computational penalty**

Sandia National Laboratories

Image Mode of PTFE* NEXAFS Data

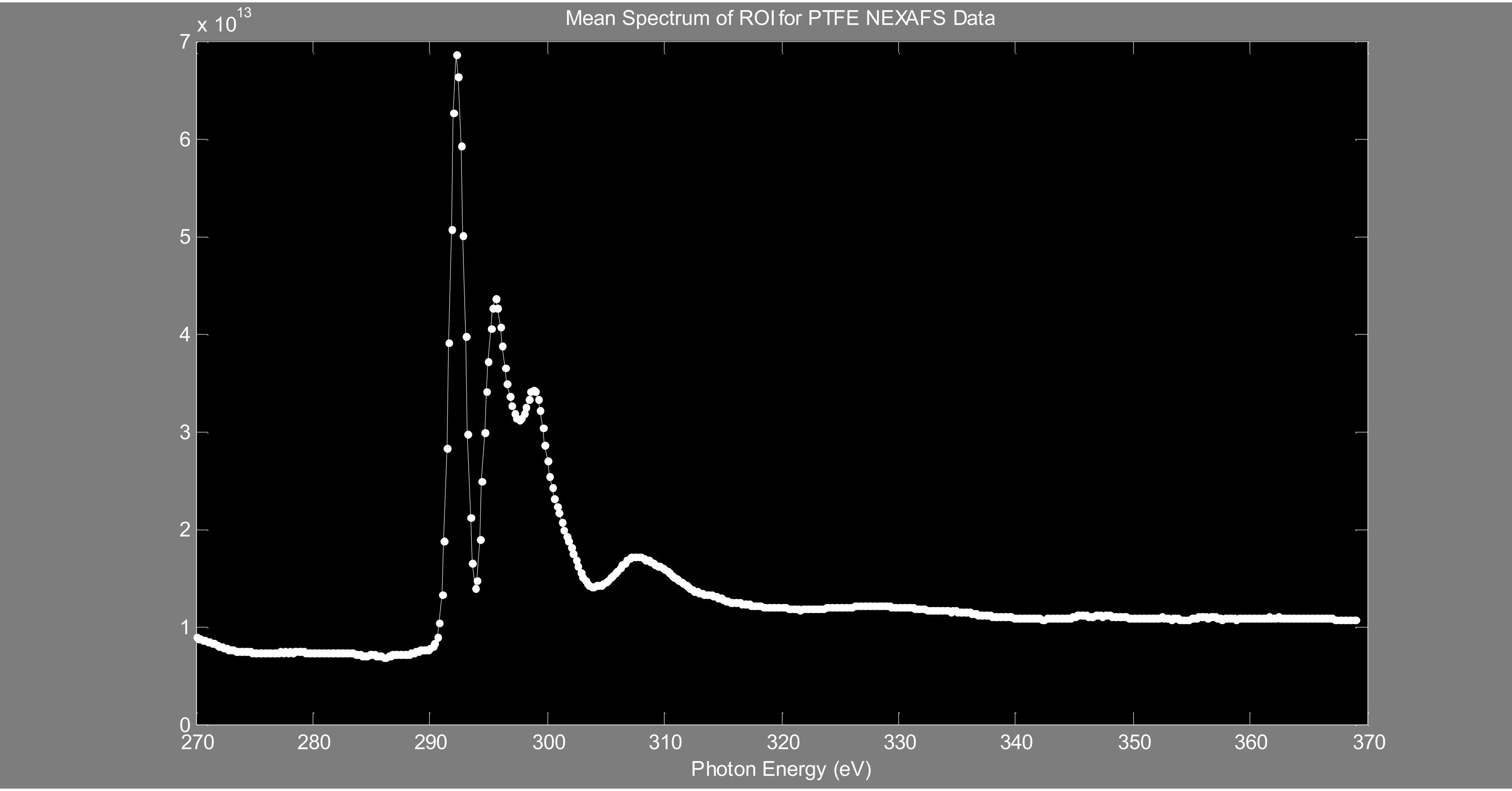




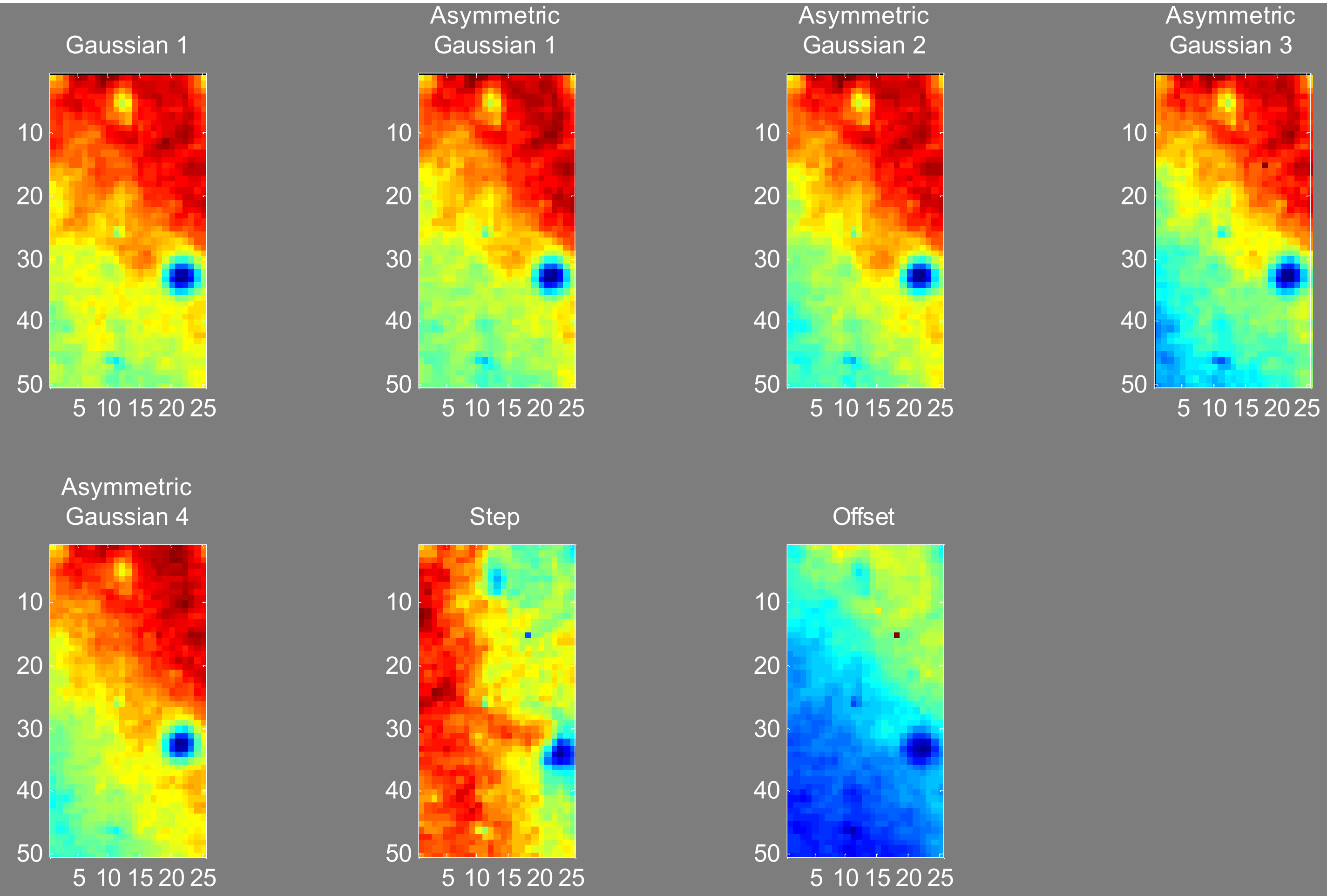
*polytetrafluoroethylene

Calculations performed using MATLAB Version: 8.3.0.532 (R2014a). Uses lsqnonlin.m in Optimization Toolbox.

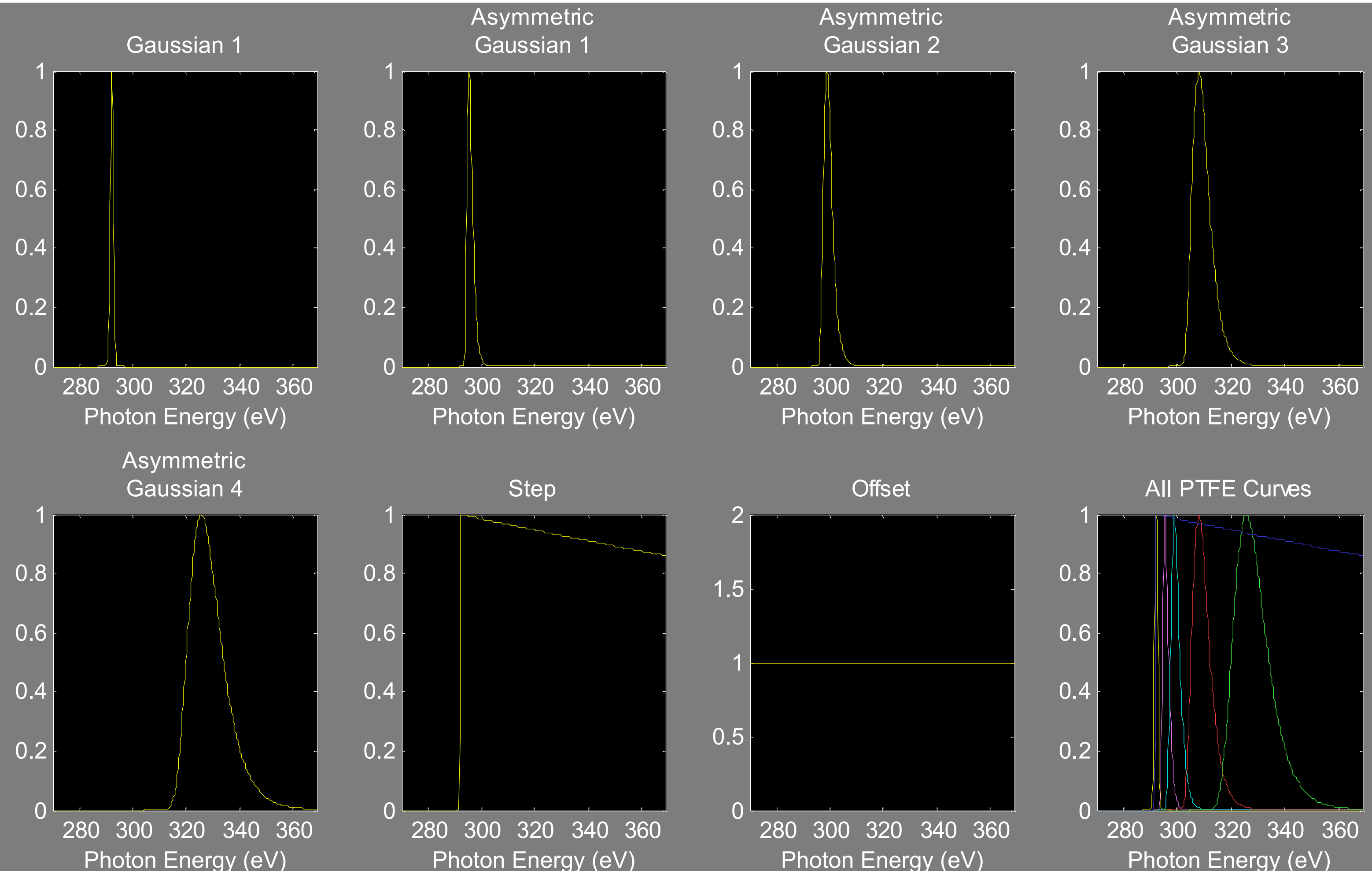
PTFE ROI Mean NEXAFS Spectrum



Estimated Image Mode Factors For PTFE



Estimated Energy Mode Peaks for PTFE



Analysis time ~2.1 seconds.

HP Elitebook 850. Intel Core i7-4600U@2.1GHz. 16GB ram. 64bit Windows 7 OS.

Sandia National Laboratories

Conclusions

- Developed and implemented a fast multivariate method of peak fitting for NEXAFS data
- Capable of fitting many spectra simultaneously
  - Currently used to fit single images
  - Could be applied to multiple images or spectra
- Takes advantage of image inhomogeneities during fitting process

Sandia National Laboratories

Acknowledgements

- Daniel Fischer, National Institute of Standards & Technology
- Cherno Jaye, National Institute of Standards & Technology
- National Synchrotron Light Source (NSLS) and Brookhaven National Laboratory (BNL)

Sandia National Laboratories