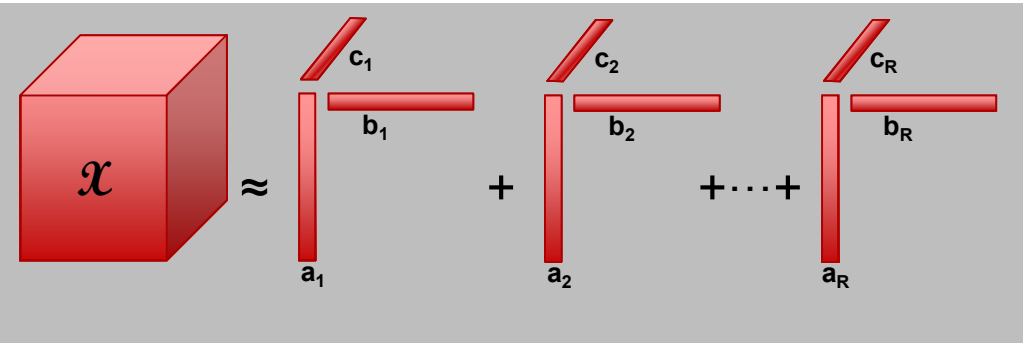
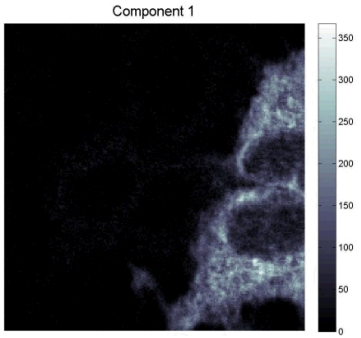
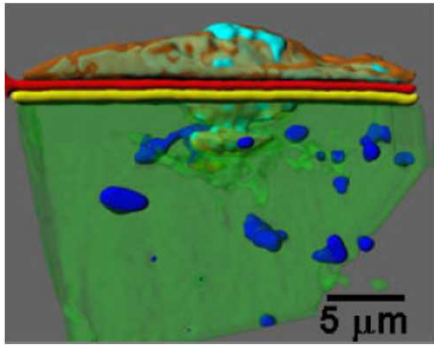
SAND2015-10674PE

Exceptional service in the national interest

Sandia National Laboratories







Multivariate Analysis: A Tool for Turning Data to Information

Mark Van Benthem

Sandia National Laboratories, Albuquerque, NM

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

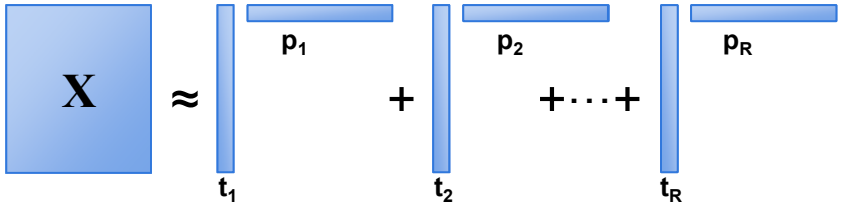
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Motivation

Sandia National Laboratories

- Multivariate analysis can reduce the time to analyze and interpret data
  - Reducing data dimensionality is the key to extracting the greatest amount of information from laboratory data
- Using the right tools in preprocessing and optimized multivariate tools is a boon to the researcher
  - Using appropriate data scaling for the nature of the data variance and the assumptions of the analytical tool
    - Optimal scaling for Poisson-distributed data
  - Fast and mathematically rigorous algorithms
    - Fast-combinatorial nonnegative least squares
    - Rigorous equality constrained least squares

Two-Way Analysis Methods



- Principal Component Analysis (PCA)
- Given a matrix containing data, D, as a first step in many analyses we want principal components

$$\mathbf{D} \cong \mathbf{T}\mathbf{P}^{\mathbf{T}}$$

- Such that T and P are an orthogonal basis sets, that is a reduced dimensional representation of D, with ordered maximized variance.
  - T is orthogonal (scores); P is orthonormal (loadings).
- Optimal Scaling-Scale D using the column and row space means

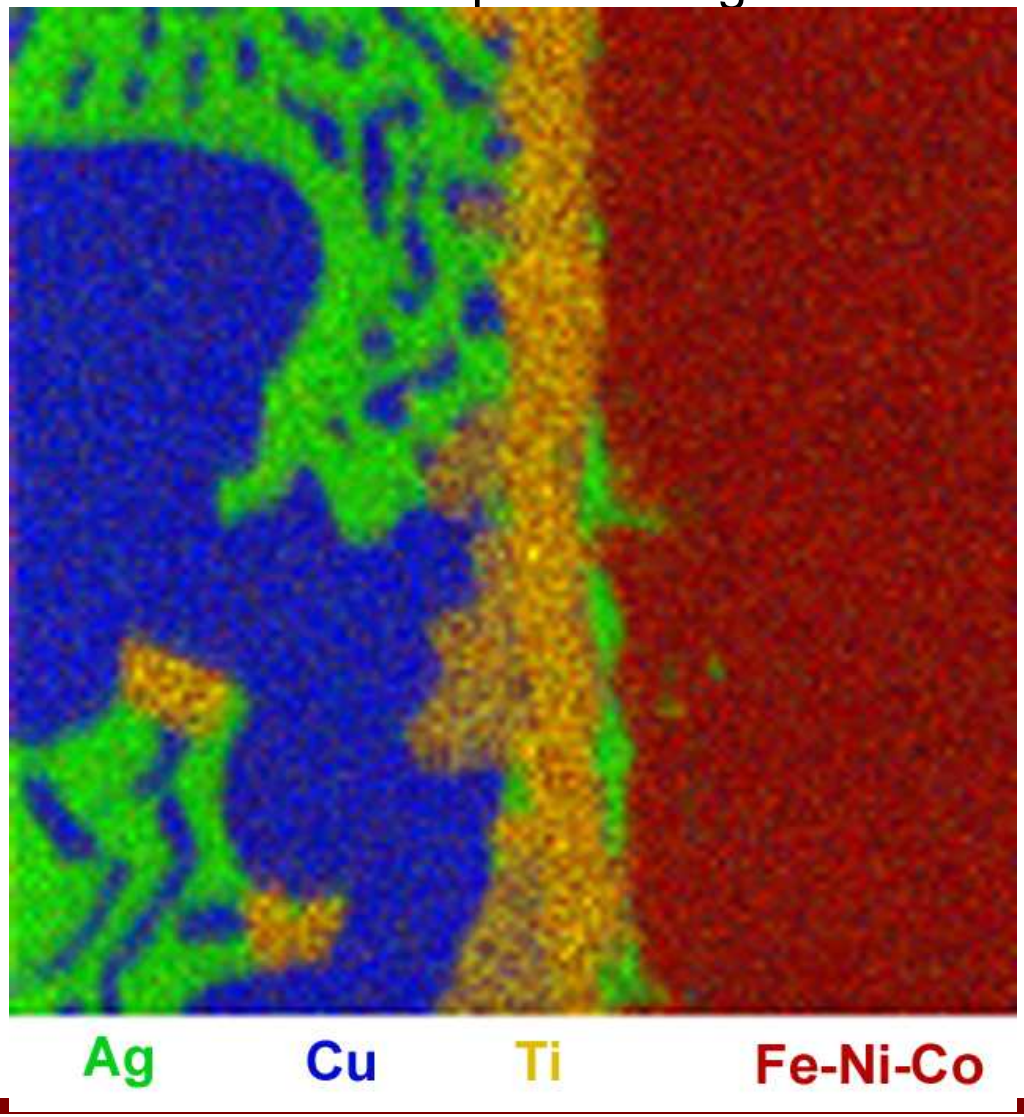
$$\tilde{\mathbf{D}} = \mathbf{GDH}$$

- Multivariate Curve Resolution (MCR)
  - Impose constraints on solution space
  - Alternating least squares employing constrains

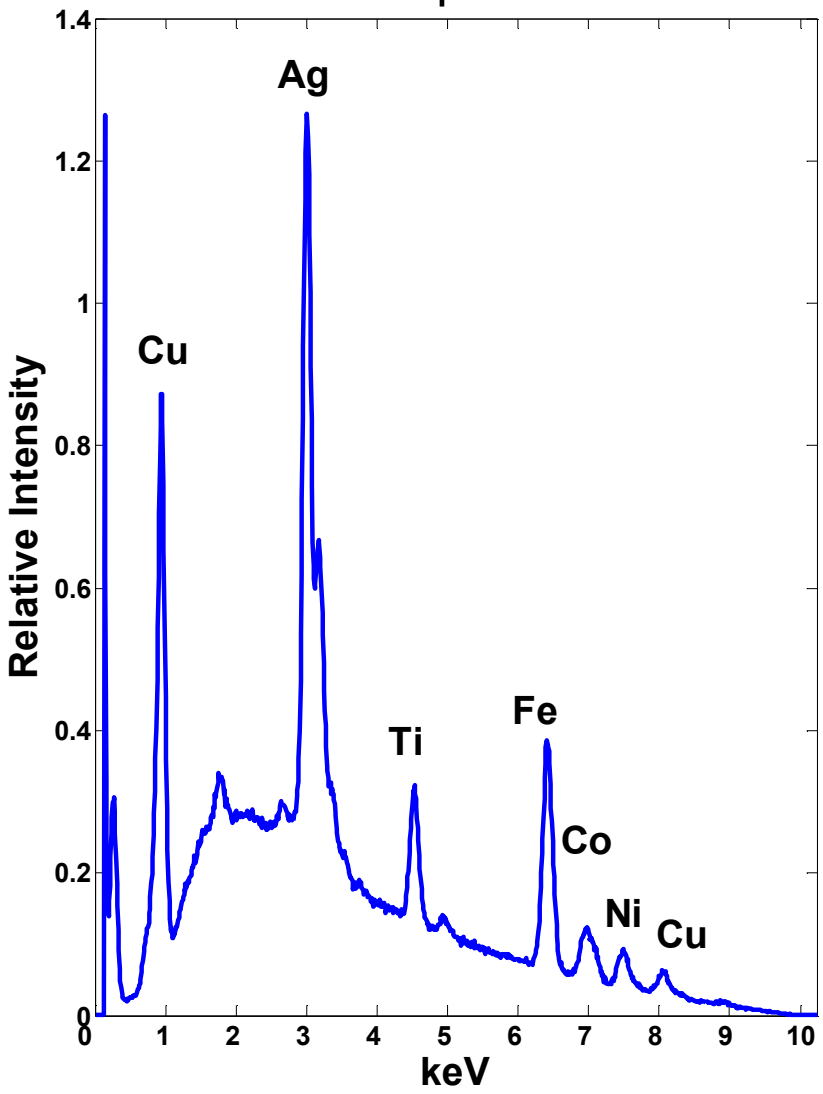
An Example: Energy Dispersive X-ray Analysis of a Braze Interface

Sandia National Laboratories

RGB Composite Image



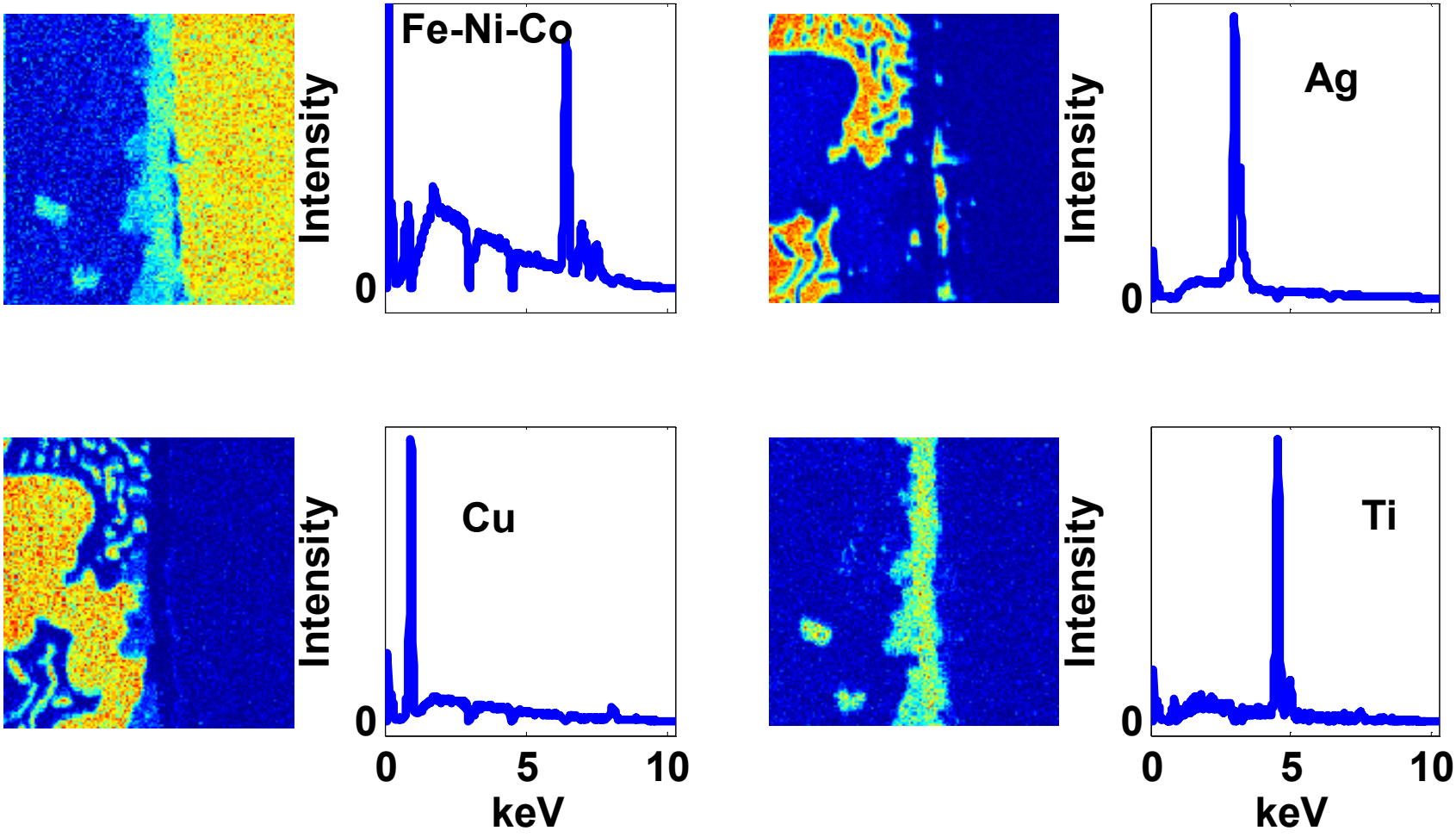
Mean Spectrum



In: Keenan, Van Benthem, Kotula *Multivariate curve resolution in the presence of non-specific spectral backgrounds*, FACSS 2002 12-17 Oct 2002.

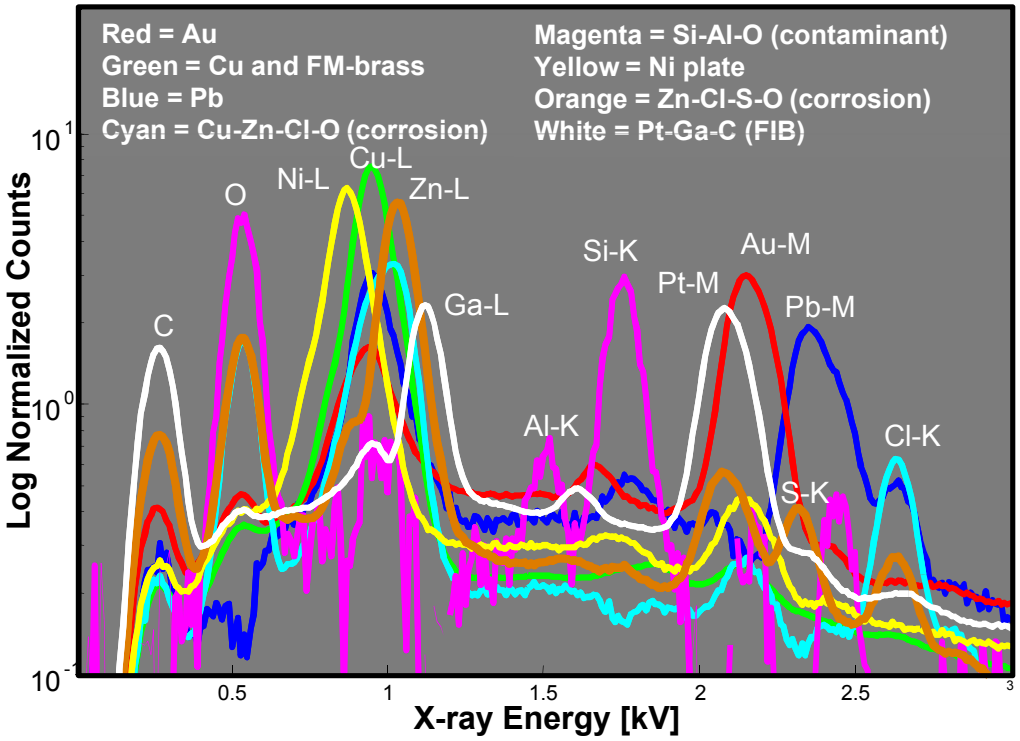
MCR-ALS

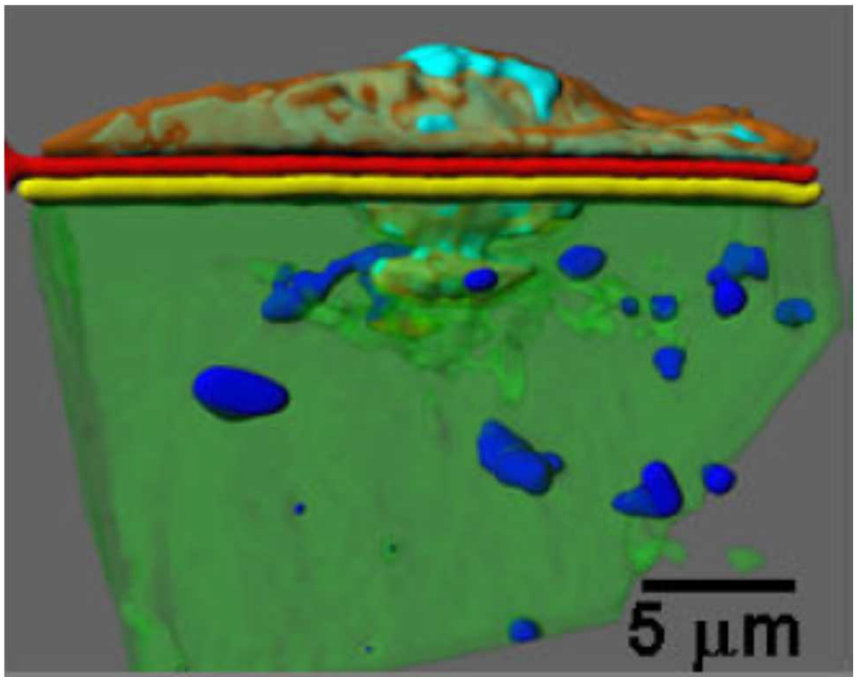
MCR-ALS with non-negativity constraints
Energy Dispersive X-ray Analysis of a Braze Interface



MVA of FIB/SEM/EDXS Tomographic Spectral Imaging (TSI)

Sandia National Laboratories





Spectral shapes from the factor analysis of the entire TSI (left). Note the significant pathological overlap of the X-ray lines which SNL-developed Automated eXpert Spectral Image Analysis (AXSIA) has unmixed. Overlay of the component images corresponding to their respective spectral shapes (right). The Cu-brass component (green) is rendered translucently so that internal porosity can be visualized.

Kotula, P.G., Van Benthem, M.H. & Sorensen, N.R. (2012). Application of FIB/SEM/EDXS Tomographic Spectral Imaging and Multivariate Statistical Analysis to the Analysis of Localized Corrosion. In *IEEE Statistical Signal Processing Workshop, pp. 672-675. Ann Arbor, MI: IEEE.*

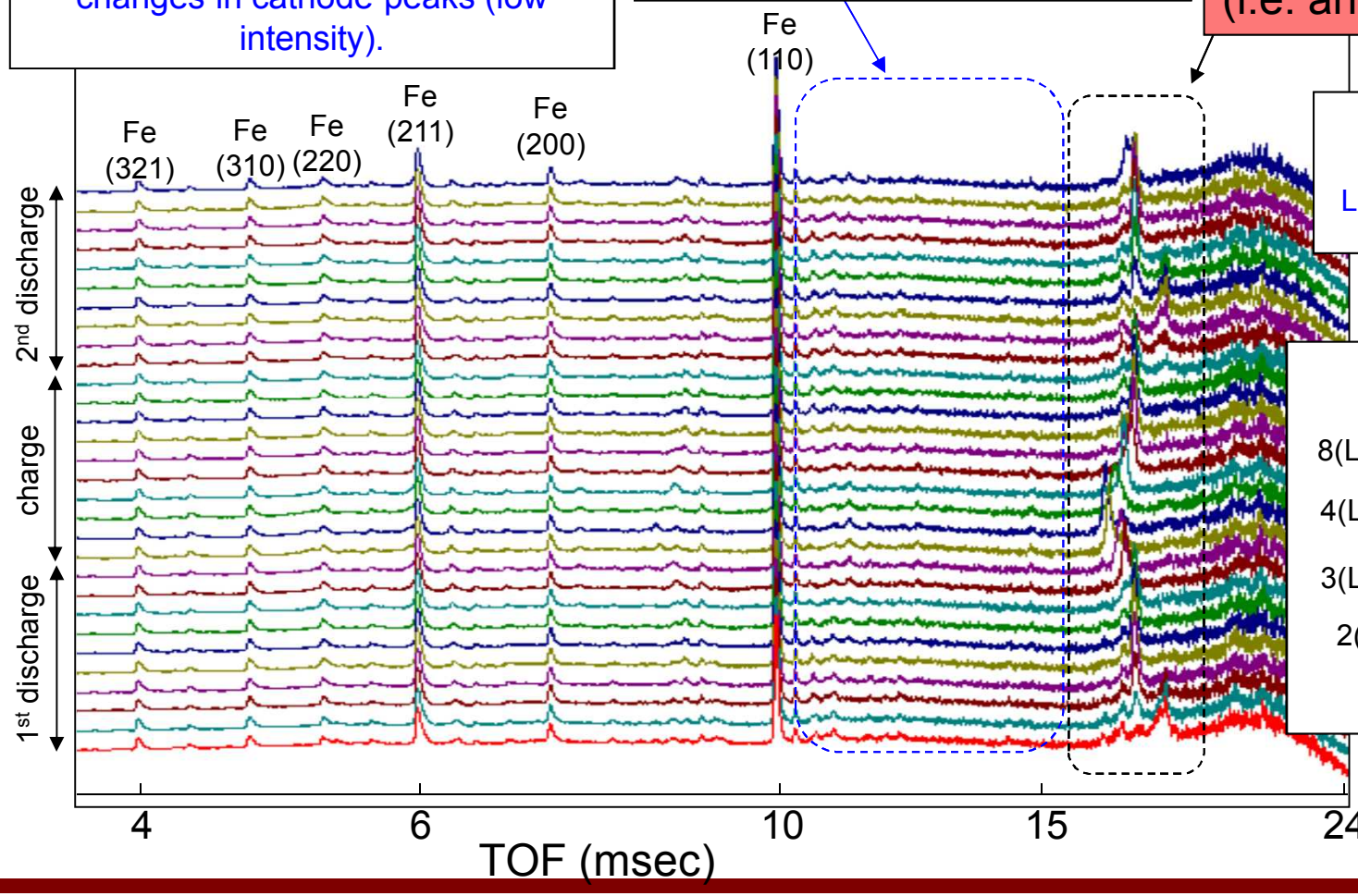
Use neutron diffraction to monitor battery in-situ (during cycling).

Sandia National Laboratories

Changes in anode were immediately observed in neutron diffraction raw data. It was hard to see changes in cathode peaks (low intensity).

Cathode information is buried deep in the noise in the ~10 -16 msec range (i.e. cathode "window")

16-20 msec range is sensitive to anode (i.e. anode "window"



Cathode Rx

$LiFePO_4 \leftrightarrow FePO_4 + (Li^+,e^-)$

Anode Rx

$8(LiC_6) \leftrightarrow 4(LiC_{12}) + 4(Li^+,e^-)$

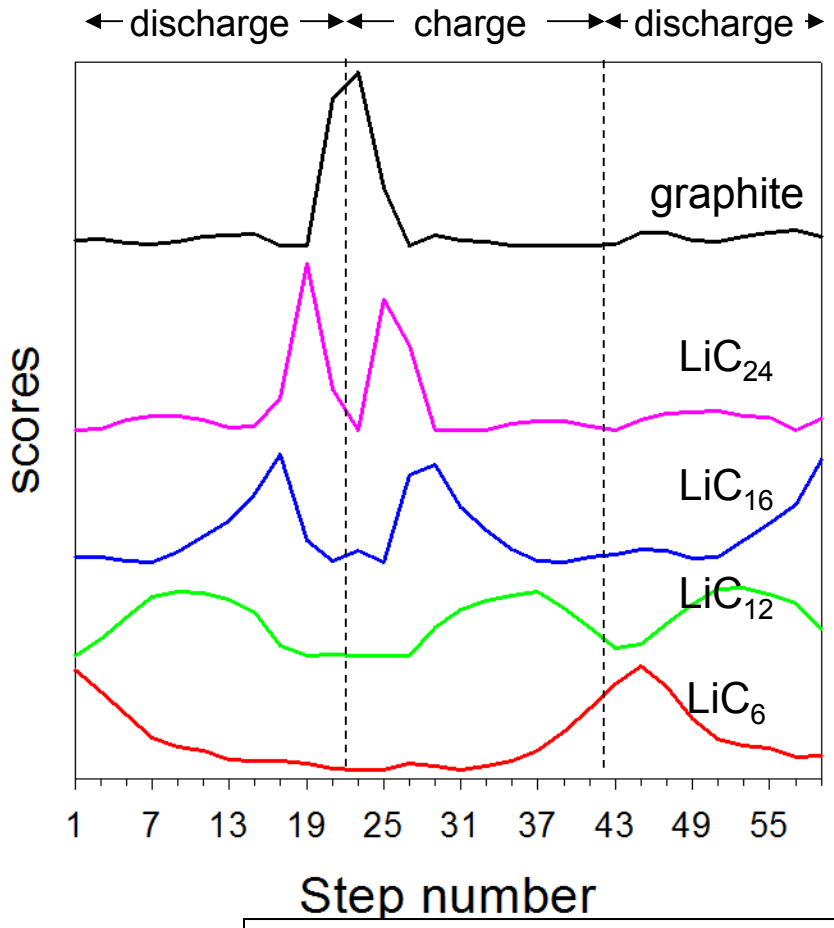
$4(LiC_{12}) \leftrightarrow 3(LiC_{16}) + (Li^+,e^-)$

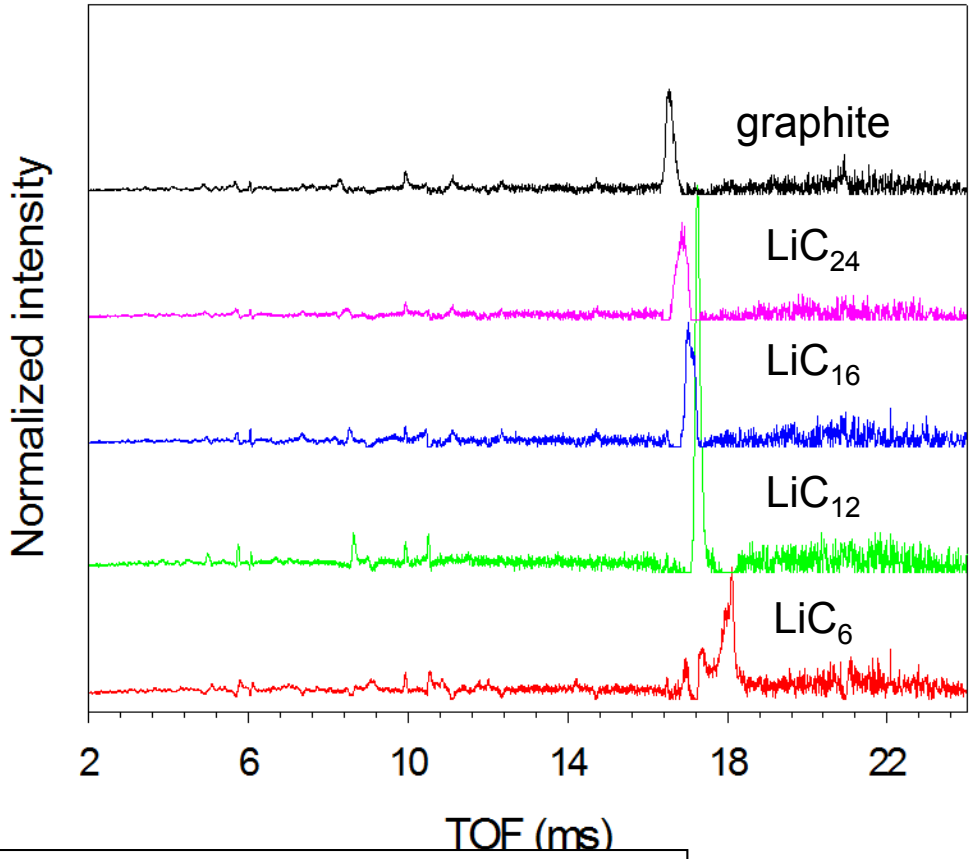
$3(LiC_{16}) \leftrightarrow 2(LiC_{24}) + (Li^+,e^-)$

$2(LiC_{24}) \leftrightarrow C(gr) + 2(Li^+,e^-)$

In: Rodriguez, Van Benthem, Ingersoll and Vogel, *In-situ Analysis of LiFePO4 Batteries: Signal Extraction by Multivariate Analysis*, 58th Annual Denver X-ray Conference , Colorado Springs, CO

- PCA with full histograms (150° banks) yielded 6 components.
  - Inert component representing average pattern
  - 5 components which can be assigned to various stagings of LiC_x

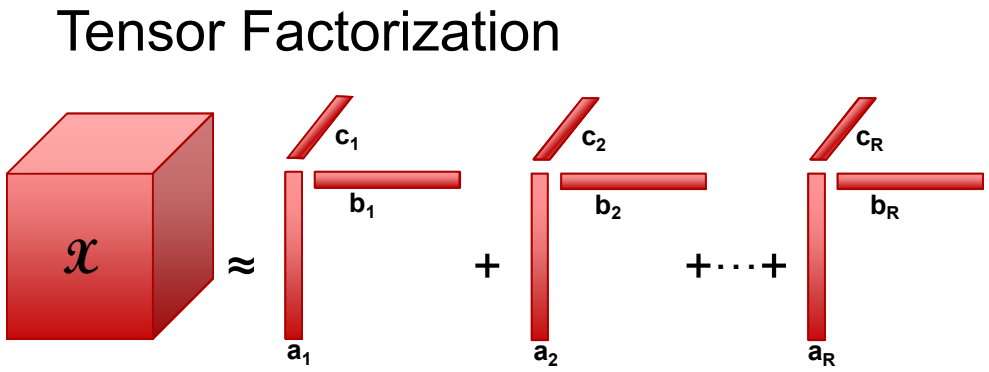




Anode staging tracks well with battery cycling

Three-Way Analysis Methods

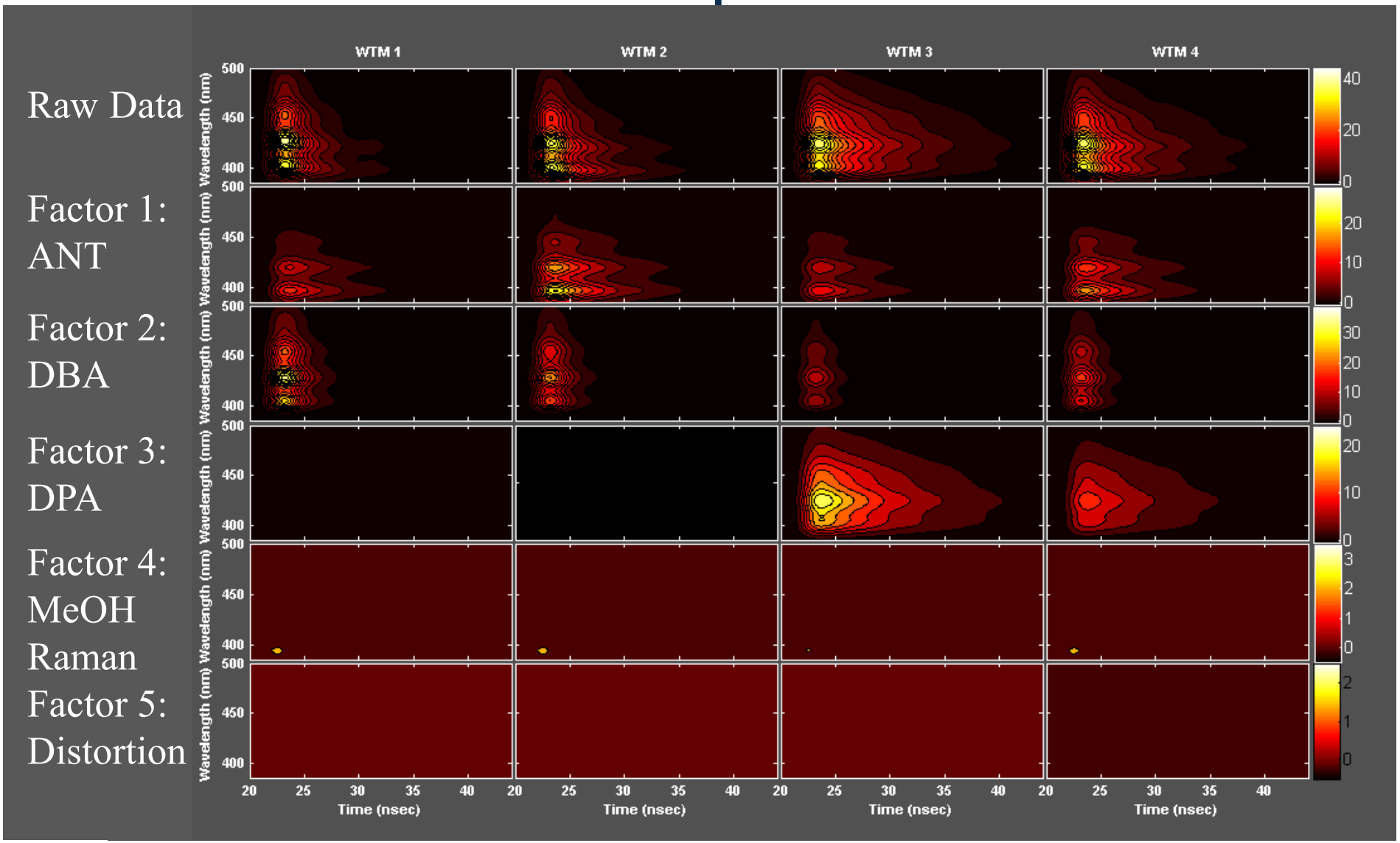
Sandia National Laboratories



- Tensor factorizations of multi-way data
  - Parallel factor analysis (PARAFAC)
  - Nonnegative tensor factorization (NTF)
  - Similar in idea to least squares matrix techniques: principal component analysis (PCA), singular value decomposition (SVD), multivariate curve resolution (MCR)
- When applied to a data array, data are modeled as a mixture of factors, each with its own triad of signature factors

Wavelength-Time Matrix Data and PARAFAC Pure Component WTMs

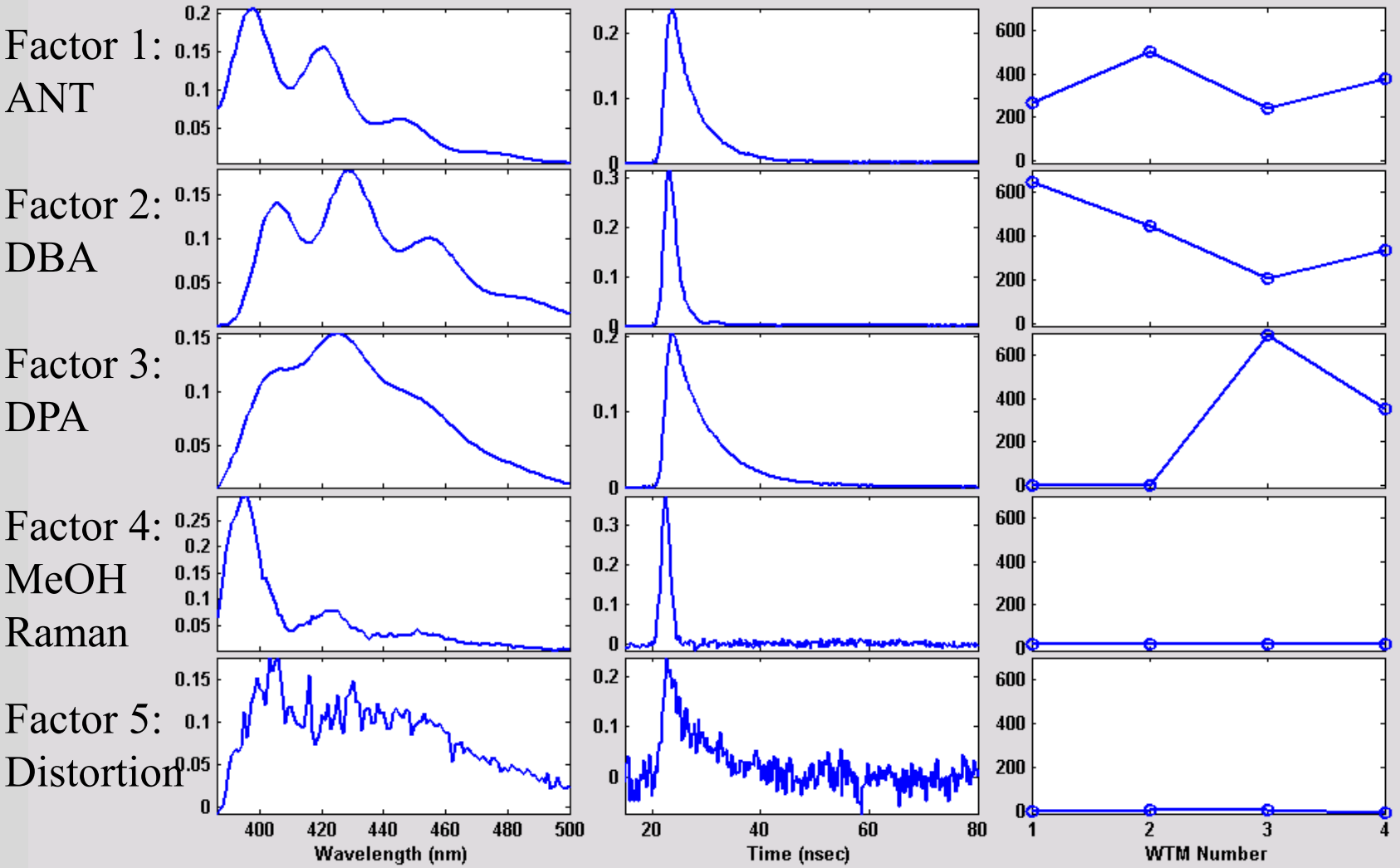
Sandia National Laboratories



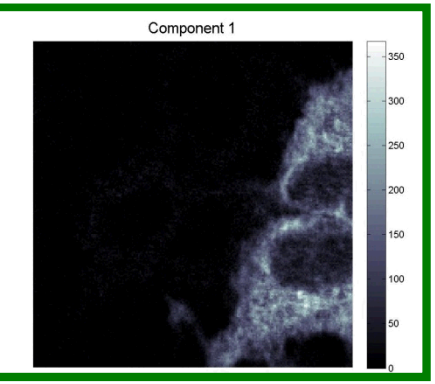
In M.H. Van Benthem and G.D. Gillispie, "*Nanosecond Time-Domain Fluorescence: A Mixture Analysis Perspective*", PittCon 2010, Orlando, FL

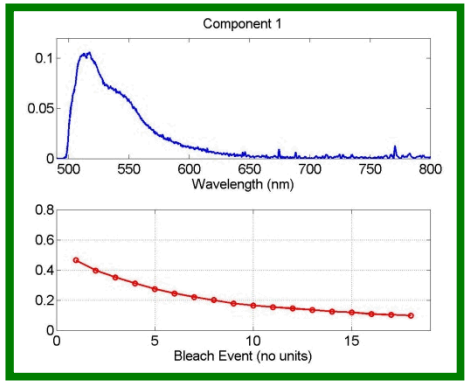
PARAFAC Factors

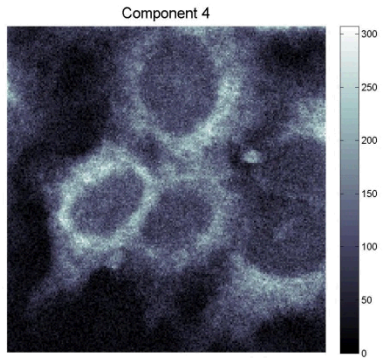
Sandia National Laboratories

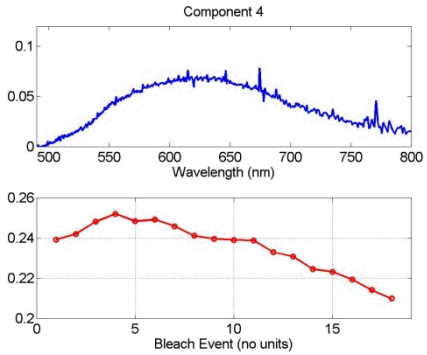


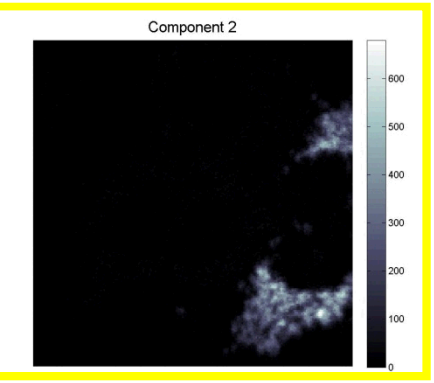
GFP-YFP-AF PARAFAC Model

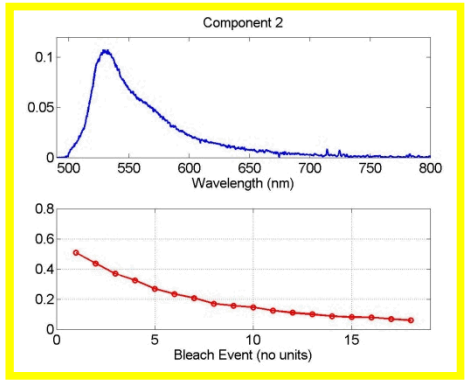


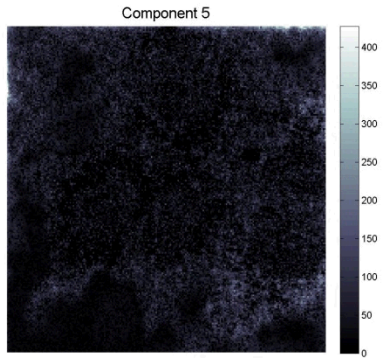


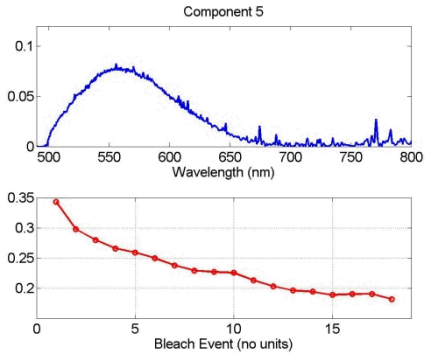


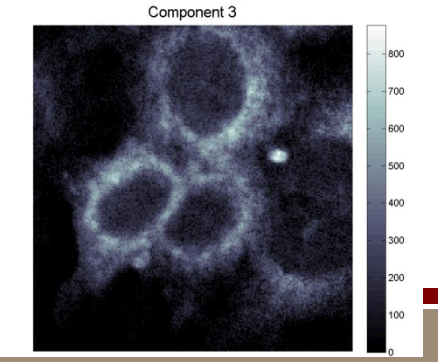


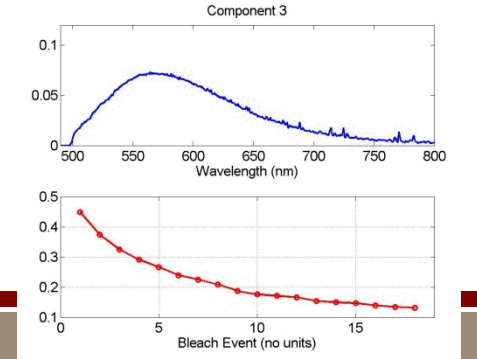












PARAFAC easily finds GFP and YFP fluors and returns their photobleach curves and their cell locations.

In: Van Benthem MH, *et al., Trilinear Analysis Of Images Obtained With A Hyperspectral Imaging Confocal Microscope. J. Chemom., 2008;* ***22 (9), 491-499.***

Acknowledgments

- This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.
- Michael R. Keenan (SNL), Paul G. Kotula (SNL), N. Rob Sorensen (SNL), Mark A. Rodriguez (SNL), David Ingersoll (SNL), Sven C. Vogel (LANL), James Griego (SNL), Kathy Rice (SNL), Ryan Davis (SNL), Howland D. T. Jones (SNL), David M. Haaland (SNL), Michael B. Sinclair (SNL), Ping Liu (UTMB), Allan R. Brasier (UTMB) and Gregory D. Gillispie (Fluorescence Innovations, Inc).

Photos placed in horizontal position with even amount of white space between photos and header

Photos placed in horizontal position
with even amount of white space
between photos and header

Backup Slides

Sandia National Laboratories

Exceptional service in the national interest

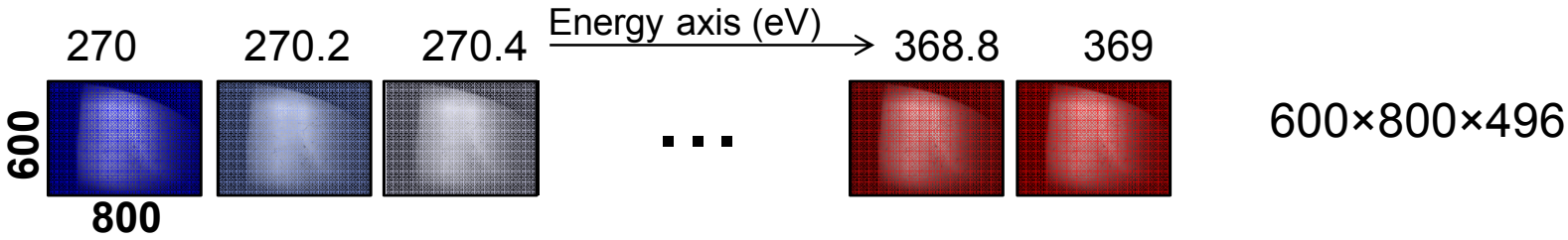
U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

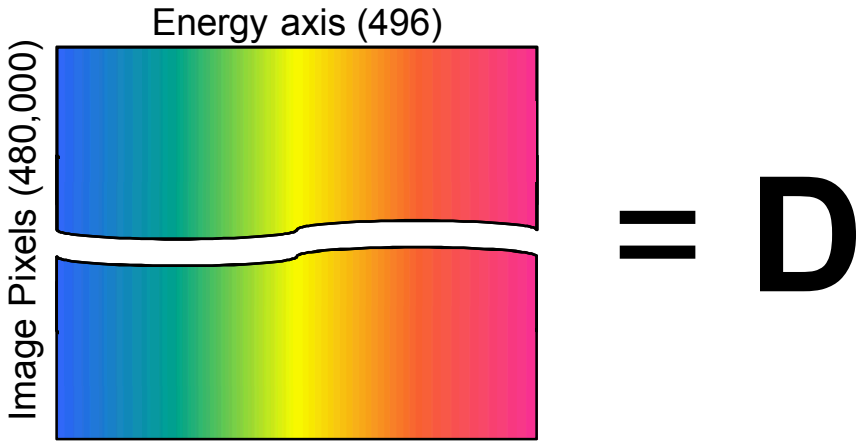
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Global Analysis: 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



Total Spectrum Fit