



Electron Microscopy in the Age of “Big Data”

CCEM Summer School on Electron Microscopy

June 5, 2019 - Hamilton, ON

Joshua Taillon

joshua.taillon@nist.gov

NIST Disclaimer

Certain commercial equipment, instruments, materials, vendors, and software are identified in this talk for example purposes and to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

A Brief Introduction

- Materials scientist by training
- Background in Materials characterization
 - TEM, FIB/SEM, EDS/EELS, etc.
- Stumbled into the world of "advanced" data analysis
 - Do not consider myself a software developer
- NIST Office of Data and Informatics
 - Not *actually* a microscopist any longer...

About NIST



Gaithersburg, MD



Boulder, CO

Office of Data and Informatics



- Provides data expertise and resources to NIST researchers
- Develops best practices to optimize FAIR data products
- Supports research into advanced manipulation, visualization, and analysis of large data sets

High-level Outline

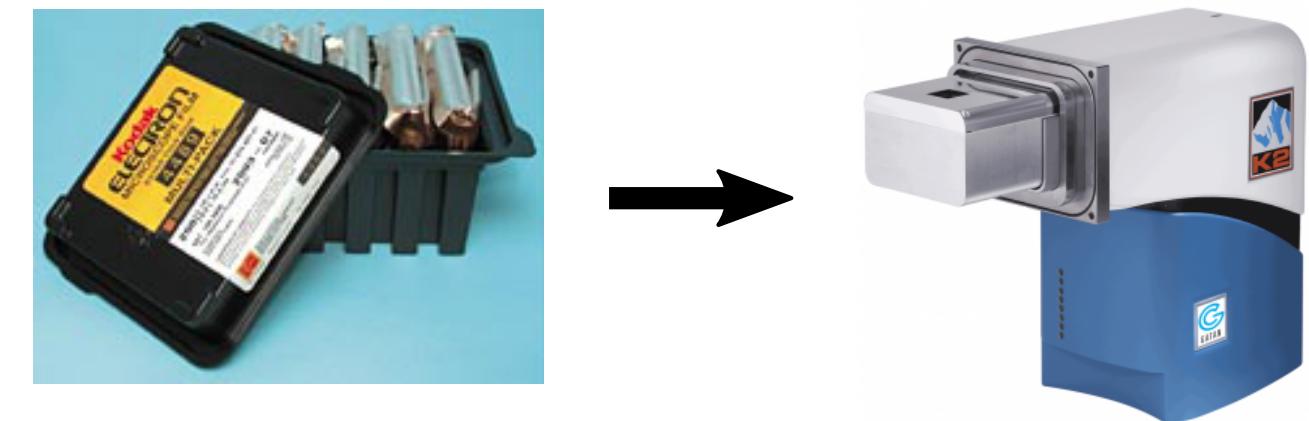
- What is *computational microscopy*?
- Real-world examples of "big data" analysis in EM
- The advent of open tools
 - Now you can get the same result as your neighbor!
- Deeper dive into "signal separation"
 - Methods, examples, gotchas, etc.

WHAT IS COMPUTATIONAL MICROSCOPY?

What is not CM?

This is personal
opinion! Feel free
to disagree...

- Digitization

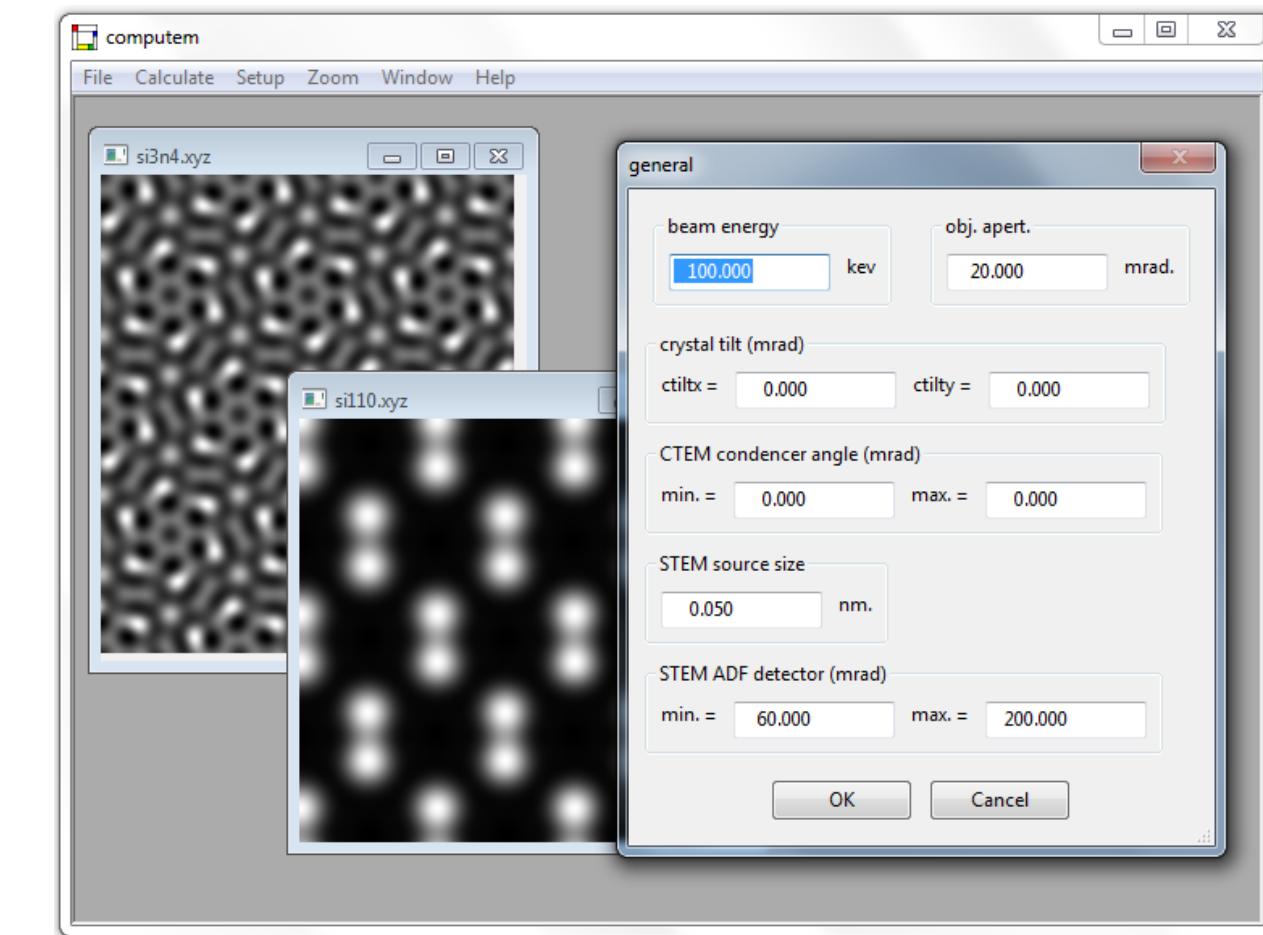


*From film to CCD
(Ted Pella; Gatan)*

What is not CM?

This is personal opinion! Feel free to disagree...

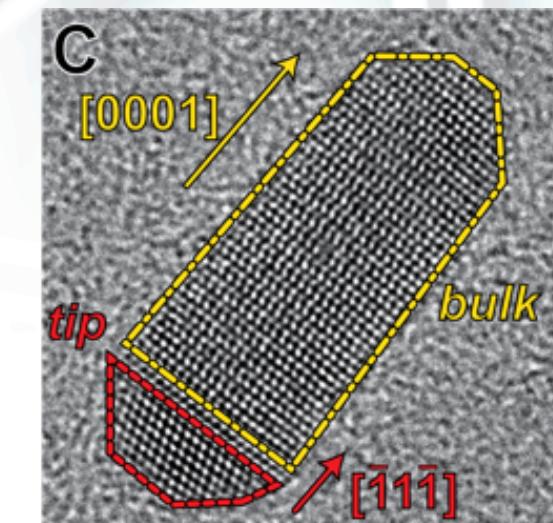
- Digitization
- Image simulation



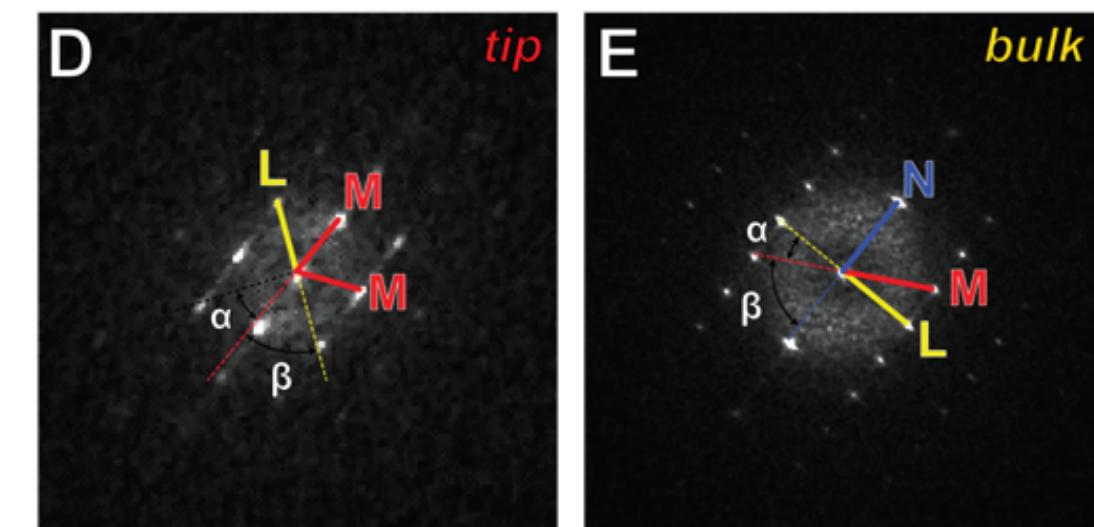
EJ Kirkland; Multislice TEM Simulation

What is not CM?

- Digitization
- Image simulation
- "Traditional" image analysis

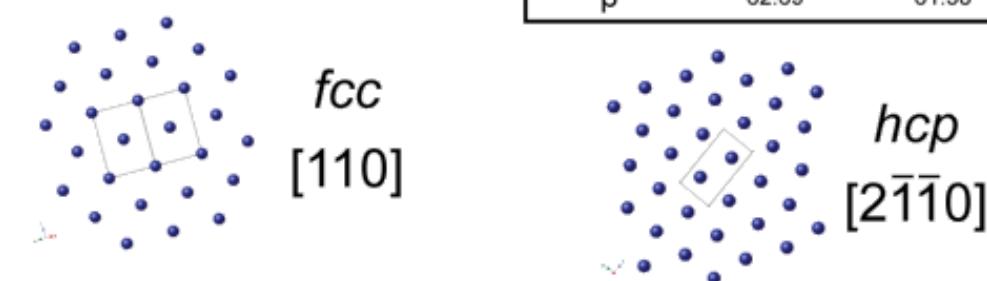


This is personal opinion! Feel free to disagree...



Parameter	Measured	Expected
L/M	1.162	1.155
α	36.16°	35.26°
β	54.63°	54.74°

Parameter	Measured	Expected
M/L	1.144	1.139
N/L	1.09	1.09
α	28.62°	28.62°
β	62.69°	61.38°



HRTEM Fourier analysis

Attempting a definition

This is personal
opinion! Feel free
to disagree...

Microscopy directed by or collected primarily for computational processes (as opposed to by/for the user directly)

Attempting a definition

This is personal
opinion! Feel free
to disagree...

- Relevant buzzwords:
 - Machine learning
 - Artificial intelligence
 - Autonomous measurement
 - Dynamic sampling
 - Compressive sensing
 - Sparse imaging

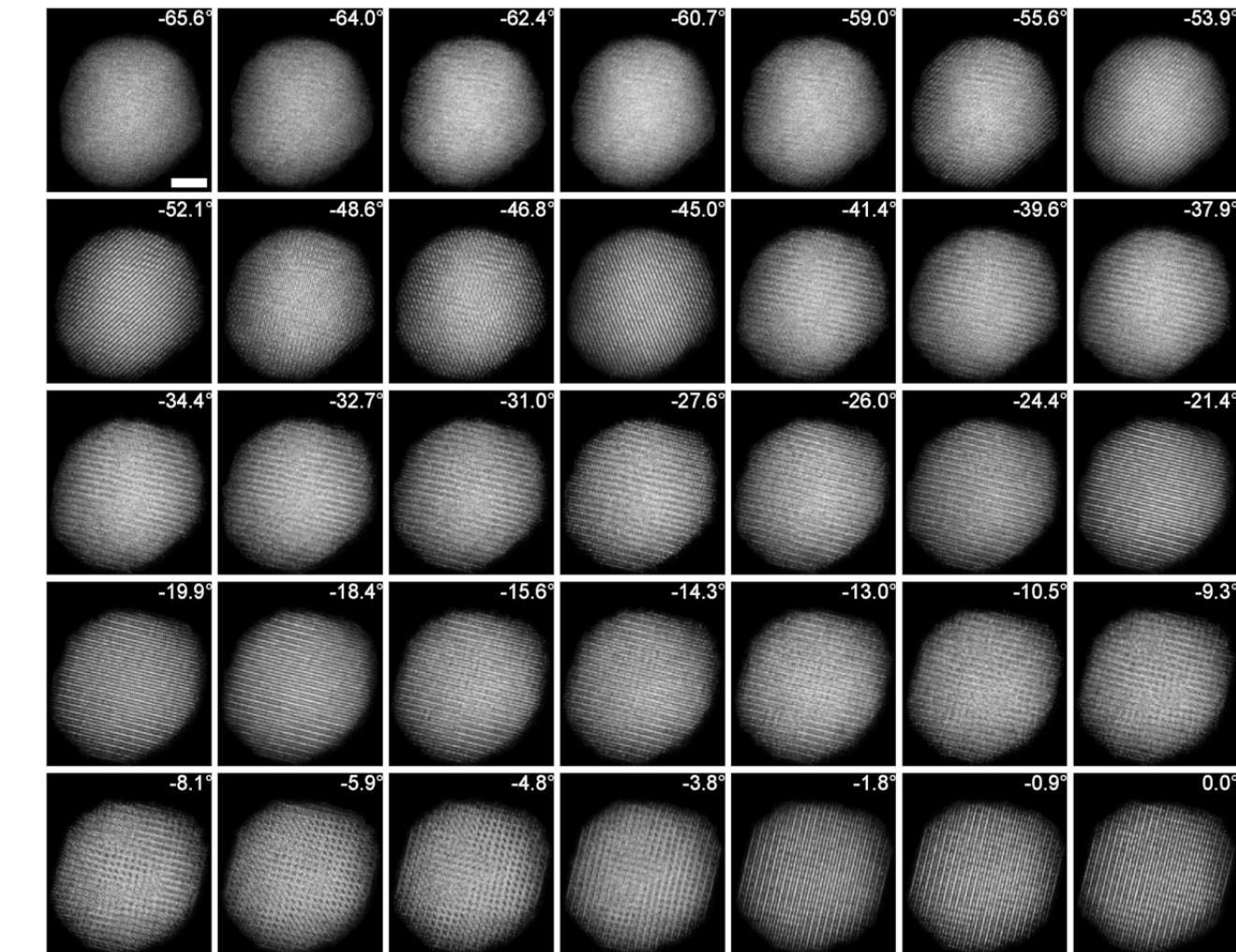
Why computational microscopy?

- **Statistical power**
 - Is your image actually *representative*?
- **Reproducibility is the default**
 - Computers only do what you tell them too (*we hope*)
- **Leverages massive advances in computational power**

SOME REAL-WORLD EXAMPLES OF COMPUTATIONAL MICROSCOPY AND "BIG DATA"

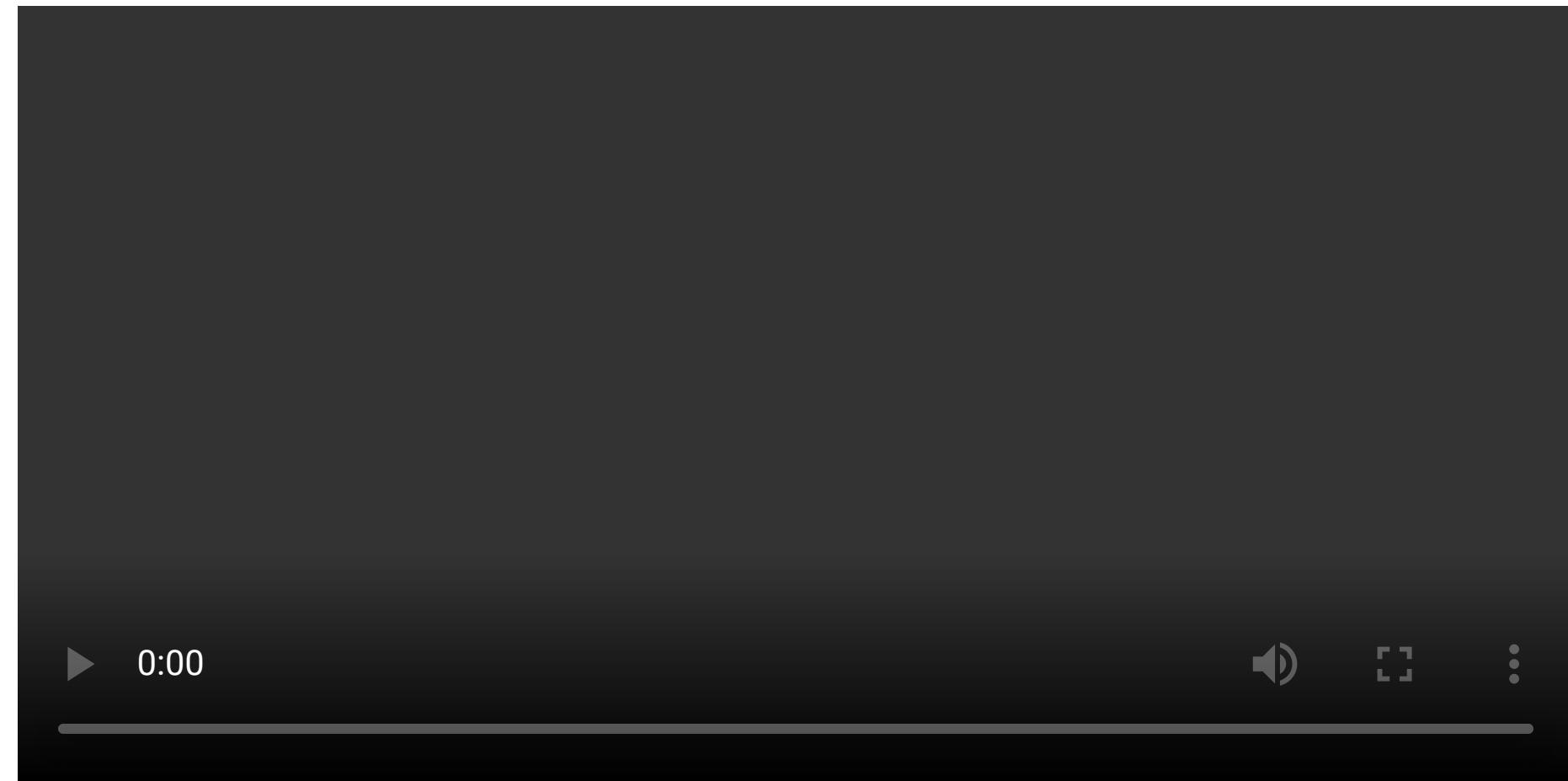
(1/5) Electron Tomography

- Y. Yang, J. Miao, et al., *Nature*, **542**, 75-79, 2017 ([link](#))
- 68 ADF-STEM images of FePt nanoparticle
- Located 6,569 Fe and 16,627 Pt atoms to 21.6 pm precision (correlated with multislice simulations)
- Calculated SROP to distinguish individual grains
- 3D information at the atomic scale



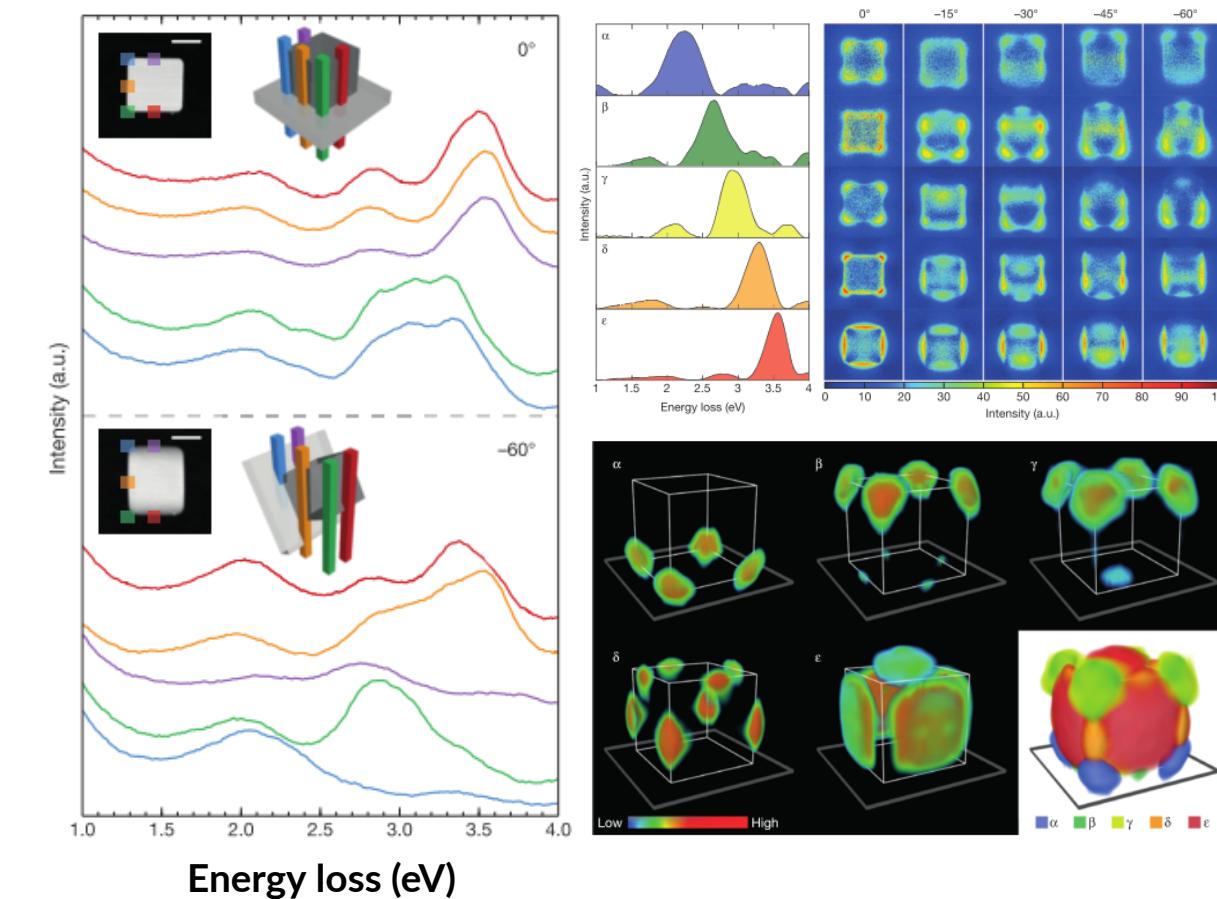
Selection of nanoparticle tilt images

(1/5) Electron Tomography



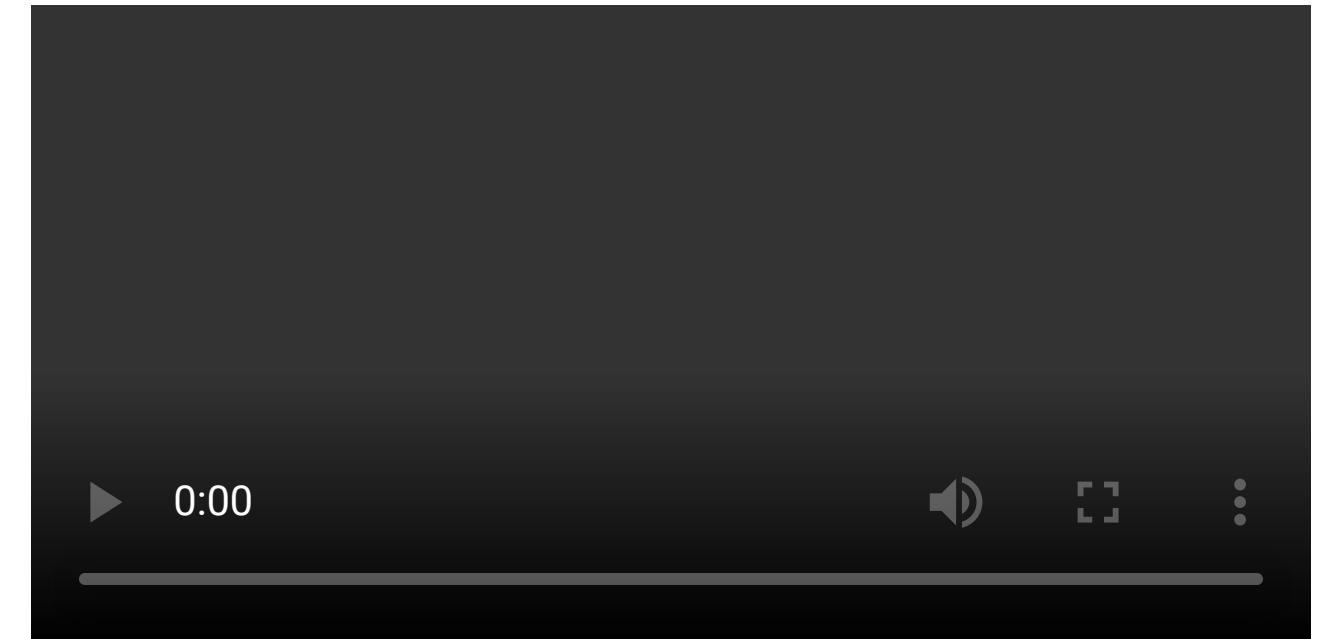
(2/5) ML for Factor Analysis

- O. Nicoletti, P. Midgley, et al., *Nature*, 502, 80-84, 2013 ([link](#))
- Non-negative matrix factorization of EELS spectra
- Identifying meaningful spectral components in a sea of overlapping signals
- Combine with tilt-tomography for 3D information
- Identified nanoparticle plasmon resonances



(2/5) ML for Factor Analysis

- O. Nicoletti, P. Midgley, et al., *Nature*, 502, 80-84, 2013 ([link](#))
- Non-negative matrix factorization of EELS spectra
- Identifying meaningful spectral components in a sea of overlapping signals
- Combine with tilt-tomography for 3D information
- Identified nanoparticle plasmon resonances



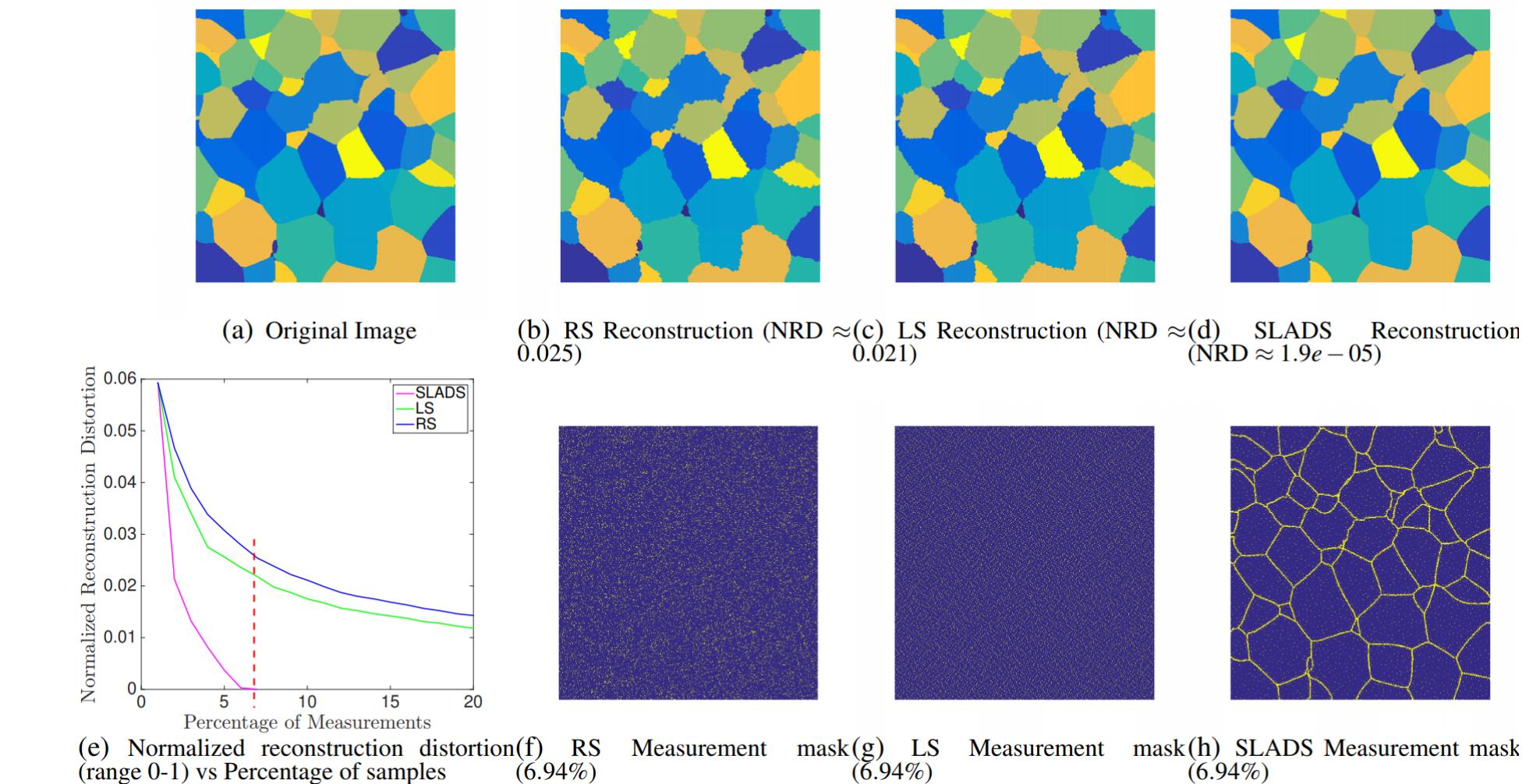
(3/5) Autonomous Metrology

- A. Kusne, I. Takeuchi, et al.,
Nanotechnology, **26**, 444002, 2015
([link](#))
- High-throughput XRD for combinatorial materials discovery
- Autonomous phase diagram mapping of composition spread wafer
- Phase diagram is estimated at each step based on collected data and physics-informed ML algorithms
- Unsupervised AI determines new composition to measure to best estimate phase diagram



(4/5) Dynamic Sampling in SEM

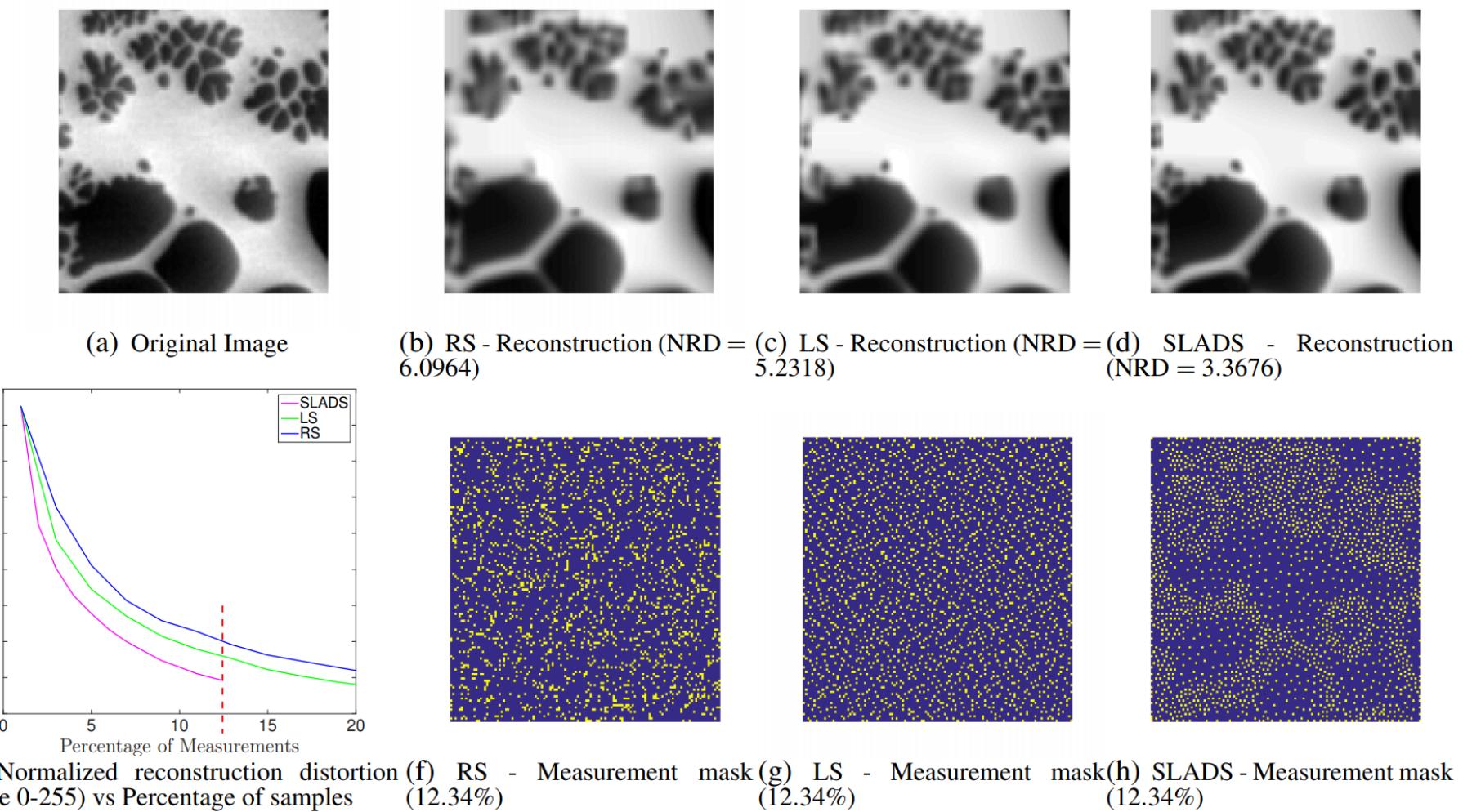
- G. Godaliyadda, C. Bouman, et al.,
Electronic Imaging, 19, 1-8, 2016 ([link](#))
- Supervised Learning Approach to Dynamic Sampling (SLADS)
- Sparse imaging and weighted inpainting reconstruction
- Train algorithm offline to measure how much certain types of pixels reduce overall distortion
- Online, pick new pixels to reduce expected distortion in reconstruction



Simulated EBSD patterns

(4/5) Dynamic Sampling in SEM

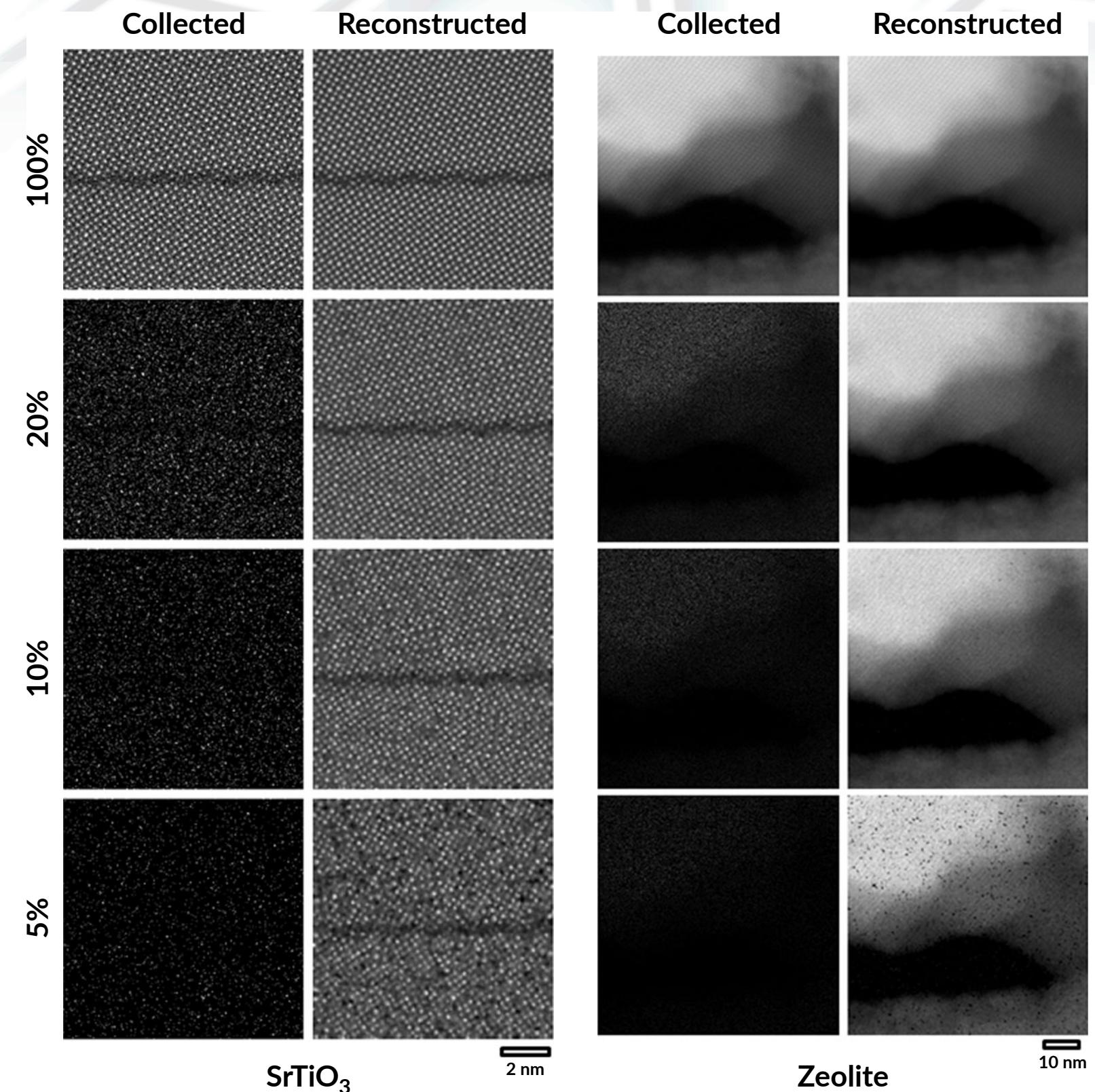
- G. Godaliyadda, C. Bouman, et al.,
Electronic Imaging, 19, 1-8, 2016 ([link](#))
- Supervised Learning Approach to Dynamic Sampling (SLADS)
- Sparse imaging and weighted inpainting reconstruction
- Train algorithm offline to measure how much certain types of pixels reduce overall distortion
- Online, pick new pixels to reduce expected distortion in reconstruction



Experimental SEM images

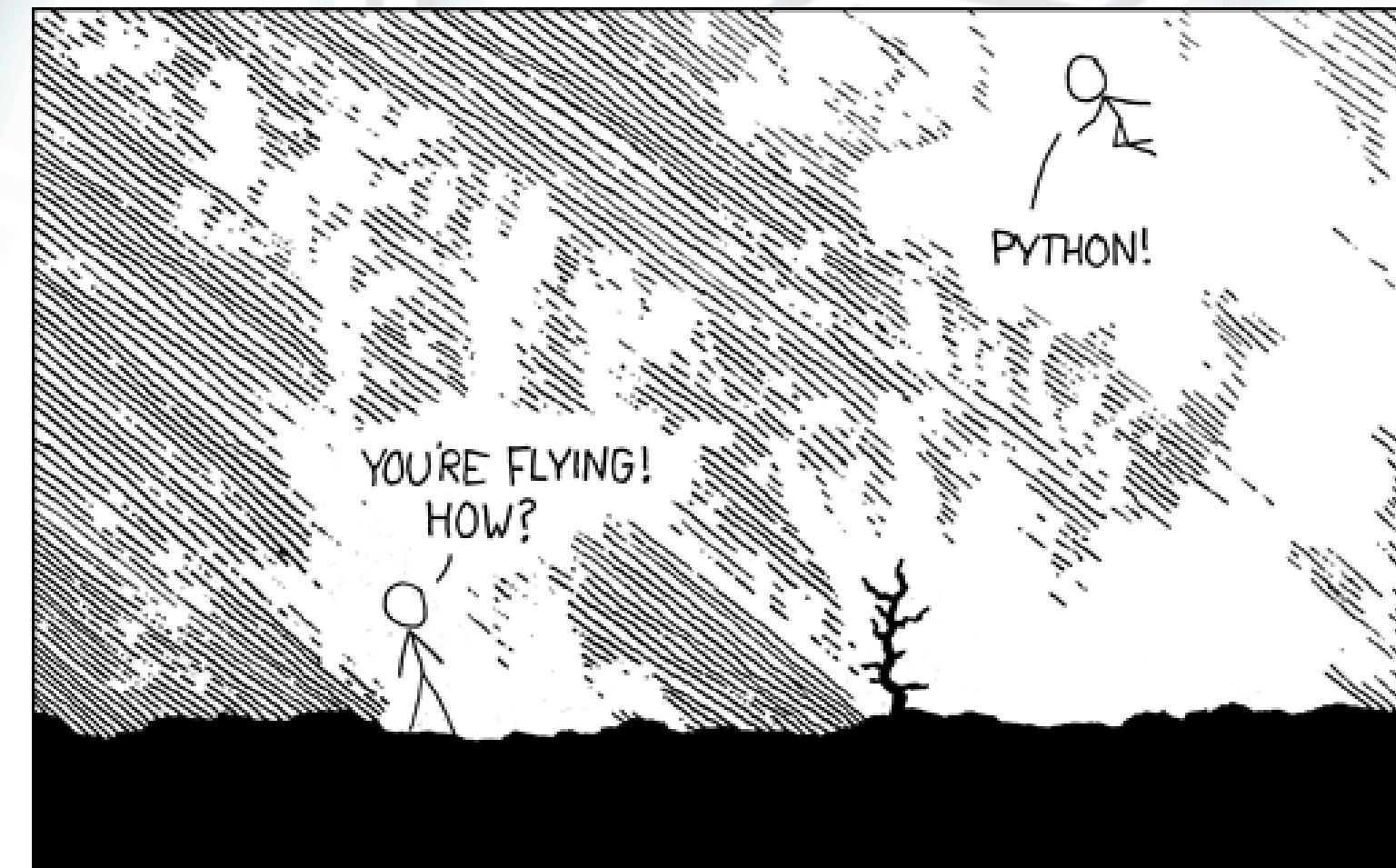
(5/5) STEM CS

- A. Stevens, N. Browning, et al.,
Microscopy, **63**, 41-51, 2014.
[\(link\)](#)
- Intentionally acquire image at severe undersampling conditions
- Use ℓ_1 -norm convex optimization to fill in the missing details
- An interesting means to get around the Nyquist-Shannon limit
- Demonstrated with random sampling in both STEM and SEM



Compressed Sensing STEM reconstructions

OPEN TOOLS FOR ELECTRON MICROSCOPY



I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!
I
HELLO WORLD IS JUST
print "Hello, world!"

I DUNNO...
DYNAMIC TYPING?
WHITESPACE?

COME JOIN US!
PROGRAMMING IS FUN AGAIN!
IT'S A WHOLE NEW WORLD UP HERE!

BUT HOW ARE YOU FLYING?

I JUST TYPED
import antigravity
THAT'S IT?

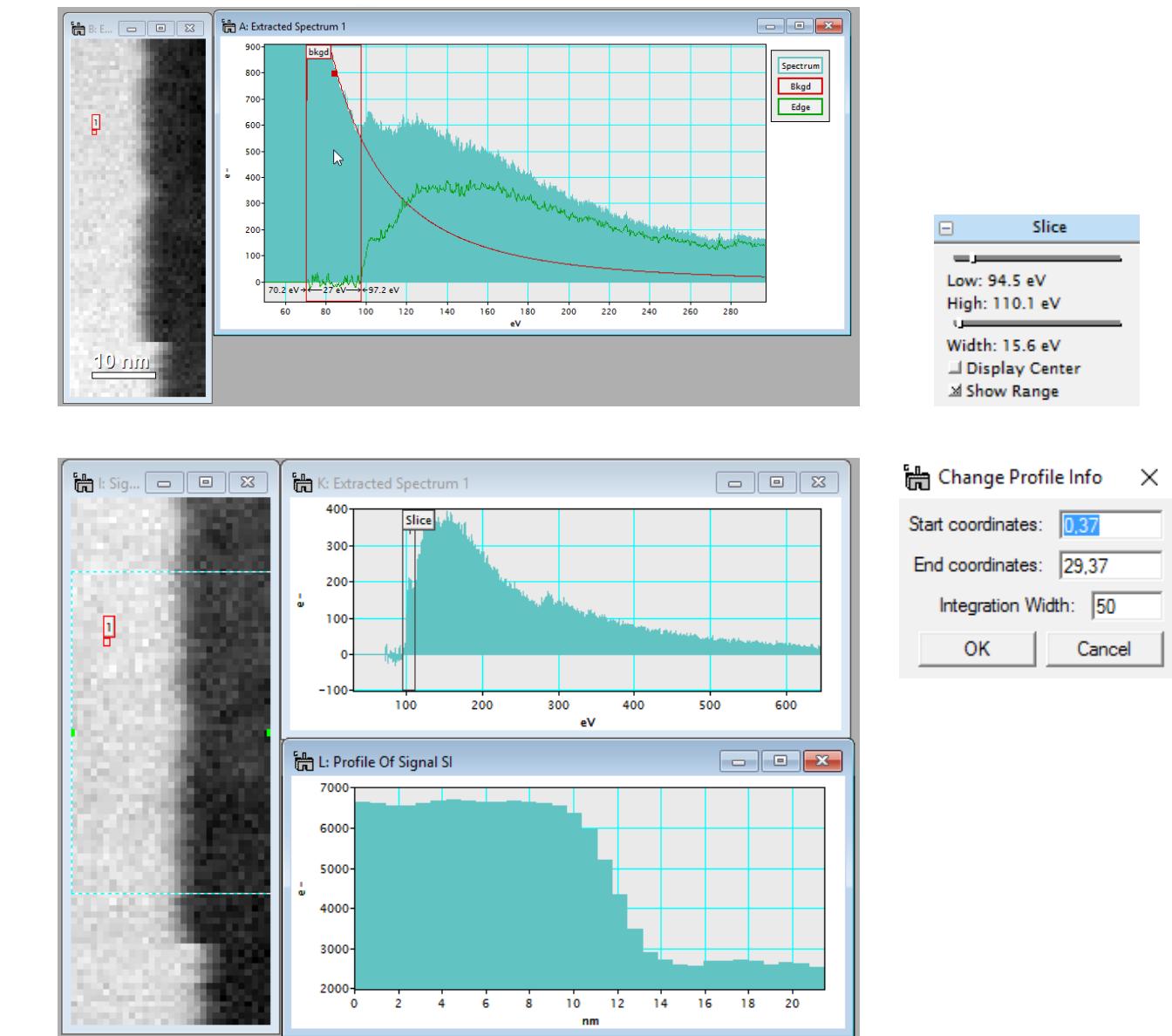
... I ALSO SAMPLED
EVERYTHING IN THE
MEDICINE CABINET
FOR COMPARISON.

BUT I THINK THIS
IS THE PYTHON.

Randall Munroe - [xkcd](#)

A "typical" EM analysis

- One or more software packages typically necessary
- Often vendor-provided
- GUI-driven with many options, sometimes “black-box”
- Typically, no log recorded
 - Hope you keep a good notebook!
 - Can you reproduce your post-doc's analysis?
- Tightly integrated with equipment/acquisition



*Extracting EELS intensity profile in Gatan
Digital Micrograph*

A better way?

- A "typical" EELS analysis, but totally reproducible
- Computation within a “notebook” environment
- Seamless mixing of notetaking, mathematics, and data analysis
- Notebook is rendered in any web browser
- Version controlled and exportable to PDF, HTML, Markdown, etc.



Notebook compared to GUI

A better way?

Please join one of the HyperSpy tutorials this week if you want to learn more!

```
import hyperspy.api as hs
s = hs.load('EELS Spectrum Image.dm3')
s_sig = s.remove_background(signal_range=(70.2, 97.2))
s_crop = s_sig.inav[:, 12:62].isig[94.5:110.1]
s_crop.sum(axis=(1,-1)).plot()
```

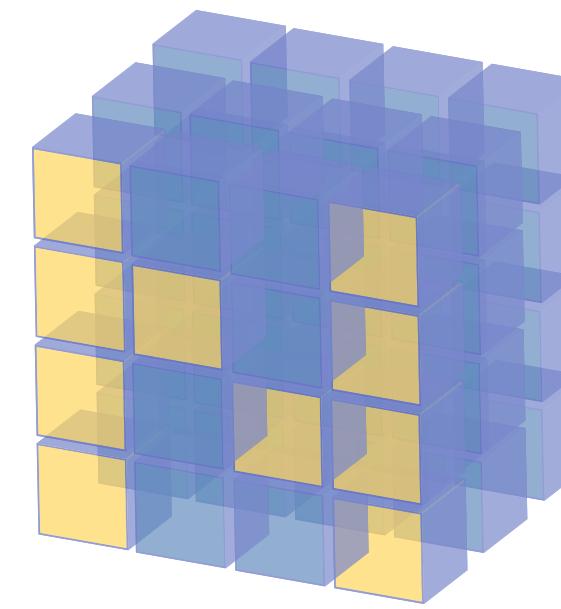
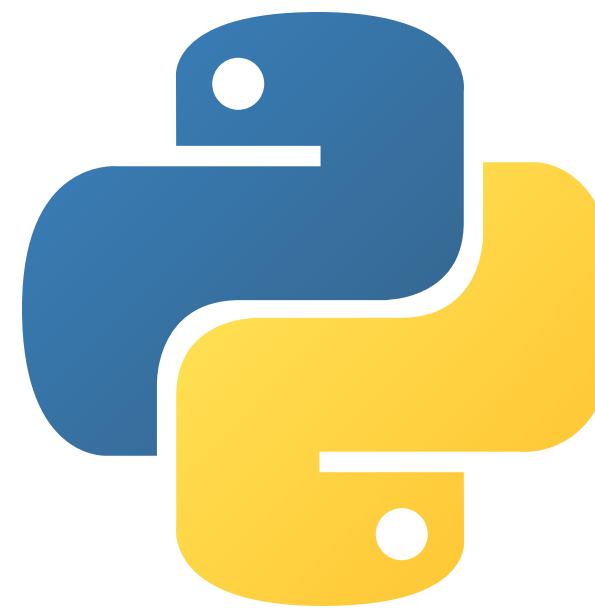
Only 5 lines of code!

A burgeoning ecosystem

General Purpose		Others	
HyperSpy	http://hyperspy.org/	PyQSTEM	https://github.com/jacobjma/PyQSTEM
Nion Swift	https://nionswift.readthedocs.io/en/stable/	HRTEMFringeAnalyzer	https://github.com/ialxn/HRTEMFringeAnalyzer
pycroscopy	https://github.com/pycroscopy/pycroscopy	Atomap	https://atomap.org/
Pixelated STEM		Tomography	
pyXem	https://pyxem.github.io/pyxem/	tomopy	https://tomopy.readthedocs.io/en/latest/
pixStem	https://pixstem.org/	tomotools	https://github.com/AndrewHerzing/tomotools
LiberTEM	https://github.com/LiberTEM/LiberTEM	tomviz	https://tomviz.org/
fpd	https://gitlab.com/fpdpy/fpd/		

And many others (all free!)...

All built on Python!



Click to visit each project

UNSUPERVISED HYPERSPECTRAL SIGNAL SEPARATION (SPECTRAL UNMIXING)

Overview

- What techniques is this applicable to?
- What is unmixing (phase mapping)?
- Vendor options vs. open solutions
- Demonstration of a few different unmixing algorithms

Applicable techniques

- **Raster-based scanning spectroscopic methods:**
 - Scanning transmission electron microscopy (STEM)
 - Electron energy-loss spectroscopy (EELS) and EDS
 - Scanning electron microscopy (SEM)
 - X-ray energy dispersive spectroscopy (EDS)
 - X-ray fluorescence spectroscopy mapping (XRF and μ XRF)
 - Infrared spectroscopy mapping (FTIR)
- **Image-based methods**
 - Time series images of kinetic behaviors
 - Through-focal series in TEM (extracting true structure)

Applicable techniques

- Raster-based scanning spectroscopic methods:
 - Scanning transmission electron microscopy (STEM)
 - Electron energy-loss spectroscopy (EELS) and EDS
 - Scanning electron microscopy (SEM)
 - X-ray energy dispersive spectroscopy (EDS)
 - X-ray fluorescence spectroscopy mapping (XRF and μ XRF)
 - Infrared spectroscopy mapping (FTIR)
- Image-based methods
 - Time series images of kinetic behaviors
 - Through-focal series in TEM (extracting true structure)

X-ray energy dispersive spectroscopy (EDS)

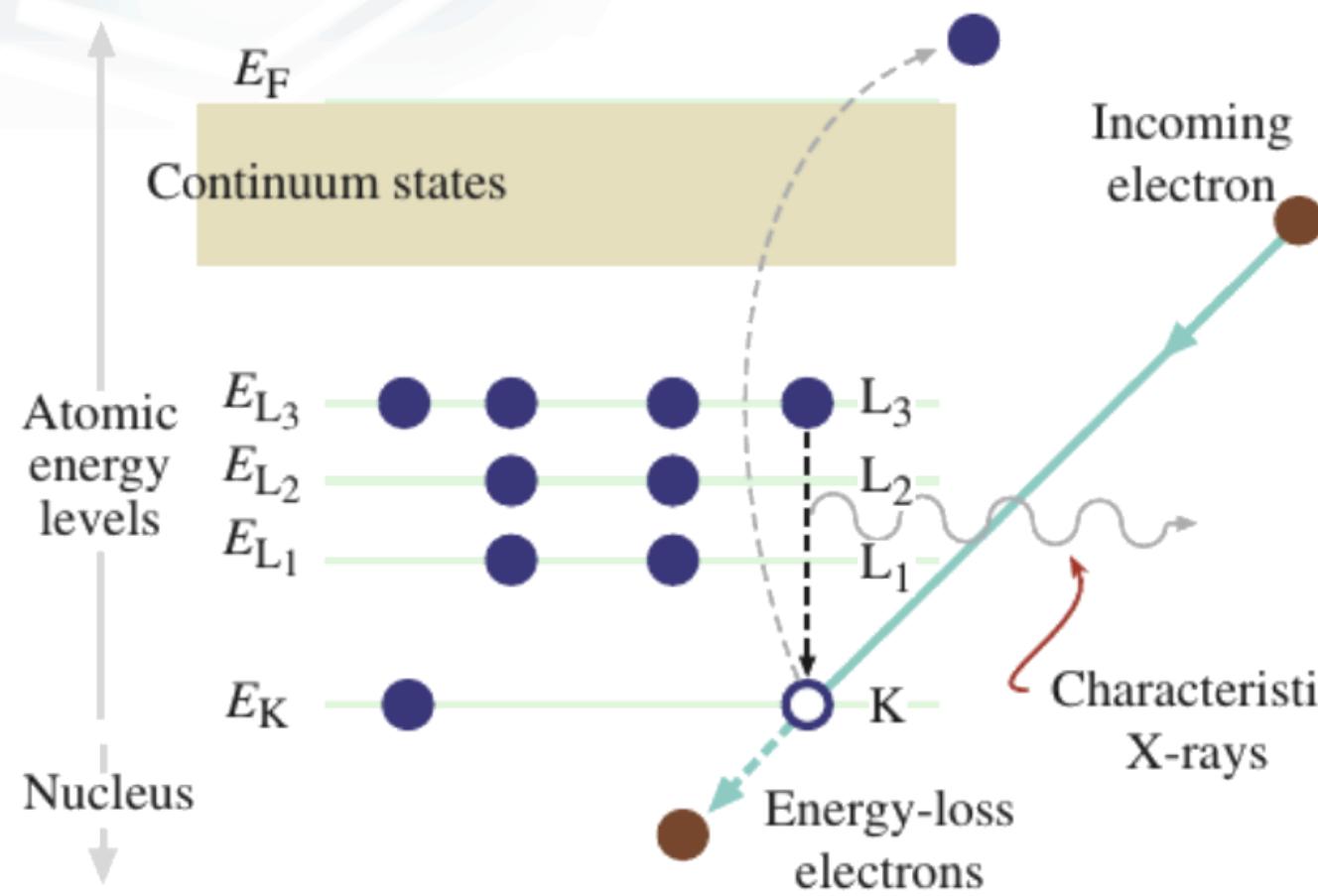
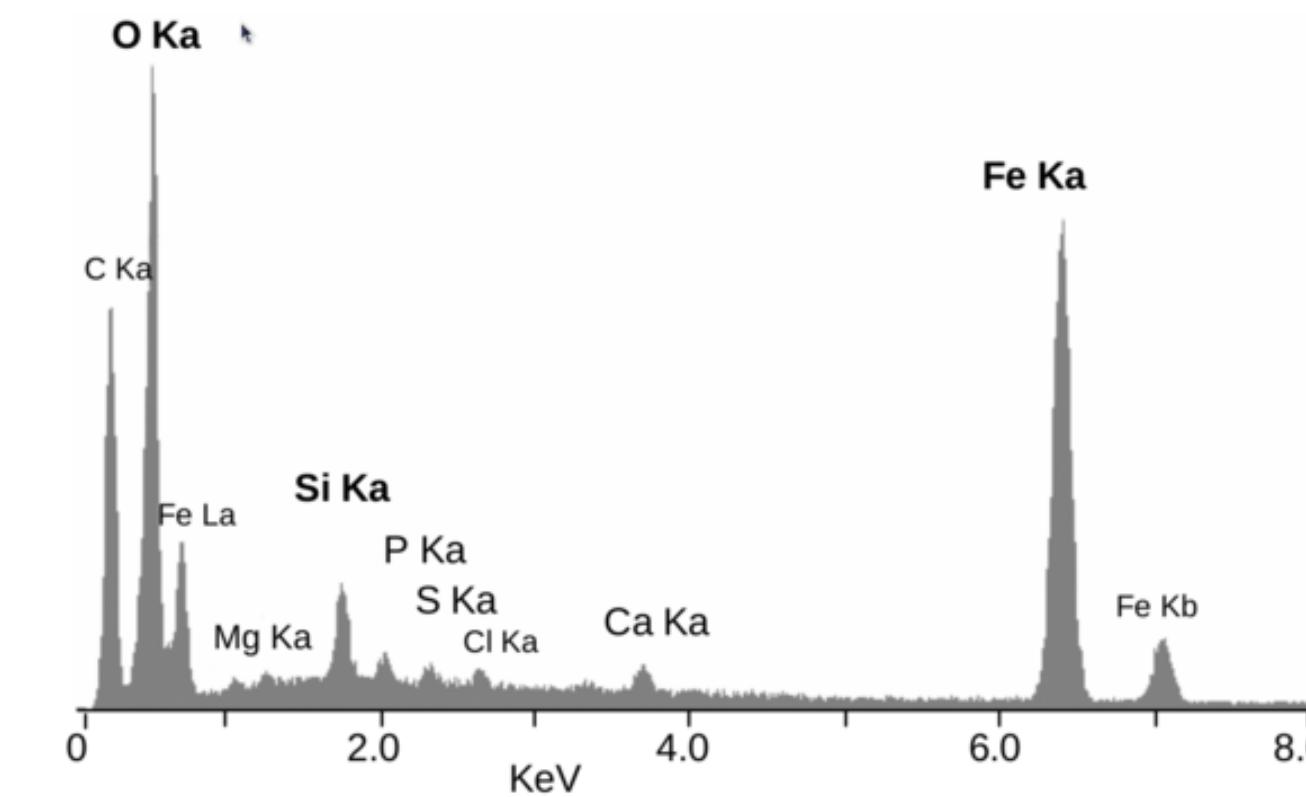


FIGURE 4.2. The ionization process. An inner (K) shell electron is ejected from the atom by a high-energy electron. When the hole in the K shell is filled by an electron from the L shell, characteristic (K_{α}) X-ray emission occurs. The beam electron loses energy but continues on through the specimen.

Williams and Carter, *Transmission Electron Microscopy*, p. 55 (2009)



EDS spectrum of the mineral crust of the vent shrimp [Rimicaris exoculata](#). Most of these peaks are X-rays given off as electrons return to the K electron shell. ([K-alpha](#) and [K-beta](#) lines) One peak is from the L shell of iron. ([source](#))

What is hyperspectral unmixing?

- Start with some hyperspectral data:

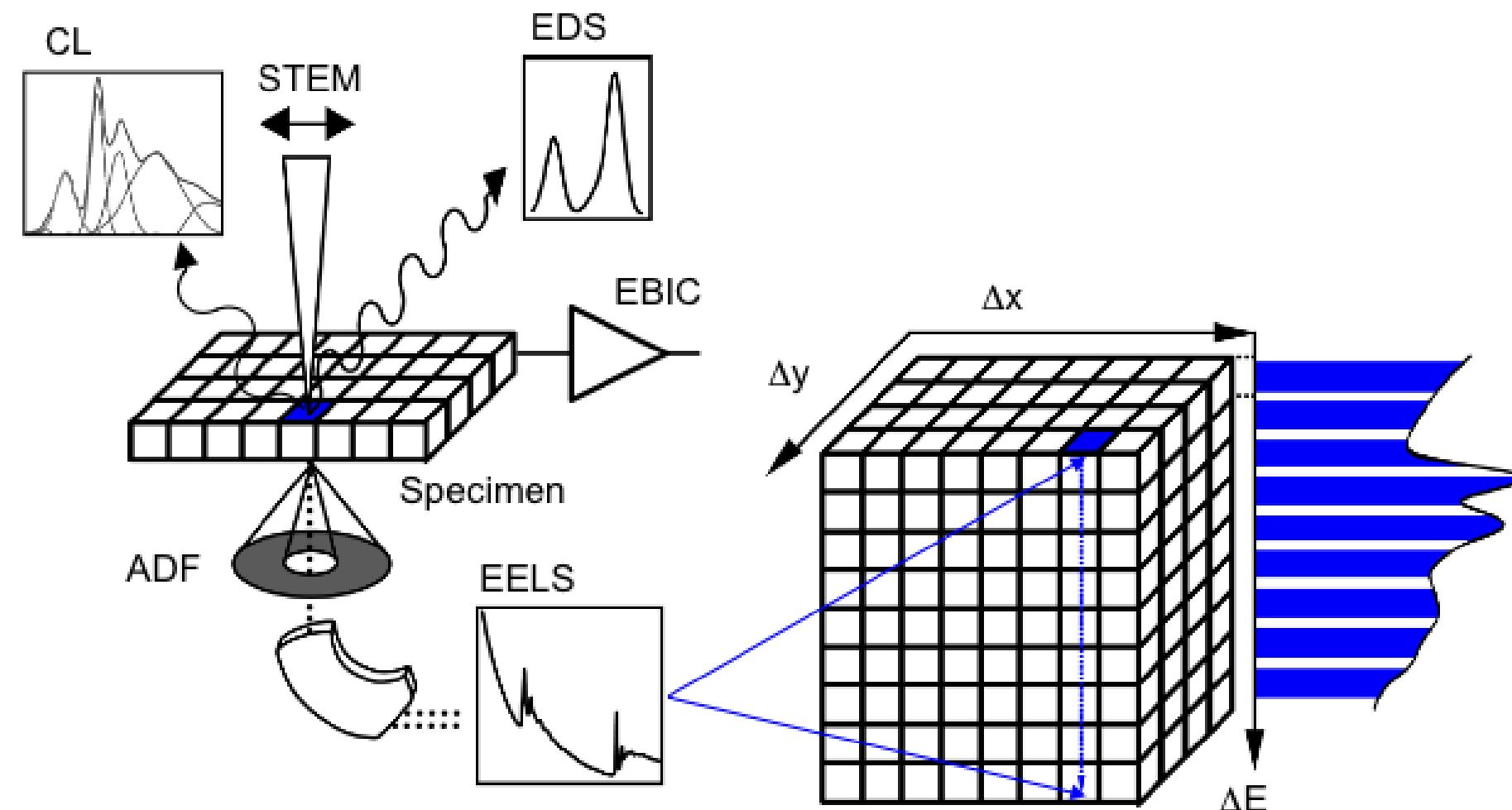
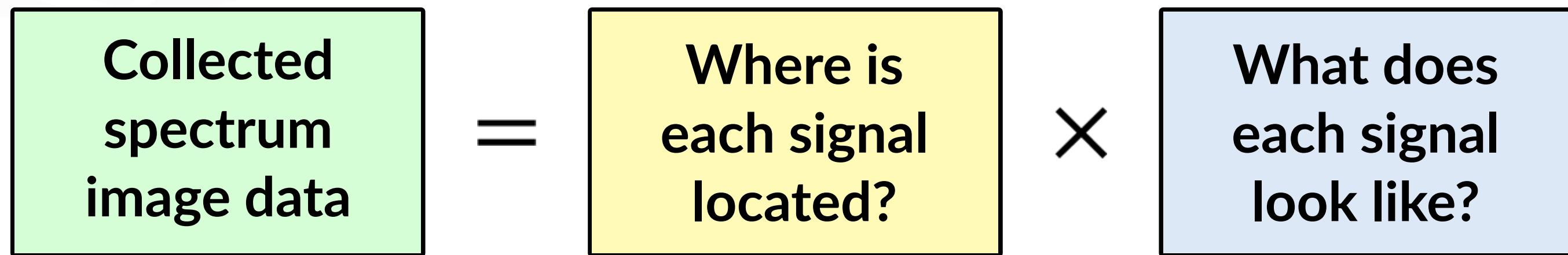


Image courtesy of Gatan, Inc.

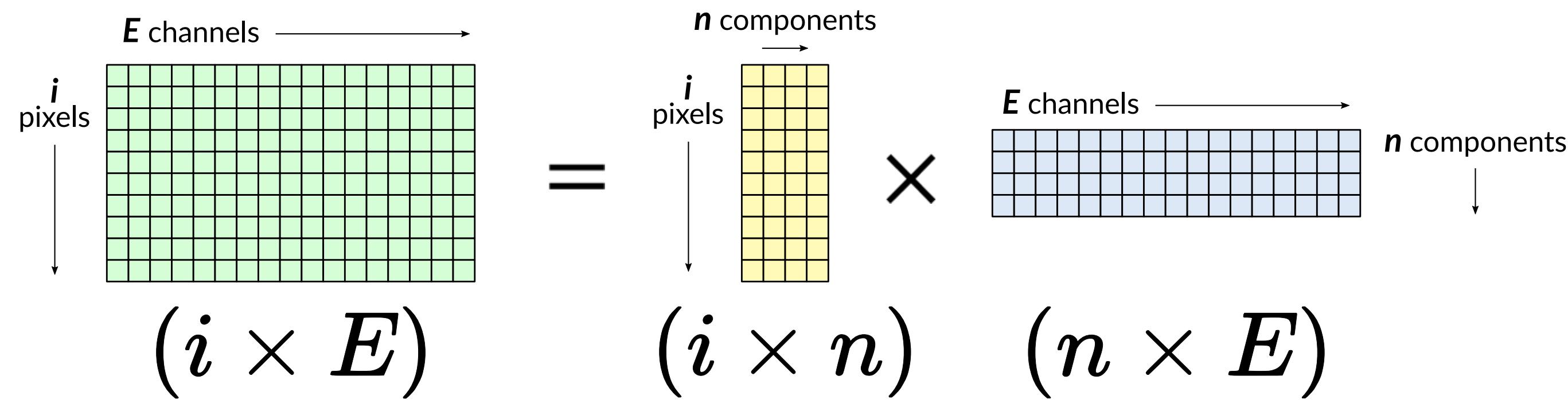
What is hyperspectral unmixing?



$$\mathbf{D}_{(x,y),E} = \mathbf{W}_{(x,y),n} \times \mathbf{S}_{n,E} + \epsilon_{(x,y),E}$$

$$\mathbf{D}_{i,E} = \mathbf{W}_{i,n} \times \mathbf{S}_{n,E} + \epsilon_{i,E}$$

What is hyperspectral unmixing?



What do the vendors offer?

- If you've used a modern EDS software package, you probably have done hyperspectral unmixing (they usually call it *phase mapping*)...

What do the vendors offer?

- If you've used a modern EDS software package, you probably have done hyperspectral unmixing (they usually call it *phase mapping*)...

Strengths/challenges of *vendor* options

The Good

- Simple point-and-click operation
- Tight integration
 - Collection, visualization, reporting, etc.
- Usually runs in real-time
- Integration with other data sources (e.g. EBSD)
- Generally "just works"

The Not So Good

- Extremely "black box"
- Reproducibility (!)
 - Configurable options with little understanding of why
- What are the uncertainties?
- Tied to software (\$)
- Choice of vendor should not change the scientific result

Strengths/challenges of *open source* options

The Not So Good

- Usually not point-and-click
- (Can be) difficult to access raw data from the vendor software
- Generally only post-processing
- Learning curve can be substantial
- Can take a lot more fiddling

The Good

- You know what's happening
- Reproducibility (!)
 - Anyone can recreate your analysis (including you!)
- Uncertainty can be understood
- Usually free
- Results do not depend on vendor

Offline "phase mapping"

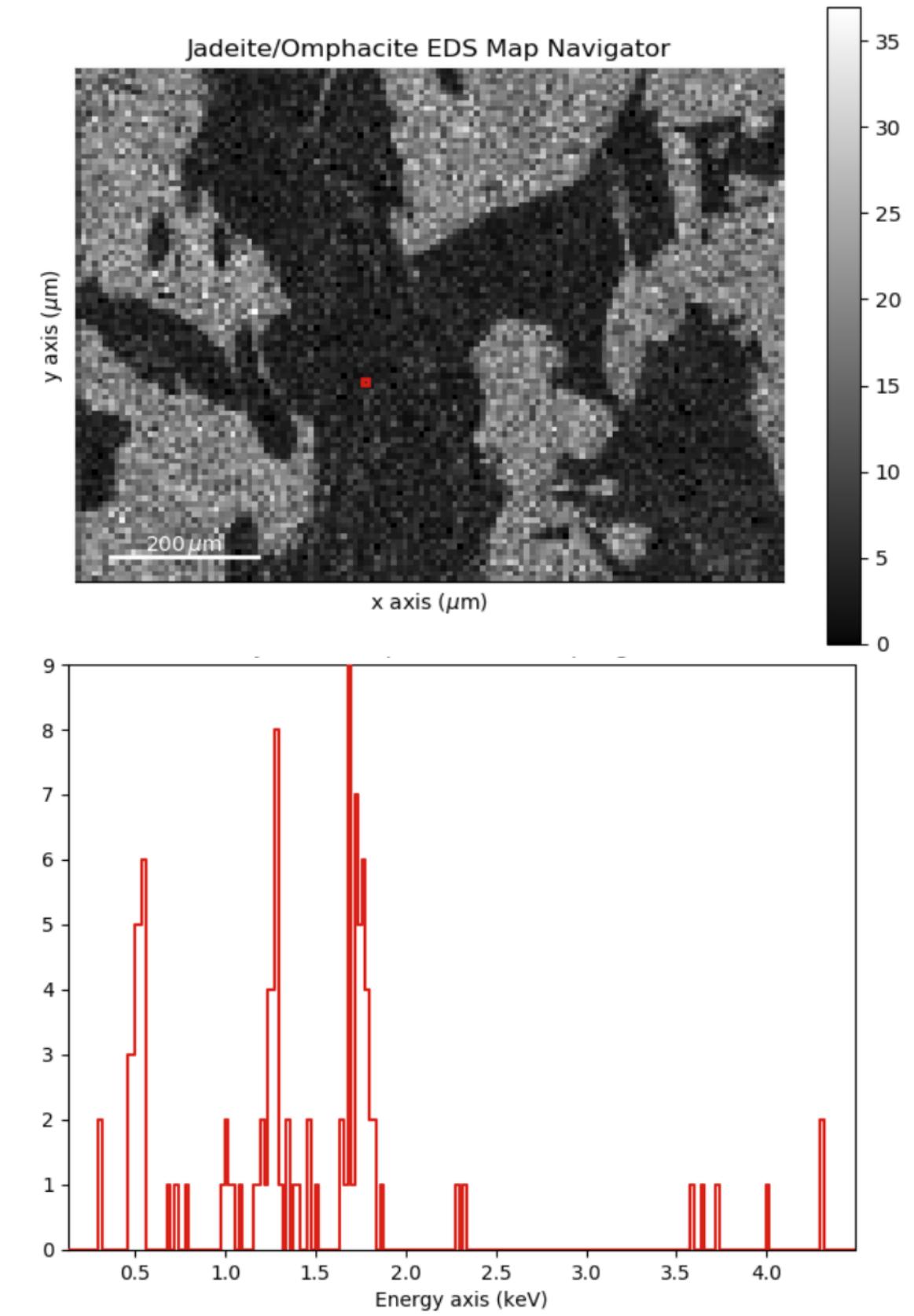
- Many algorithms exist to solve: $\mathbf{D}_{(x,y),E} = \mathbf{W}_{(x,y)} \times \mathbf{S}_E$
 - Assumptions implicit in each affect their suitability for EDS
- Primary methods:
 - Principal component analysis (PCA) – finds non-physical spectra that describe the most variance in the datacube
 - Independent component analysis (ICA) – maximizes independence between spectral results
 - Multivariate curve resolution (MCR) and non-negative matrix factorization (NMF) – enforce positivity in spectral components and weights

Offline "phase mapping"

- **Many algorithms exist to solve:** $\mathbf{D}_{(x,y),E} = \mathbf{W}_{(x,y)} \times \mathbf{S}_E$
 - Assumptions implicit in each affect their suitability for EDS
- **Other methods:**
 - Geometric methods – Vertex component analysis (VCA), Minimum volume simplex analysis (MVSA), and others...
 - Monte Carlo methods – Bayesian linear unmixing (BLU)
 - Clustering methods – k-means, Gaussian mixture modeling (GMM)

Looking at some real data...

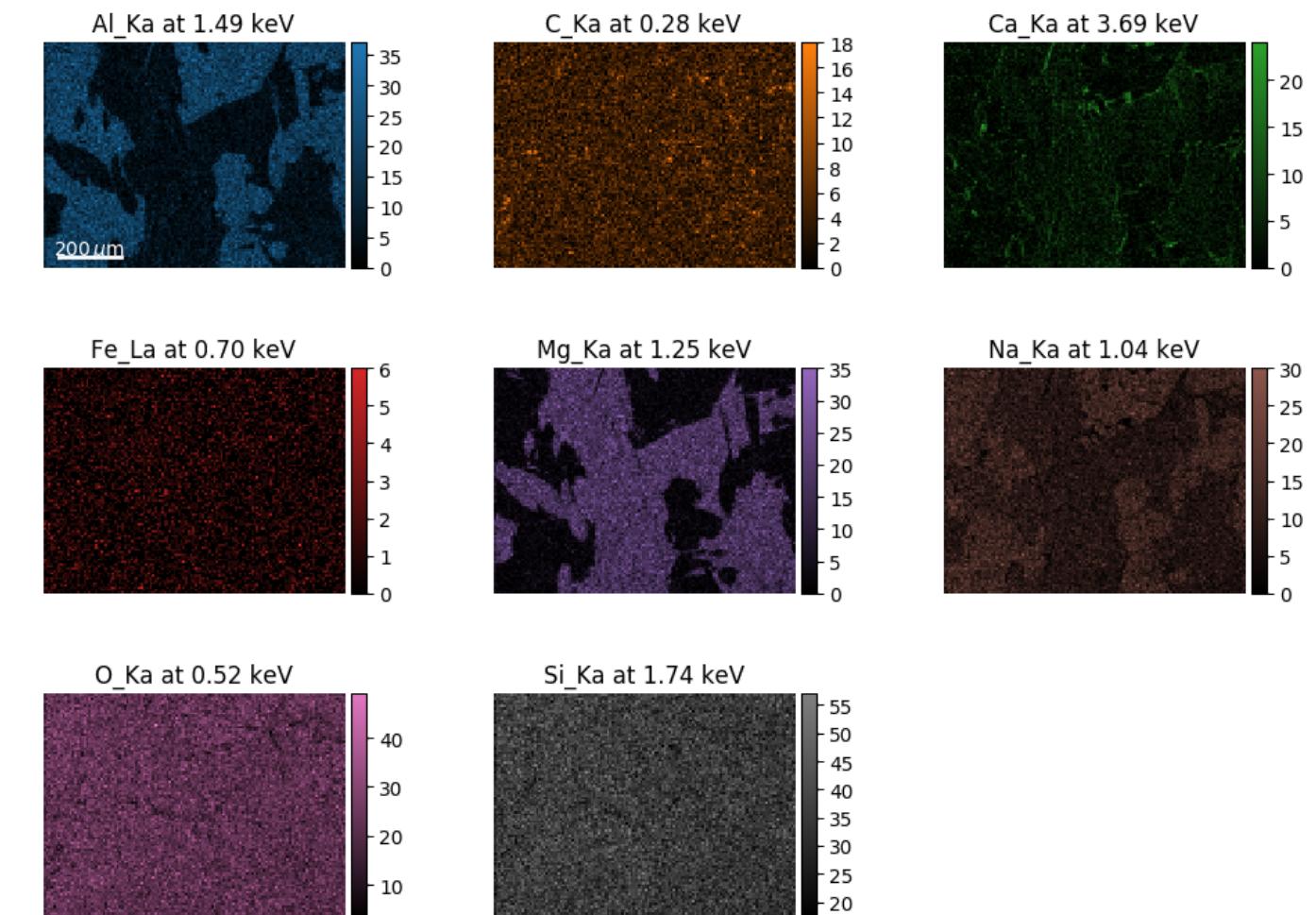
- SEM-EDS mapping data
- Japan 1957 Research Specimen from Freer Gallery of Art
 - Data from an ongoing research project at Smithsonian's Museum Conservation Institute
 - Courtesy of Thomas Lam and Edward P. Vicenzi
- Map specifics:
 - 30 keV primary beam
 - 512 x 384 pixels; 1564 spectral channels
 - Jadeite ($\text{NaAlSi}_2\text{O}_6$) and Omphacite ($(\text{Ca},\text{Na})(\text{Mg},\text{Fe}^{2+},\text{Al})\text{Si}_2\text{O}_6$)



Looking at some real data...

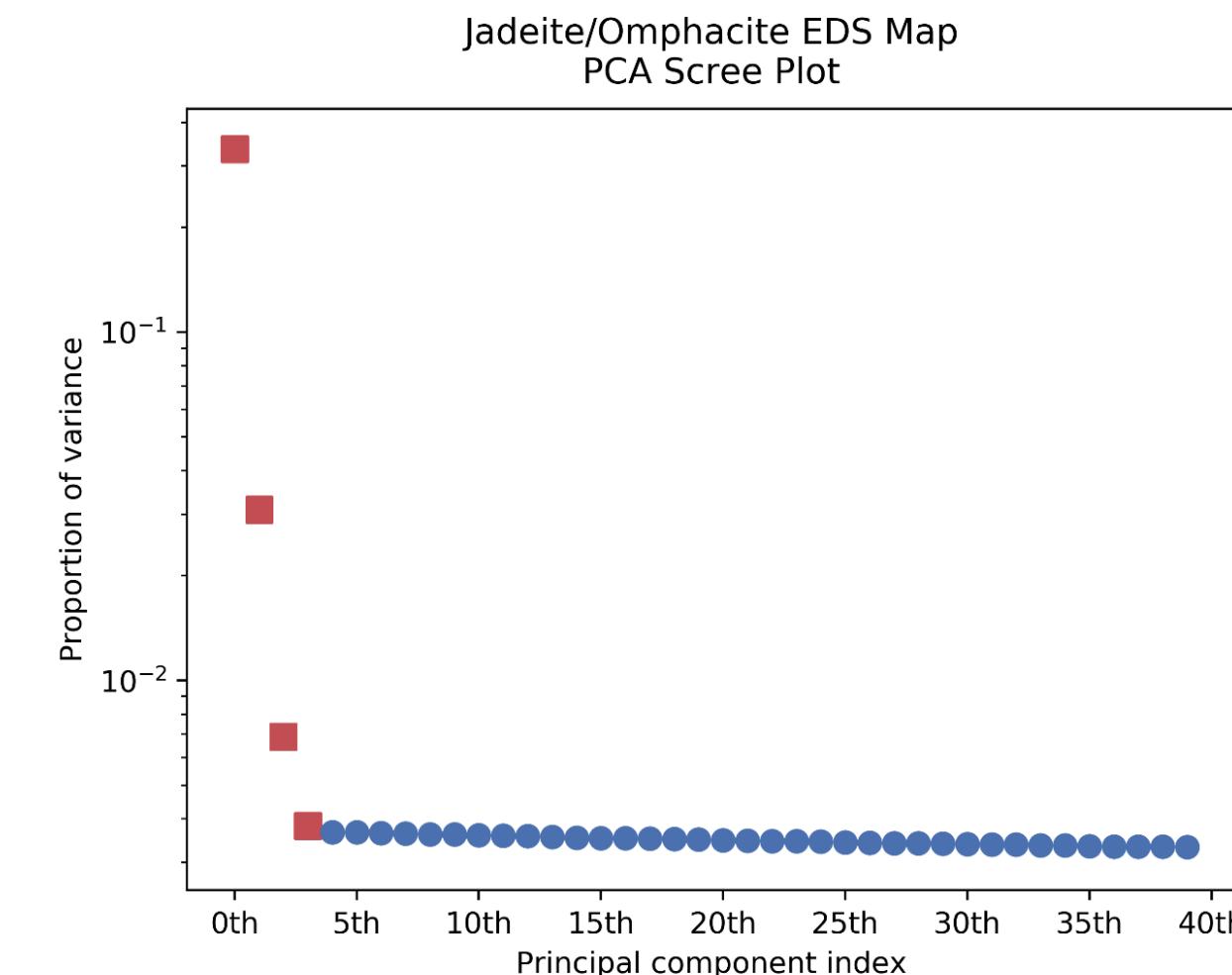
- SEM-EDS mapping data
- Japan 1957 Research Specimen from Freer Gallery of Art
 - Data from an ongoing research project at Smithsonian's Museum Conservation Institute
 - Courtesy of Thomas Lam and Edward P. Vicenzi
- Map specifics:
 - 30 keV primary beam
 - 512 x 384 pixels; 1564 spectral channels
 - Jadeite ($\text{NaAlSi}_2\text{O}_6$) and Omphacite ($(\text{Ca},\text{Na})(\text{Mg},\text{Fe}^{2+},\text{Al})\text{Si}_2\text{O}_6$)

Raw EDS line intensities



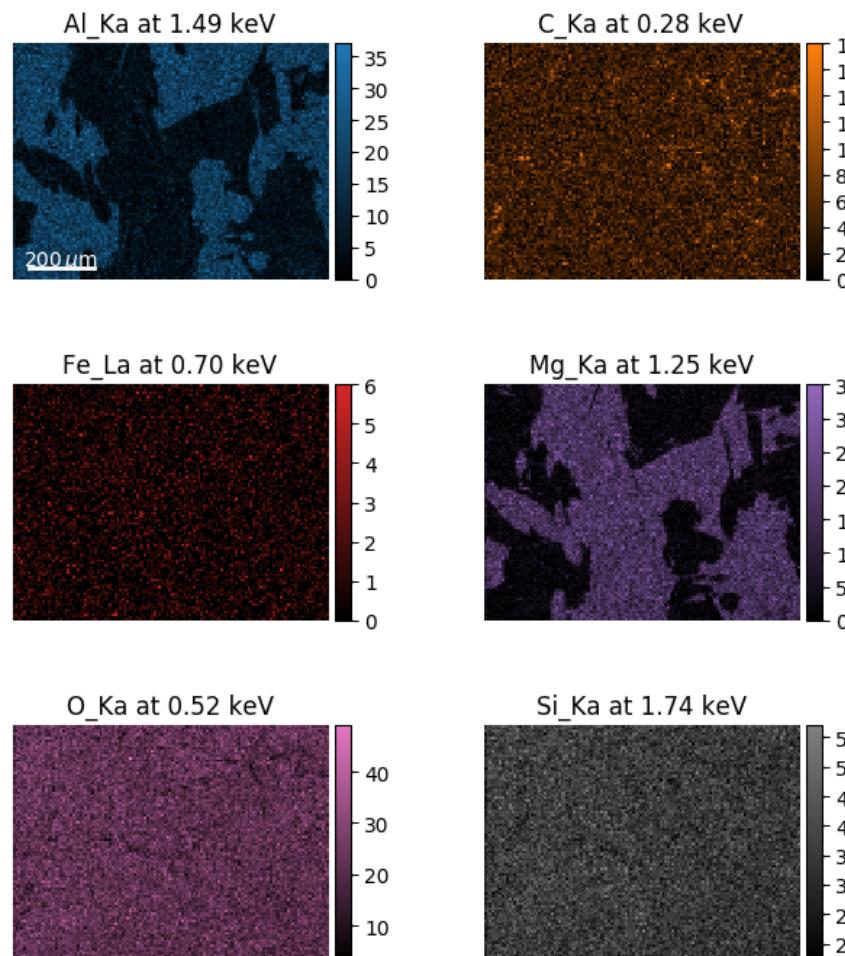
How many components are needed?

- PCA helps determine the answer with a scree plot
- Order the components by decreasing amount of contained variance on logarithmic scale
- "Correct" number of components generally at the discontinuity in the scree plot

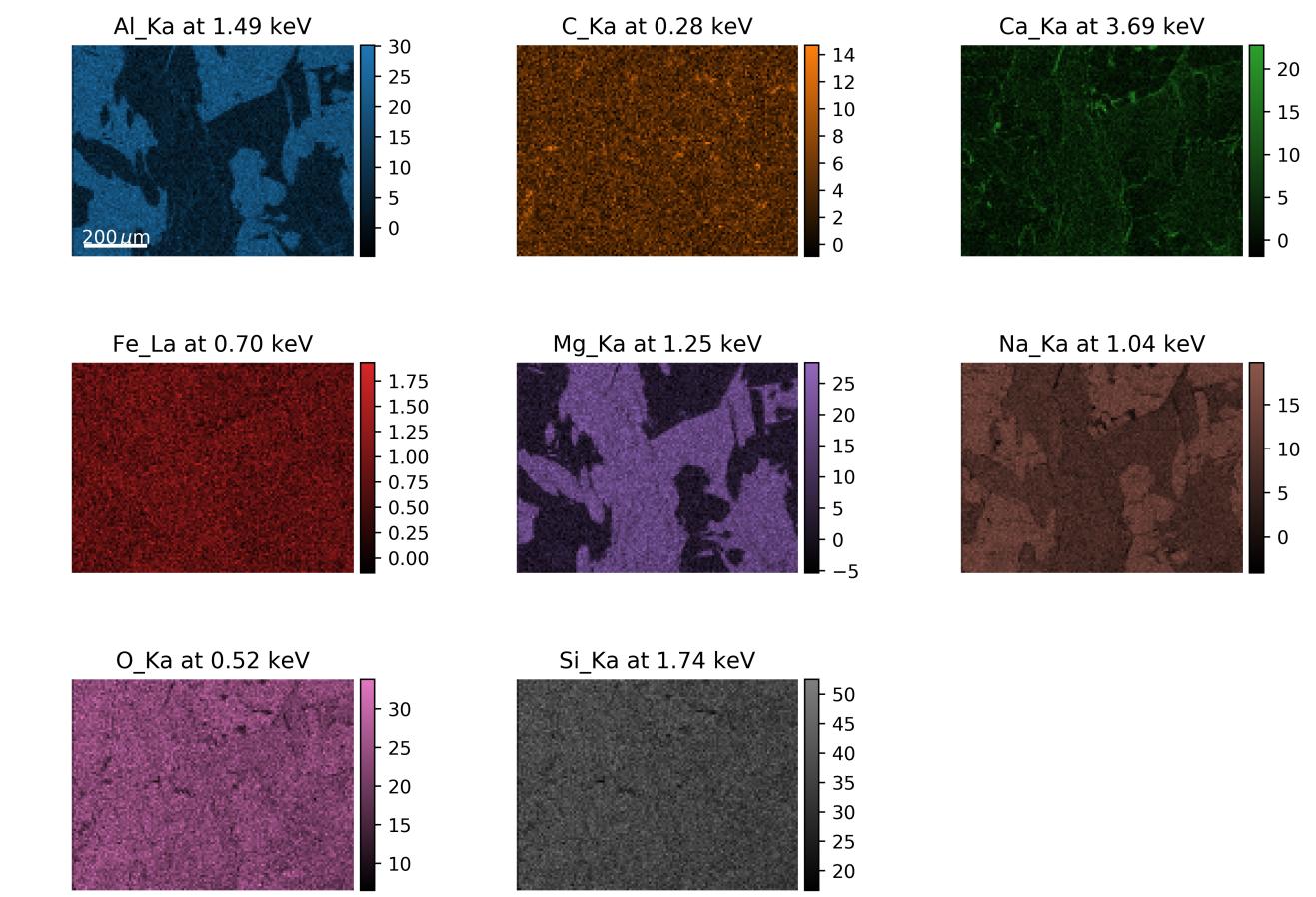


PCA Denoising ($n = 4$)

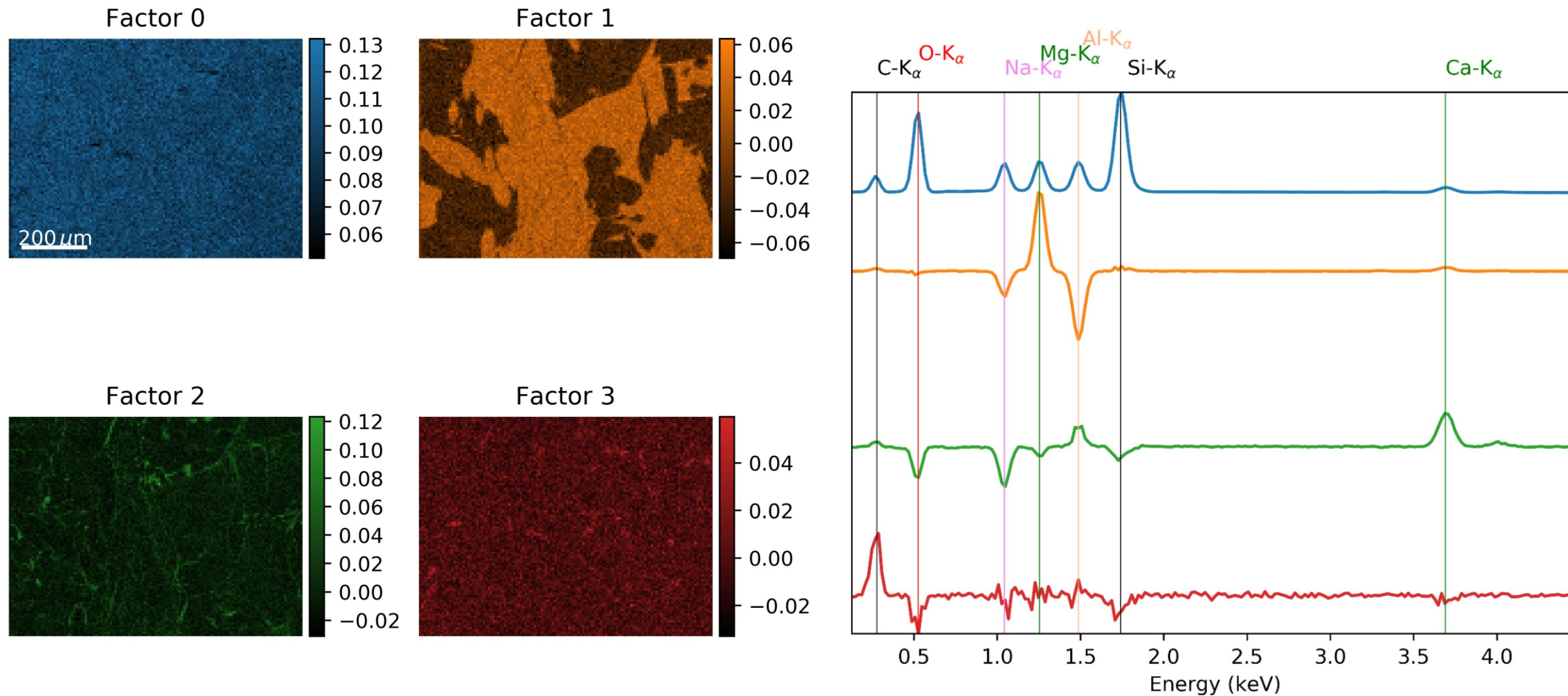
Raw EDS line intensities



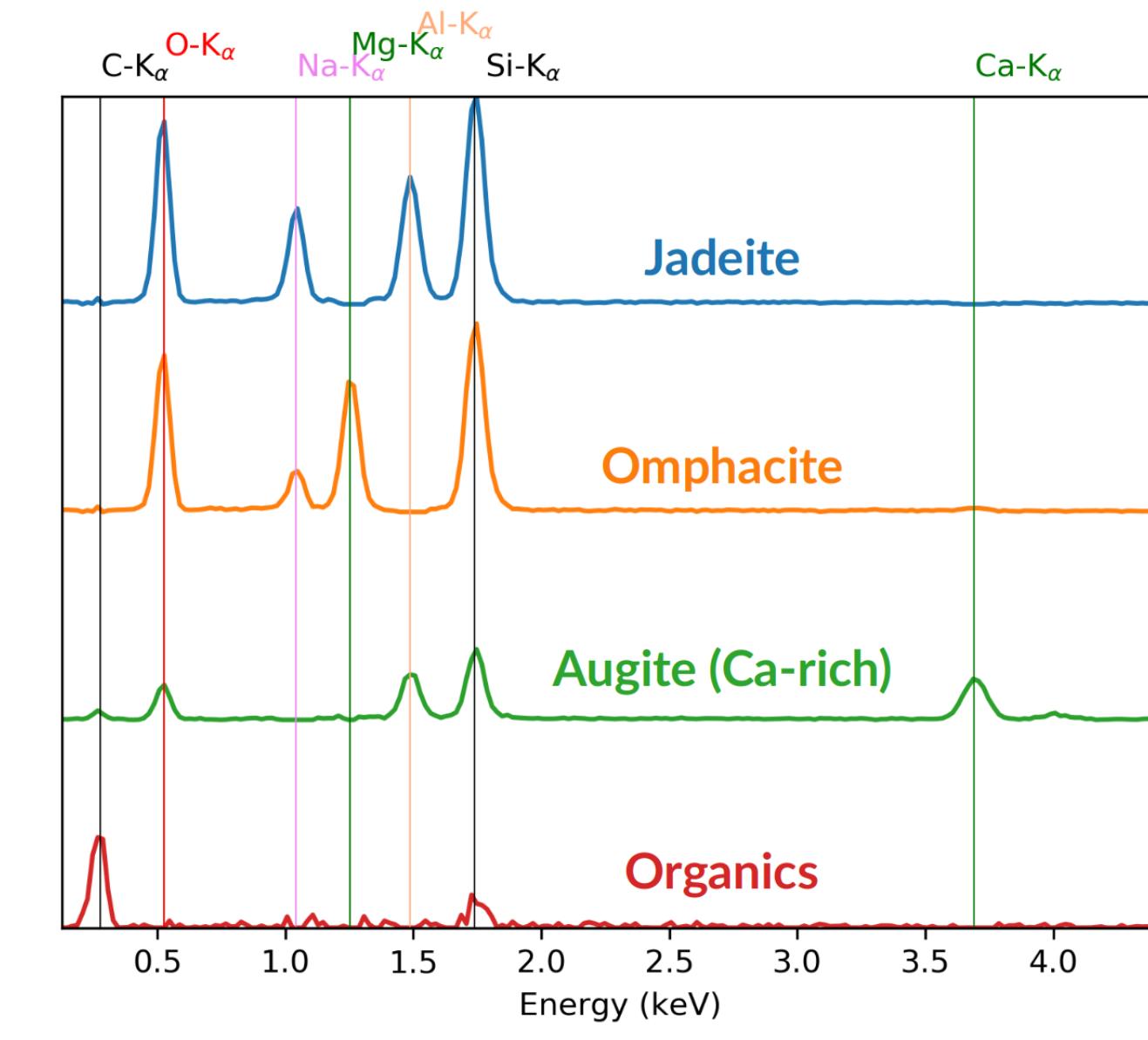
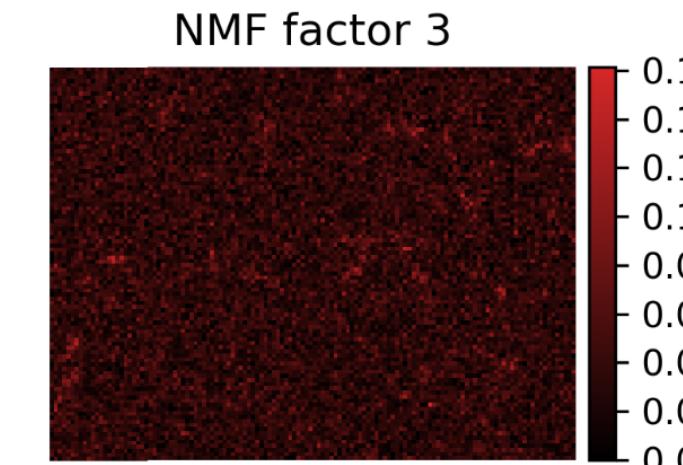
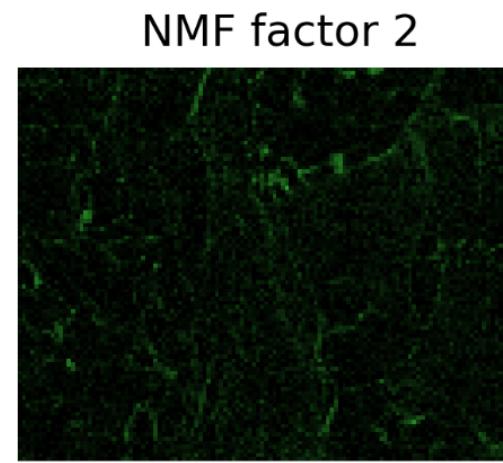
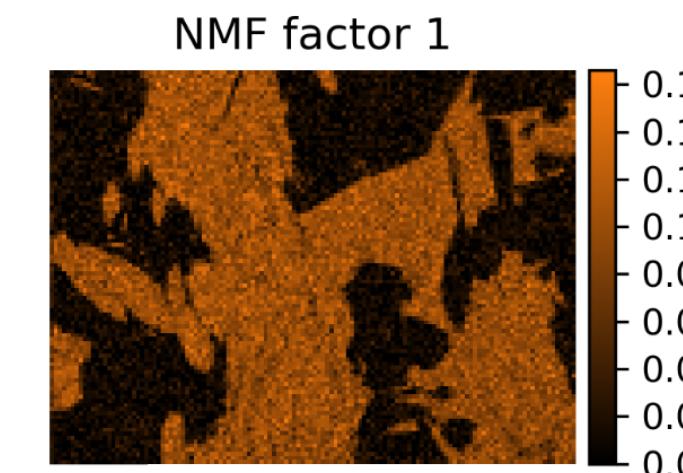
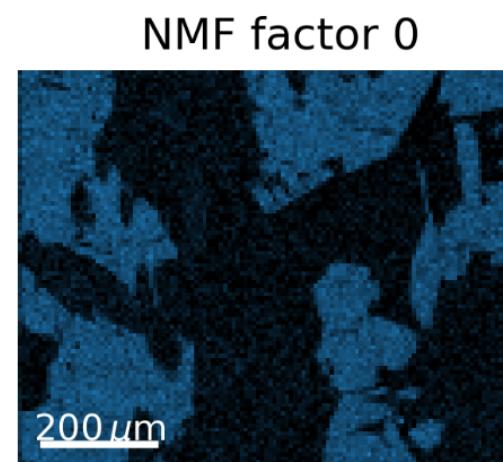
PCA-denoised EDS line intensities



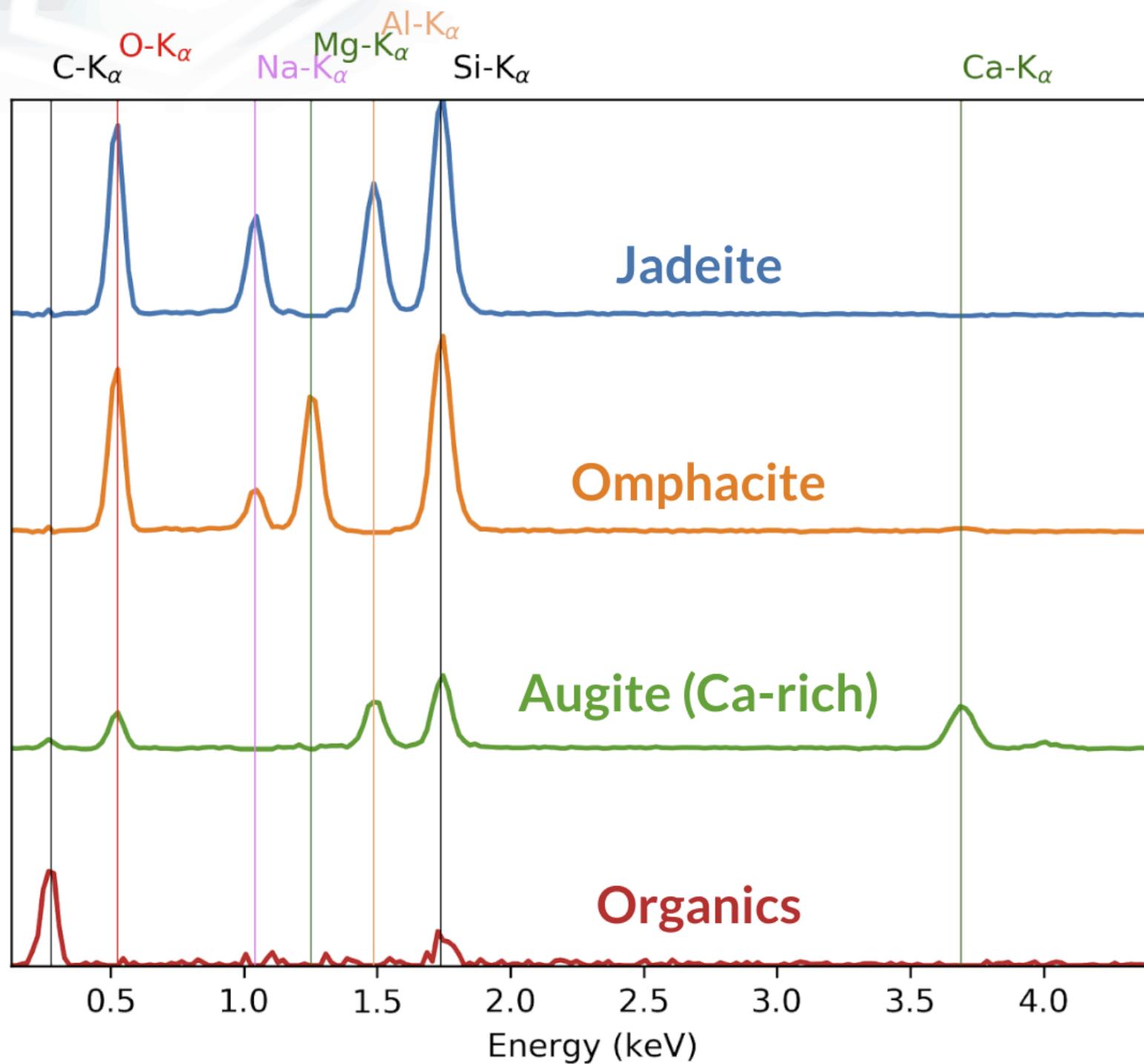
Principal component analysis ($n = 4$)



Non-negative Matrix Factorization (NMF)

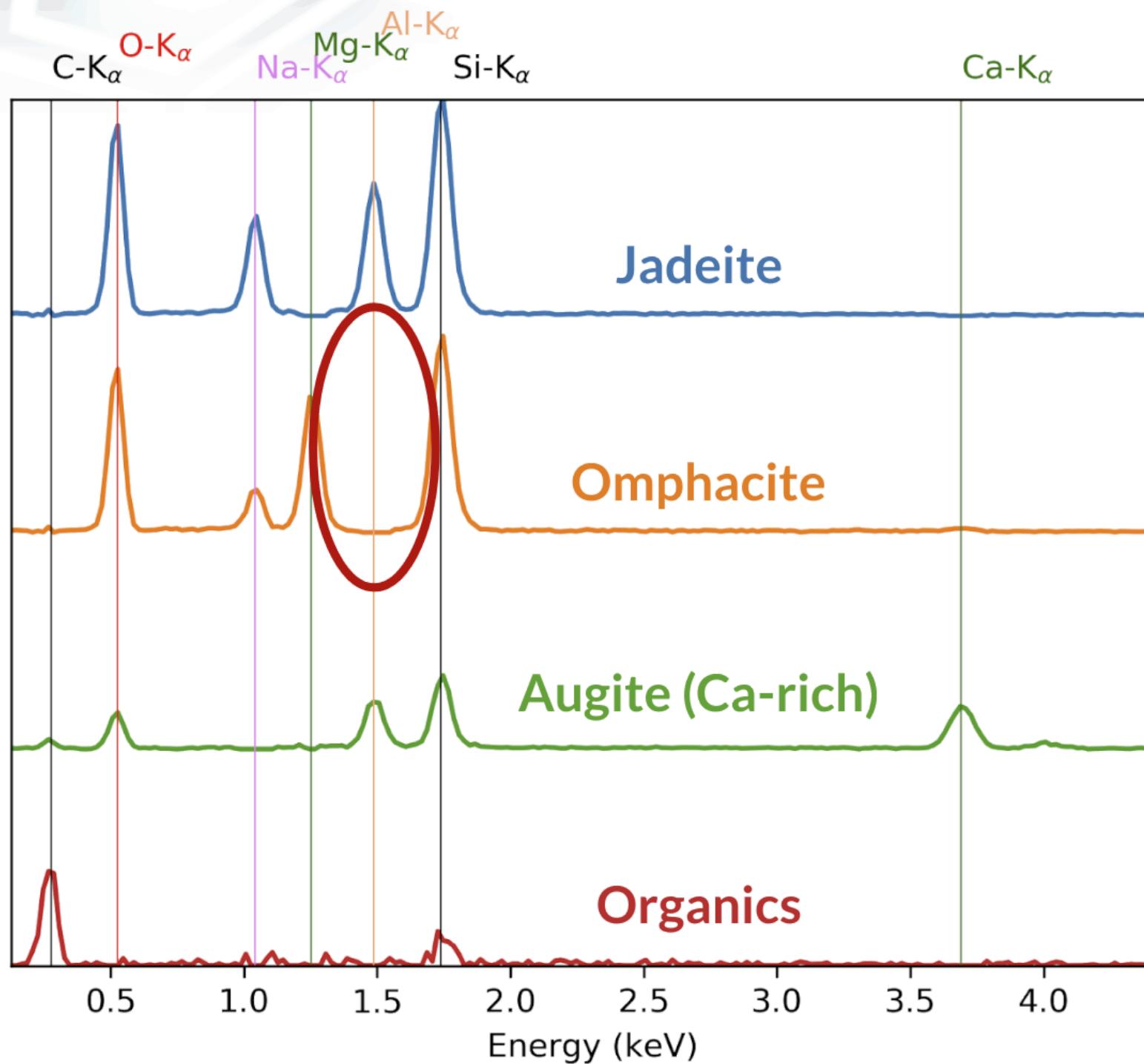


A peculiar result



Mostly carbon, some residual silicon signal...

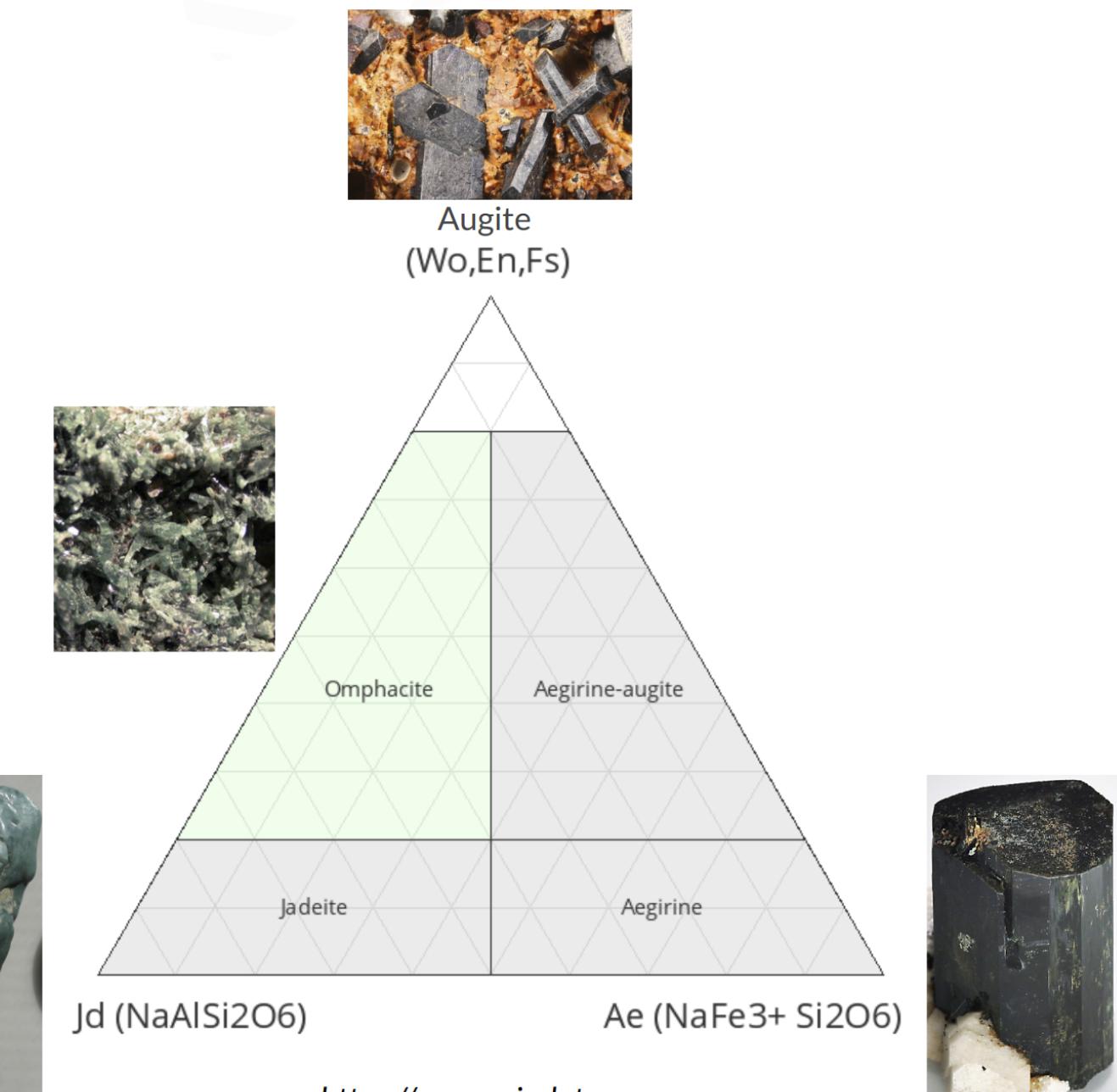
A peculiar result



Mostly carbon, some residual silicon signal...

A solid solution explains the result!

- Omphacite is solid solution of:
 - Jadeite – $\text{Na}(\text{Al},\text{Fe}^{3+})\text{Si}_2\text{O}_6$
 - Augite – $(\text{Ca}_x\text{Mg}_y\text{Fe}_z)(\text{Mg}_{y_1}\text{Fe}_{z_1})\text{Si}_2\text{O}_6$
 - Aegirine – $\text{NaFe}^{3+}\text{Si}_2\text{O}_6$



Remaining Questions

- Still not really known what this means for more precise quantification
 - Unmixing results are not always deterministic
 - Dependent on constraints, assumptions, and signal quality (noise level)
- Do count statistics, noise characteristics, etc. hold valid after signal separation?
- No (known) rigorous studies of quantification/unmixing
- Lots of opportunities for further research

Thank you!

Joshua Taillon

joshua.taillon@nist.gov