Compressed Sensing Applications in Microscopy and Microanalysis

Joshua Taillon

CS-Bio-Met meeting - May 18, 2017

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

MATERIAL
MEASUREMENT
LABORATORY

Outline

- **Introduction (about me)**
- **Brief (and basic) introduction to CS**
- **Existing applications of CS in microscopy**
- **Our work (in progress)**
- **Future ideas**

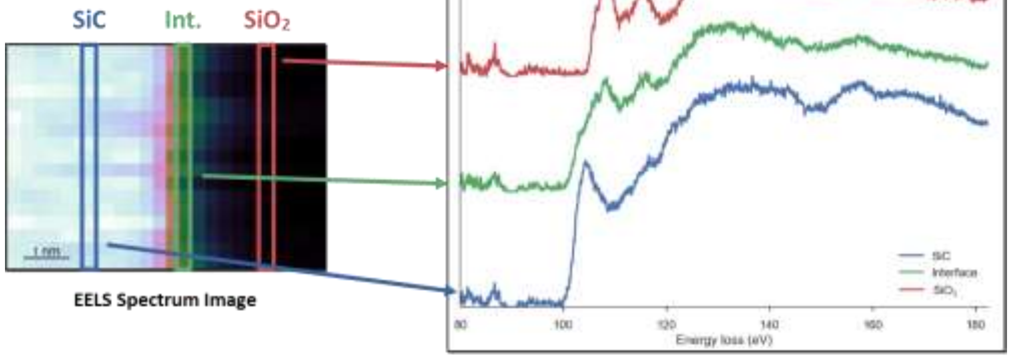
Introduction

Introduction (About Me)

- **NRC Postdoc in Materials Measurements Science Division**
  - Microscopy and Microanalysis Research Group

- **Background in Materials Characterization:**
  - TEM, FIB/SEM, EELS/EDS spectroscopies

ML Factor analysis and spectral unmixing:



EELS Spectrum Image

Si-$L_{2,3}$ ELNES signal

Introduction (About Me)

- **NRC Postdoc in Materials Measurements Science Division**
  - Microscopy and Microanalysis Research Group

- **Background in Materials Characterization:**
  - TEM, FIB/SEM, EELS/EDS spectroscopies
  - HyperSpy multi-dimensional data analysis

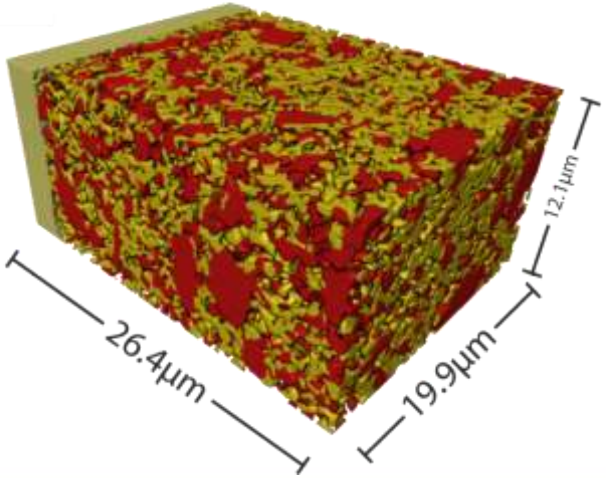
    python™

Introduction (About Me)

- **NRC Postdoc in Materials Measurements Science Division**
  - Microscopy and Microanalysis Research Group

- **Background in Materials Characterization:**
  - TEM, FIB/SEM, EELS/EDS spectroscopies

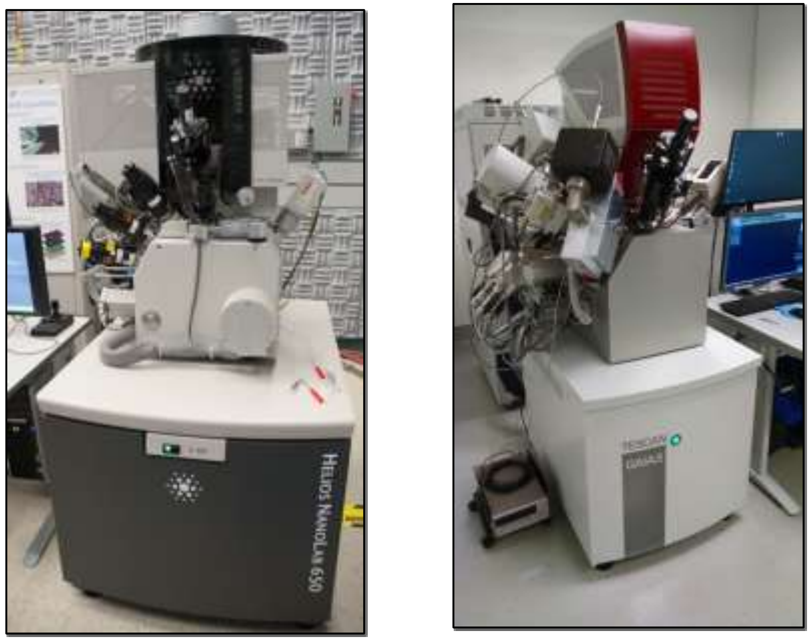
Three dimensional nanotomography and microstructure analysis:



Introduction (About Me)

- **NRC Postdoc in Materials Measurements Science Division**
  - Microscopy and Microanalysis Research Group

- **Background in Materials Characterization:**
  - TEM, FIB/SEM, EELS/EDS spectroscopies
  - ***TCL*** ***Matlab***
  - python



Research Interests

- **Lots of opportunity at intersection of microscopy and computer science/mathematics**
  - Active learning
  - Novel data analysis methods
  - Automated tool control
  - Compressed sensing

Brief intro to compressed sensing

Motivating example

- **Assume you have 7 coins**
  - One is counterfeit with a different mass than the others
  - Easy solution would be to measure the mass of each individual coin
  - Can we do better?

Good introductions in: E.J. Candès and M.B. Wakin, An Introduction To Compressive Sampling, *IEEE Signal Process. Mag*. **25**, 21 (2008).
K. Bryan and T. Leise, Making Do with Less: An Introduction to Compressed Sensing, *SIAM Rev*. **55**, 547 (2013).

Motivating example

- **Assume you have 7 coins**
  - One is counterfeit with a different mass than the others
  - Easy solution would be to measure the mass of each individual coin
  - Can we do better?

- **Measure groupings of coins instead:**
  - Only 3 measurements needed

$$\begin{array}{cc} \xrightarrow{\quad N = 7 \text{ coins} \quad} & \\ \Phi = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} & \Big\downarrow \; n = 3 \text{ measurements} \end{array}$$

Good introductions in: E.J. Candès and M.B. Wakin, An Introduction To Compressive Sampling, *IEEE Signal Process. Mag*. **25**, 21 (2008).
K. Bryan and T. Leise, Making Do with Less: An Introduction to Compressed Sensing, *SIAM Rev*. **55**, 547 (2013).

Motivating example

- **Assume you have 7 coins**
  - One is counterfeit with a different mass than the others
  - Easy solution would be to measure the mass of each individual coin
  - Can we do better?

- **Measure groupings of coins instead:**
  - Only 3 measurements needed
  - More generally,

    $n \approx \log_2(N) \ll N$

$N = 7$ coins

$$\Phi = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$n = 3$ measurements

Good introductions in: E.J. Candès and M.B. Wakin, An Introduction To Compressive Sampling, *IEEE Signal Process. Mag*. **25**, 21 (2008).
K. Bryan and T. Leise, Making Do with Less: An Introduction to Compressed Sensing, *SIAM Rev*. **55**, 547 (2013).

More general CS formalism

- **$\boldsymbol{\Phi}$ is $\boldsymbol{n} \times \boldsymbol{N}$ “sensing matrix”**
  - We are trying to recover an unknown *sparse* vector $\mathbf{x} \in \mathbb{R}^{\boldsymbol{N}}$ with a measurement vector $\mathbf{b}$ and a known sensing matrix $\boldsymbol{\Phi}$
  - $\mathbf{x}$ is sparse (mostly zeros), and describes which coin is different
  - We want to find a sparse solution that satisfies $\boldsymbol{\Phi}\mathbf{x} = \mathbf{b}$, given:

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \end{bmatrix}$$

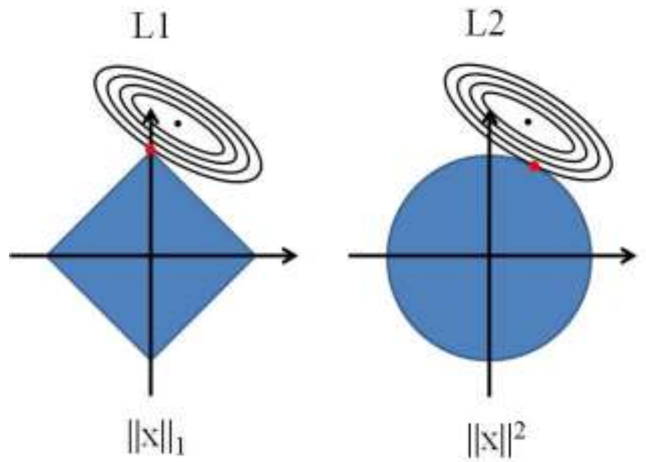
More general CS formalism

- **How do we determine $\mathbf{\Phi}$?**
  - Strictly speaking, it must obey the “restricted isometry principle” (RIP)
    - RIP means that if we select K random columns from $\mathbf{\Phi}$, that submatrix is full rank
    - Designing a $\mathbf{\Phi}$ to satisfy the RIP is actually NP-hard, but it turns out it’s not *necessary* for a successful recovery of $\mathbf{x}$, just *sufficient*

  - Turns out that a random binary matrix will satisfy RIP (with high probability)
    - In practically all implementations of CS, this random sampling is what is used

More general CS formalism

- **$\mathbf{\Phi}$ has fewer rows than columns, so $\mathbf{\Phi x} = \mathbf{b}$ is underdetermined**
  - This means there are an infinite number of solutions
- **The "magic" of compressed sensing is the use of the ℓ_1 norm for convex optimization to find the "best" $\mathbf{x}$ (when using the right $\mathbf{\Phi}$)**

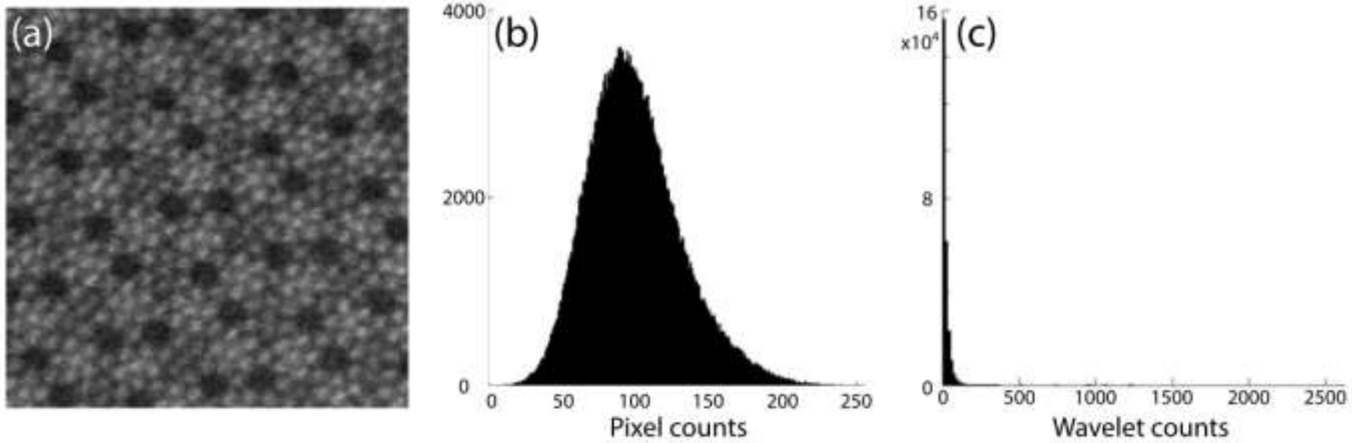


T. Zhou. "Compressed Sensing Review (1): Reconstruction Algorithms." *Tianyi Zhou's Research Blog*, (2010).

	Minimizing Norm	**Definition**	**Description**
NP-hard! →	$\boldsymbol{\ell_0}$	$\|x\|_0 = \sqrt[0]{\Sigma_i x_i^0} = \#(i\|x_i \neq 0)$	Number of non-zero components
	$\boldsymbol{\ell_1}$	$\|x\|_1 = \Sigma_i \|x_i\|$	*Manhattan* distance (Sum of Absolute Difference)
Bad results →	$\boldsymbol{\ell_2}$	$\|x\|_2 = \sqrt{\Sigma_i x_i^2}$	Euclidean distance (Least squares difference)

Why “compressed” sensing?

- **Consider a “compressible” HRTEM image:**
  - In the pixel basis, many coefficients needed to encode the image
  - In the wavelet basis, very few coefficients are needed
  - Idea behind JPEG, etc.



Adapted from P. Binev *et al.*, “Compressed Sensing and Electron Microscopy” in *Modeling Nanoscale Imaging Electron Microscopy* (2012).

- **What if we could measure the sparse basis directly?**

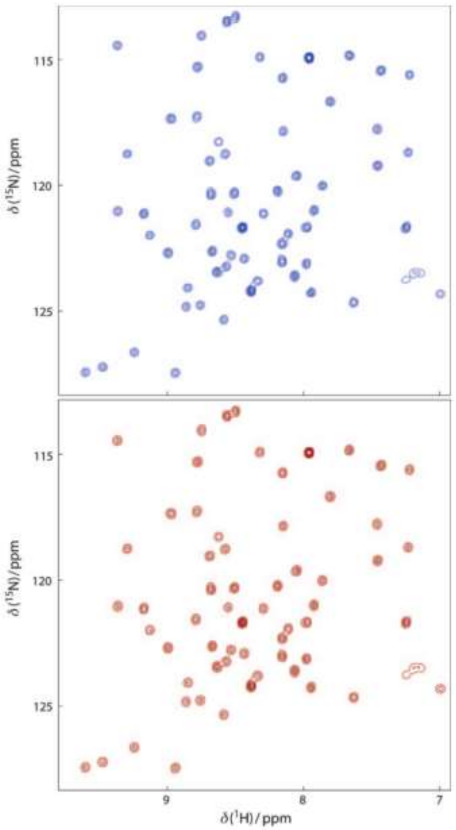
Areas of initial success - MRI

Comics from Michael Lustig

- **Subsampling *k*-space greatly reduces number of acquisitions needed, and time for patient**
  - Greatly expands indications for MRI (pediatrics, compromised health, etc.)
  - After 10 years, first FDA-approved applications of CS-MRI on the market in Feb. 2017 (link)

- **Compressed sensing saves lives!**
  - J. Ellenberg, "Fill in the Blanks: Using Math to Turn Lo-Res Datasets Into Hi-Res Samples," *Wired* (2010).

Areas of initial success – NMR Spectroscopy



30% sampling

Full sampling

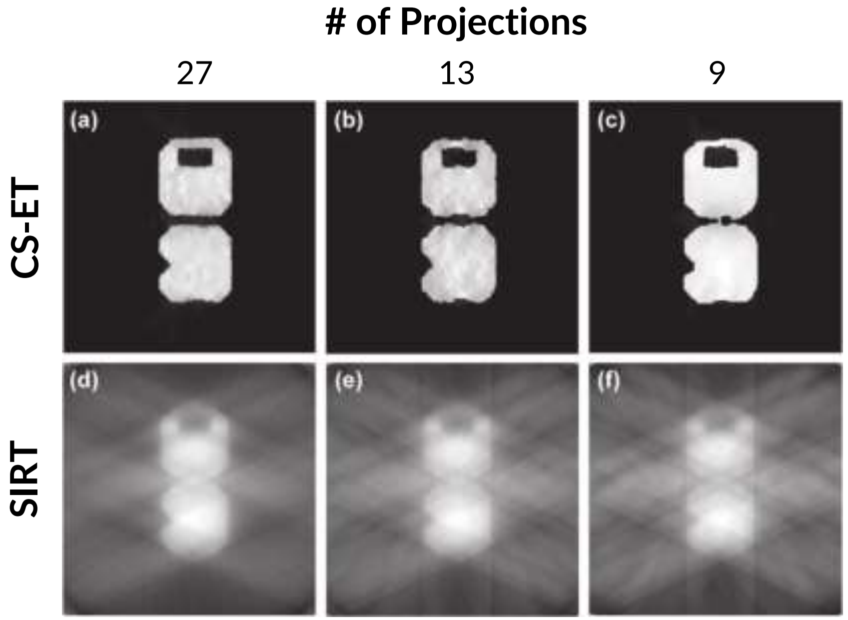
- **Multidimensional nuclear magnetic resonance**
  - NMR spectra are typically sparse – few isolated peaks in the Fourier domain
  - Measured 30% of the 128 complex data pairs
  - Measurement time: 165 min → 50 min
- **Observed very high fidelity reconstructions w.r.t. peak positions and shape/intensity**

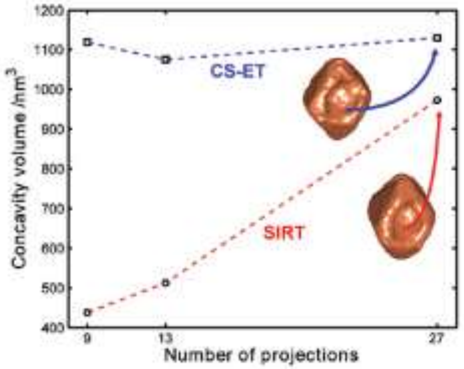
D.J. Holland, *et al.*, Fast Multidimensional NMR Spectroscopy Using Compressed Sensing, *Angew. Chemie Int. Ed.* **50**, 6548 (2011).

Efforts in CS for electron microscopy

CS in TEM (electron tomography – CS-ET)

- **Tilt-tomography finds 3D representations of objects in the TEM by acquiring 2D images at many tilt angles**
  - Can reduce number of tilted images needed using CS principles and compare results to the standard *simultaneous iterative reconstruction technique* (SIRT)
- **Tomography of iron oxide nanoparticles:**
  - CS-ET performs significantly better than SIRT, at all signal levels



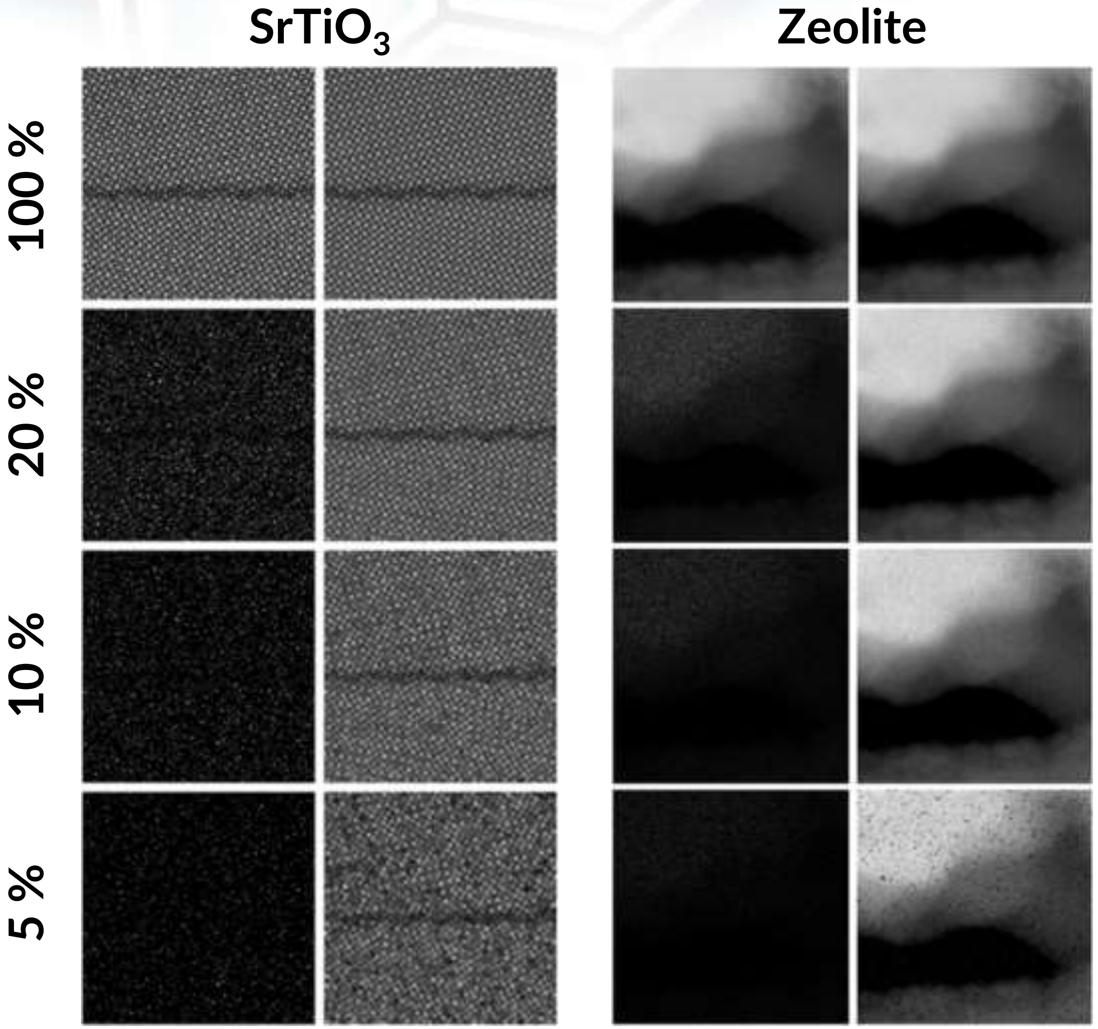


Z. Saghi, *et al.* Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. A compressed sensing-electron tomography (CS-ET) approach, *Nano Lett*. **11**, 4666 (2011).

R.K. Leary, *et al.*, Compressed sensing electron tomography, *Ultramicroscopy*. **131**, 70 (2013).

CS in STEM (imaging)

- **Benefits of reduced sampling in STEM:**
  - Reduced dose (for electron-sensitive materials)
  - Reduced time/increased throughput
- **Random sampling in pixel-domain**
  - Bayesian factor analysis to find sparse representation (BPFA)
  - Sampling done with beam blanker or meandering beam



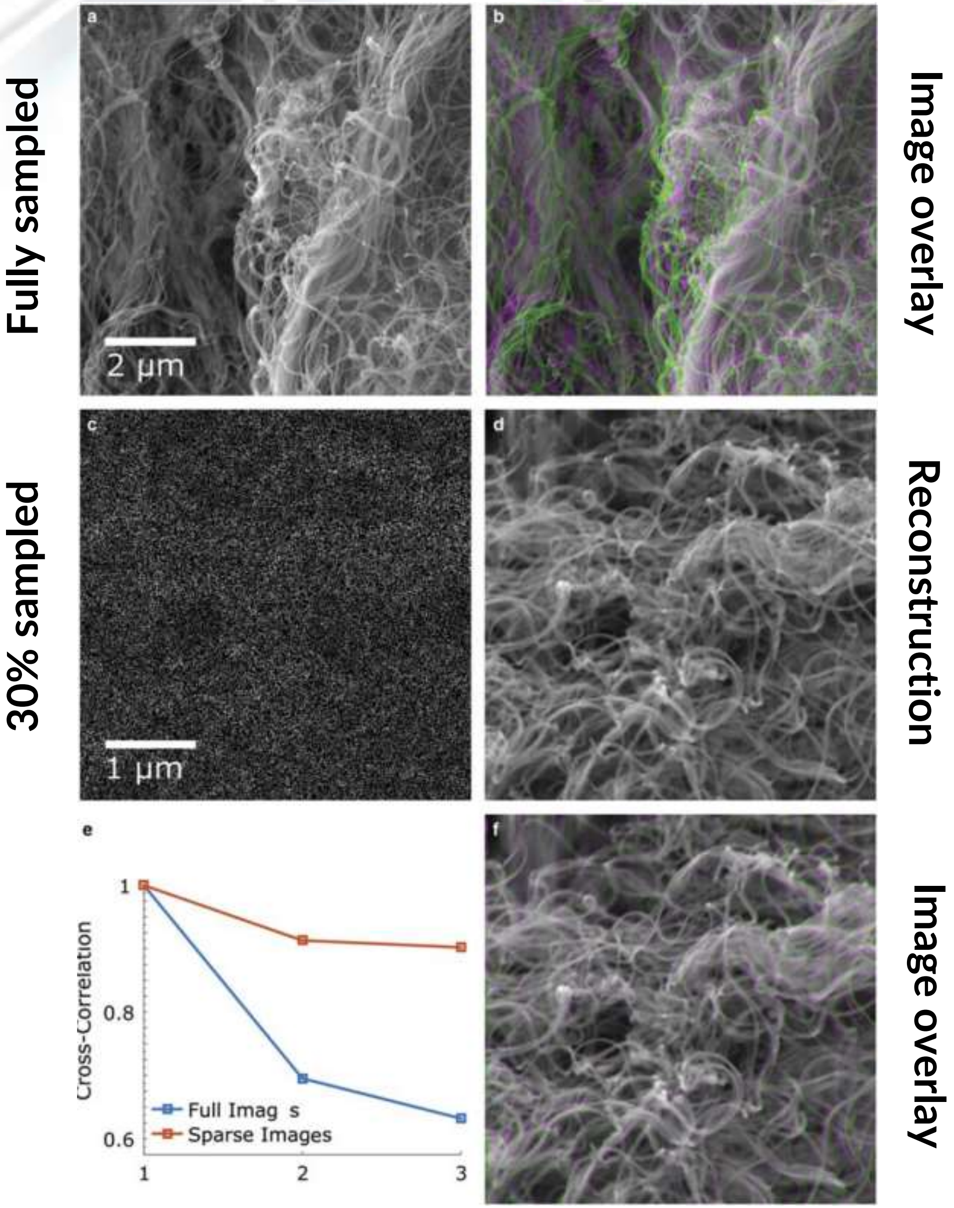
A. Stevens, *et al.*, The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images, Microscopy. 63, 41 (2014).

CS in SEM

- **Two published implementations:**
  - H.S. Anderson, *et al.*, Sparse imaging for fast electron microscopy, Proc. SPIE-IS&T Electron. Imaging. 8657 (2013).
    - Use a split Bregman formulation of basis-pursuit for ℓ_1 minimization
    - Drive beam to random pixels, taking care to account for scan coil dynamics
    - 10x imaging speedup when imaging only 10% of pixel locations (linear response)
  - K. Hujsak, *et al.*, Suppressing Electron Exposure Artifacts: An Electron Scanning Paradigm with Bayesian Machine Learning, Microscopy and Microanalysis, 1–11 (2016).
    - Use the same BPFA as in Stevens' CS-STEM
    - No scan coil modifications to SEM (just need high-speed beam blanker)
    - Did not demonstrate significant speed gains, but did significantly reduce dose

CS in SEM

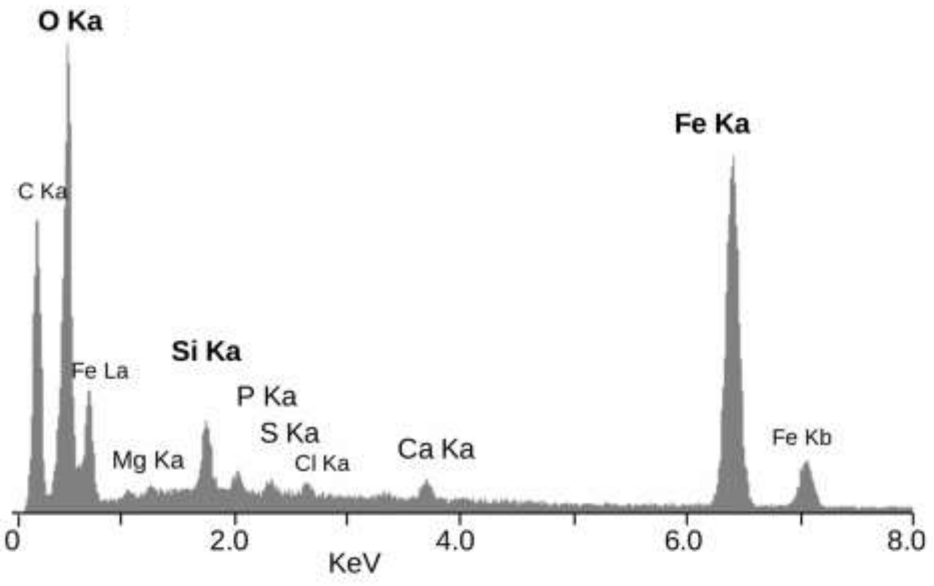
- K. Hujsak, *et al*., Suppressing Electron Exposure Artifacts: An Electron Scanning Paradigm with Bayesian Machine Learning, Microscopy and Microanalysis, 1–11 (2016).
  - Analyzed effect of reduced dose from CS-SEM on electron-sensitive human collagen sample
  - Significant reduction in sample modification due to beam-damage
- Also tested various scan patterns
  - Random sampling was found to perform better than spiral, Lissajous, and random line sampling
  - Highest reconstructed PSNR and least number of scanning artifacts



Our initial work

Our goals

- **Originally set out to implement what was published by Hujsak**
  - Have pivoted instead to see what impact we can have in the analytical realm
  - One conference paper on CS in STEM-EELS, but nothing else (yet)
  - Given our preference for FIB/SEM, we'll focus on X-ray analysis (EDS)
    - Long dwell times needed
    - Acquisitions will damage beam-sensitive materials
    - Seems ripe for "disruption"



Strategy

- **Identify existing reconstruction algorithm that is purpose-built for hyperspectral imaging (HSI) data**
  - The hyperspectral remote sensing community has been much faster to adopt advanced data analysis methods (like CS) than the microscopy community
- **Demonstrate control and implement on existing HSI data**
- **Apply to simulated EDS data to determine/quantify effectiveness**
- **Determine what (if any) performance enhancement can be gained on live experiments**

BPFA = Beta-Bernoulli Process Factor Analysis

- **Factor analysis**
  - Decompose signal into a linear approximation of factors and weights

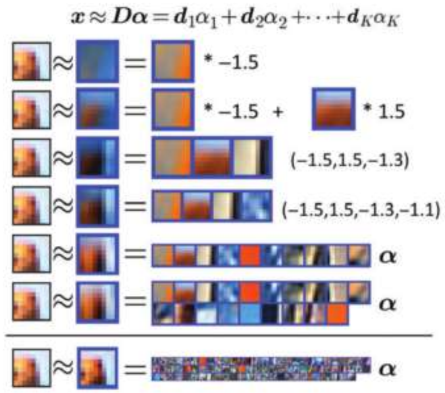
$$\boldsymbol{x}_i = \mathbf{D}\,\boldsymbol{w}_i + \boldsymbol{\epsilon}_i$$

Noise and residual

Original data samples

Dictionary matrix of “prototypical” signals

Dictionary element weights (vector)



An image patch ($\boldsymbol{x}_i$) can be represented by a dictionary ($\mathbf{D}$) of representative patches, with each element weighted by a factor from $\boldsymbol{w}$ (plus some noise $\boldsymbol{\epsilon}$)

Figure from: A. Stevens, et al. *Microscopy*. **63**, 41 (2014).
Note: $w_i = \alpha_i$ *in this paper’s notation*

BPFA = Beta-Bernoulli Process Factor Analysis

- **How to find $\mathbf{D}$ and $\boldsymbol{w}$?**
  - Bayesian beta-Bernoulli Process

- **We infer the underlying signal $\boldsymbol{x}_i = \mathbf{D}\boldsymbol{w}_i$ by:**
  - Placing Bayesian priors on $\mathbf{D}$, $\boldsymbol{w}_i$, and $\boldsymbol{\epsilon}_i$
  - Assuming that $\boldsymbol{w}_i$ is sparse
  - Iterate on each parameter to improve the estimation of their values (based on observed data)

Details about algorithms in:

- M. Zhou, et al., Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images, *IEEE Trans. Image Process*. **21**, 130–144 (2012)
- Z. Xing, et al., Dictionary Learning for Noisy and Incomplete Hyperspectral Images, *SIAM J. Imaging Sci*. **5**, 33–56 (2012).

BPFA Formalism

- **What does beta-Bernoulli Process mean in practice?**
    - BP is a strategy for updating $\boldsymbol{w}_i$ in a Bayesian manner

- **Define $\boldsymbol{w}_i$ as $\boldsymbol{z}_i \odot \boldsymbol{s}_i$:**
    - $\odot$ = element-wise multiplication
    - $\boldsymbol{s}_i$ are the dictionary weights
    - $\boldsymbol{z}_i$ are “binary indicators”, specifying which of the K columns of $\boldsymbol{D}$ are used to represent $\boldsymbol{x}_i$:
        - $\boldsymbol{z}_i$ are drawn from a Bernoulli distribution:

$$\boldsymbol{z}_i \sim \prod_{k=1}^{K} \text{Bernoulli}\,(\pi_k)$$

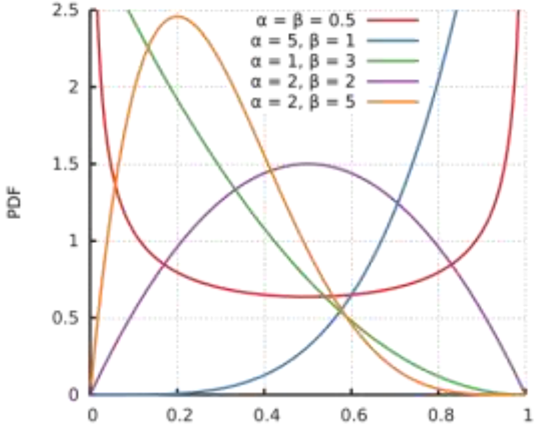
        - Where π_k is the kth component of:

$$\boldsymbol{\pi} \sim \prod_{k=1}^{K} \text{Beta}\left(\frac{a}{K}, \frac{b(K-1)}{K}\right)$$

Bernoulli Distribution

$$f(k;p)\begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0. \end{cases}$$

Beta Distribution



BPFA Formalism

- **Full model parameters:**
    - $\boldsymbol{x}_i = \mathbf{D}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i$
    - $\boldsymbol{w}_i = \boldsymbol{z}_i \odot \boldsymbol{s}_i$
    - $\boldsymbol{d}_k \sim \mathcal{N}(0, P^{-1}\mathbf{I}_{\mathrm{P}})$
    - $\boldsymbol{s}_{\boldsymbol{i}} \sim \mathcal{N}(0, \gamma_s^{-1}\mathbf{I}_K)$
    - $\boldsymbol{\epsilon}_{\boldsymbol{i}} \sim \mathcal{N}(0, \gamma_\epsilon^{-1}\ \mathbf{I}_P)$
    - $\boldsymbol{z}_i \sim \prod_{k=1}^{K} \mathrm{Bernoulli}\ (\pi_k)$
    - $\boldsymbol{\pi} \sim \prod_{k=1}^{K} \mathrm{Beta}\left(\frac{a}{K}, \frac{b(K-1)}{K}\right)$

- More definitions:
    - $\sim$ means a Bayesian variable drawn from the specified prior distribution
    - $\mathcal{N}(i, j)$ specifies a normal distribution with mean i and variance j
    - γ_i are conjugate hyperpriors of the form $\gamma_s \sim \mathrm{Gamma}(c, d)$ and $\gamma_\epsilon \sim \mathrm{Gamma}(e, f)$
    - $\boldsymbol{d}_k$ represents the kth component (column) of $\mathbf{D}$

How does it work?

- **The model is inferred using Gibbs Sampling**
  - Markov chain Monte Carlo algorithm used to sample the full Bayesian likelihood
  - Samples from the posterior distribution of each random variable are estimated by iteratively sampling from the conditional distributions of each variable (given all the others)
  - After "enough" iterations, the estimated variables are likely close to the true model parameters
- **Such a method determines both the dictionary and weights iteratively**
  - Can be used on plain imagery (like Anderson and Hujsak)
  - Also developed for hyperspectral imagery (by Xing)

What does it look like in practice?

- **Example data from algorithm authors:**
    - Hyperspectral imagery (HSI) from satellite imaging of an urban environment
    - 150 x 150 spatial pixels – 210 spectral bands
    - Artificially remove some large fraction of voxels, and reconstruct using BPFA:

What does it look like in practice?

Original data

2% sampling

2x2 patch size

4x4 patch size

Spectral Channel 100

PSNR = 8.77 dB

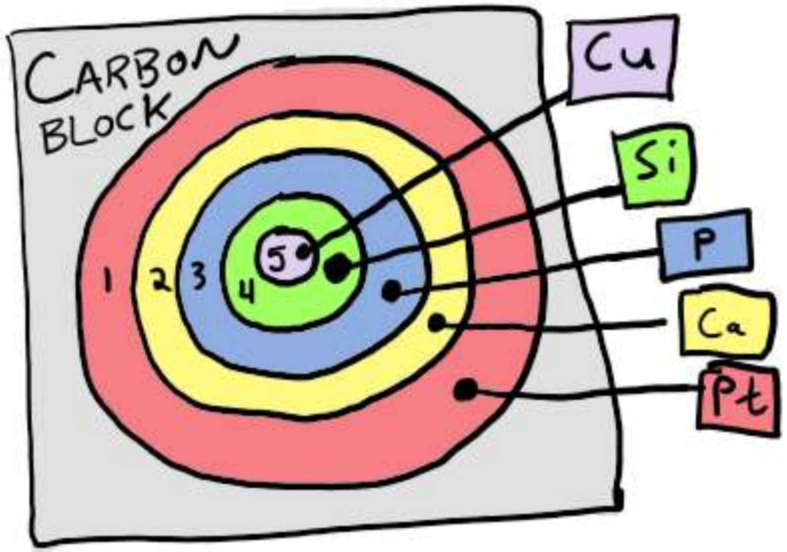
PSNR = 21.36 dB

PSNR = 24.00 dB

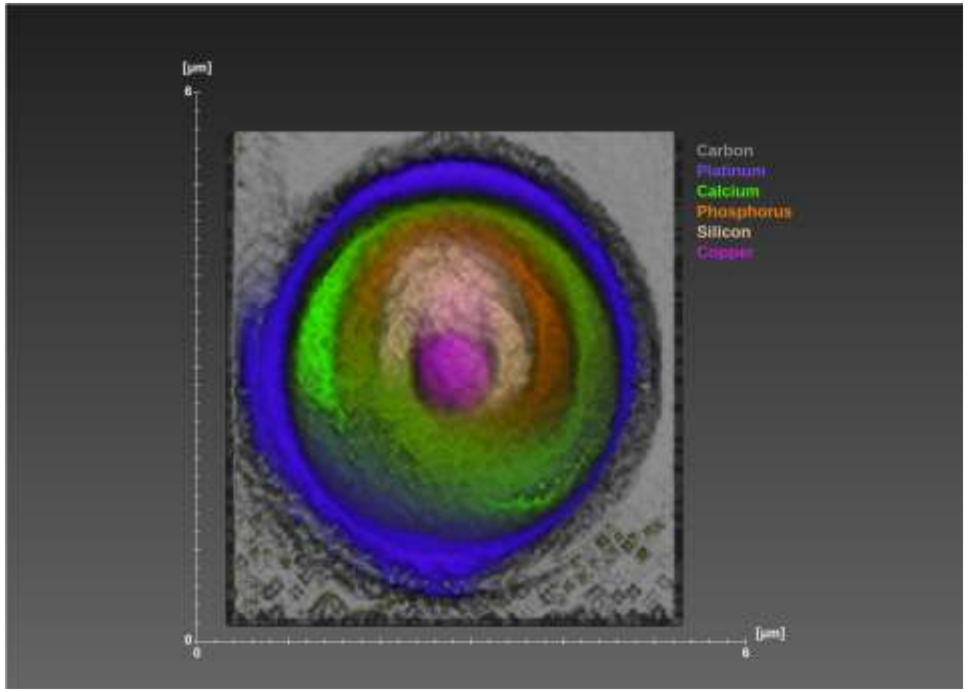
Center pixel spectrum

How about EDS?

- **Simulated 3D data cube using DTSA-II:**
  - 100 x 100 spatial – 2048 spectral channels



Schematic of spherical geometry

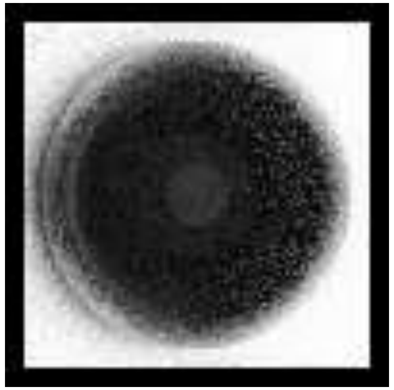


3D visualization of quantified block

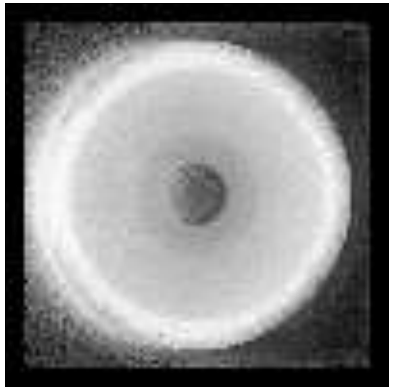
How about EDS?

- **Simulated 3D data cube using DTSA-II:**
  - 100 x 100 map – 2048 spectral channels

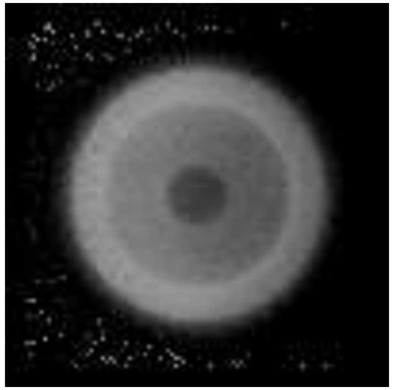
- **Fully sampled quantification data:**



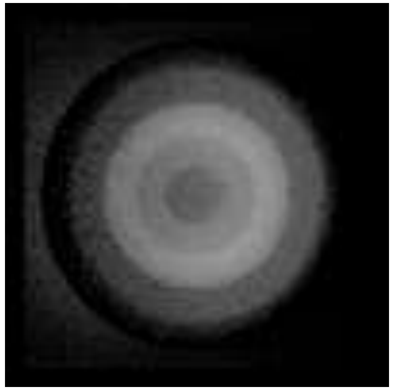
Carbon



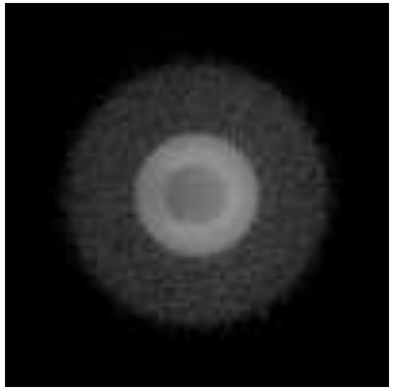
Platinum



Calcium



Phosphorus

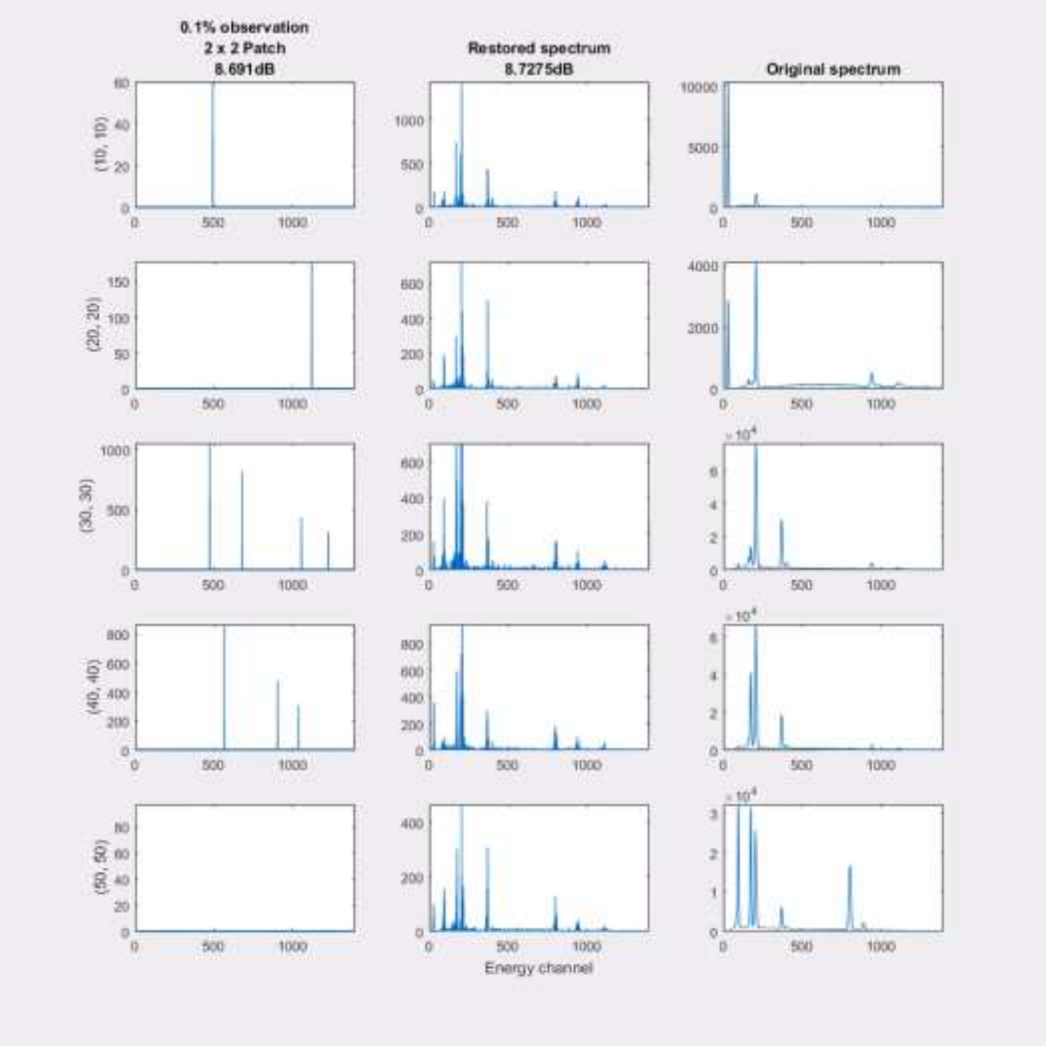


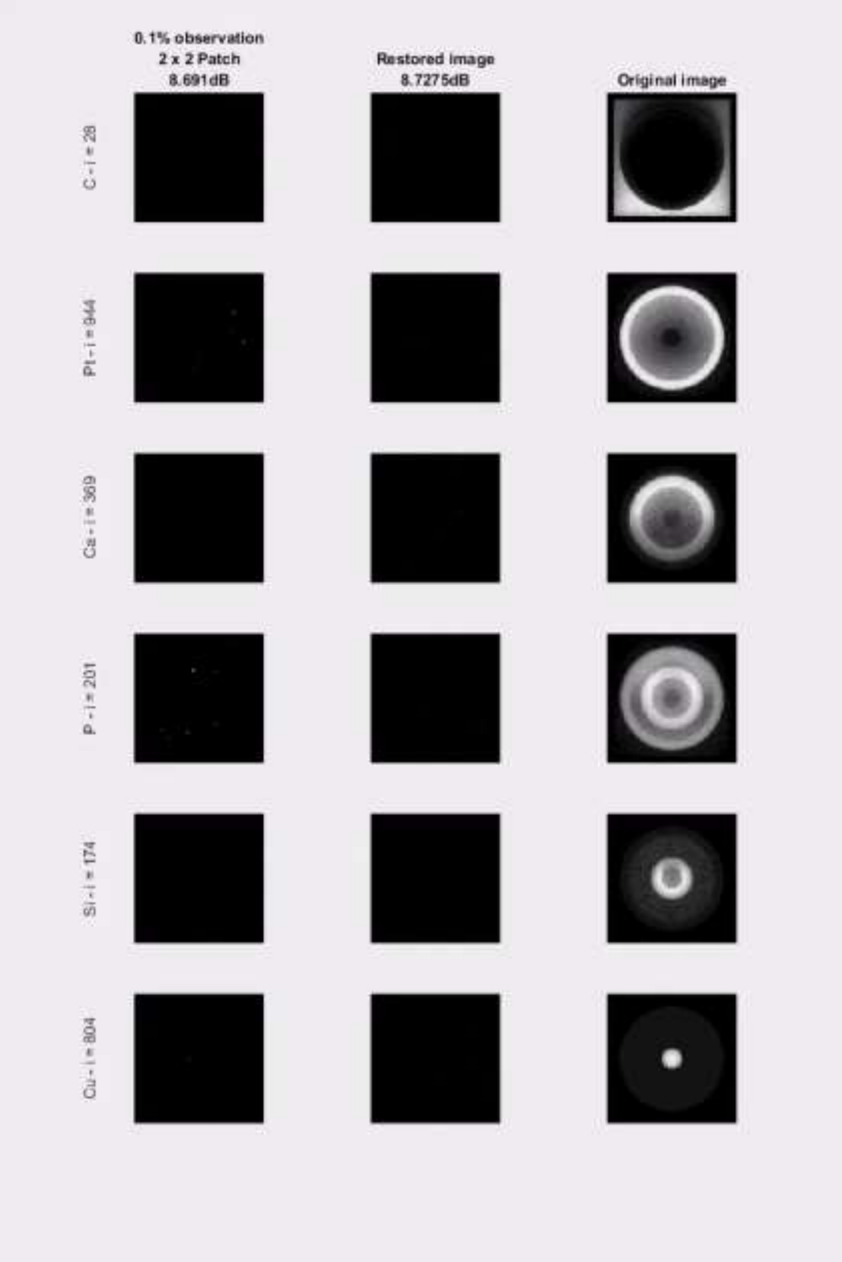
Silicon

Copper

BPFA reconstruction of subsampled EDS

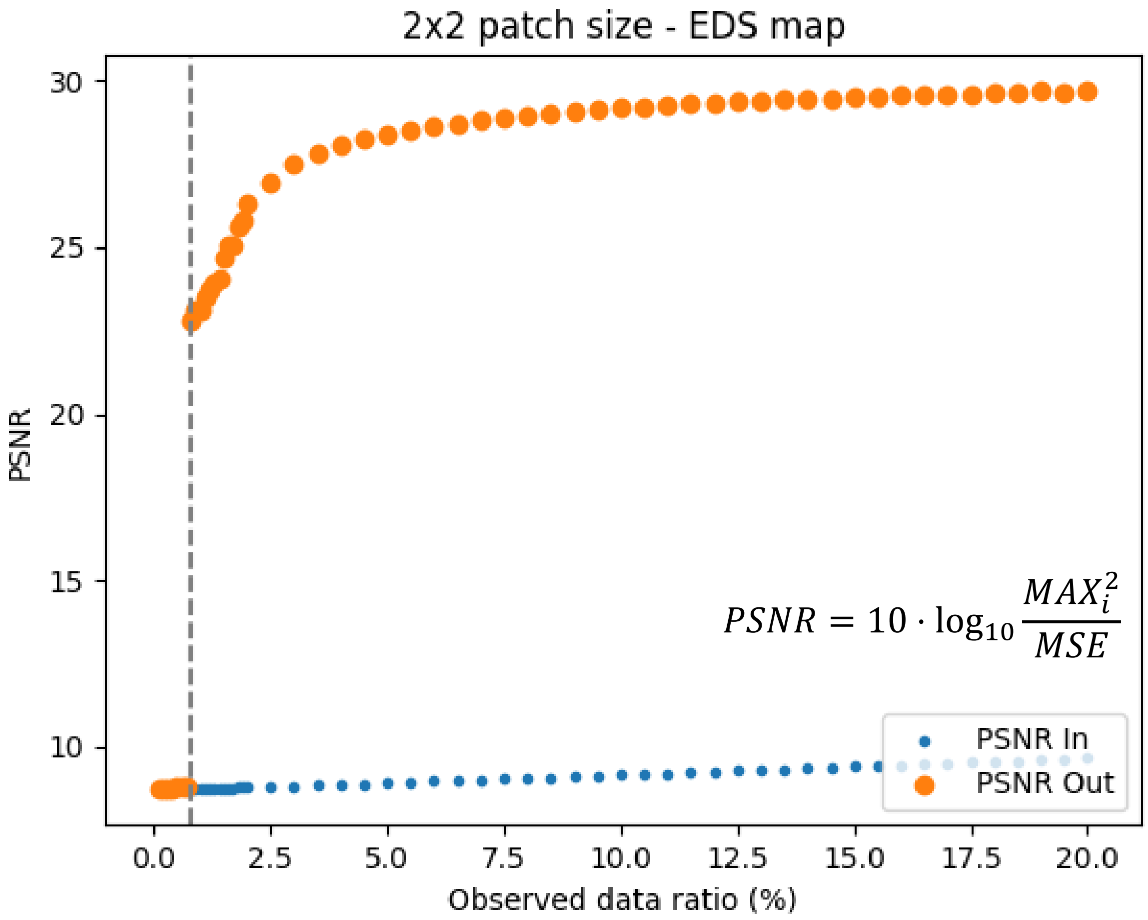
- **Tested both 2 x 2 and 4 x 4 patch sizes**
- **Subsample by zeroing out a large fraction of the simulated voxels**
- **Analyzed reconstruction output as function of observation ratio**





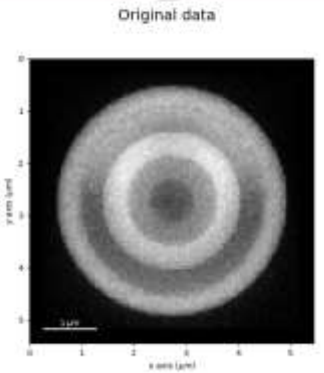
Reconstruction "phase transition"

- **Donoho-Tanner transition**
  - Reconstruction either fails completely, or does pretty well
  - Abrupt onset of success at a certain signal level

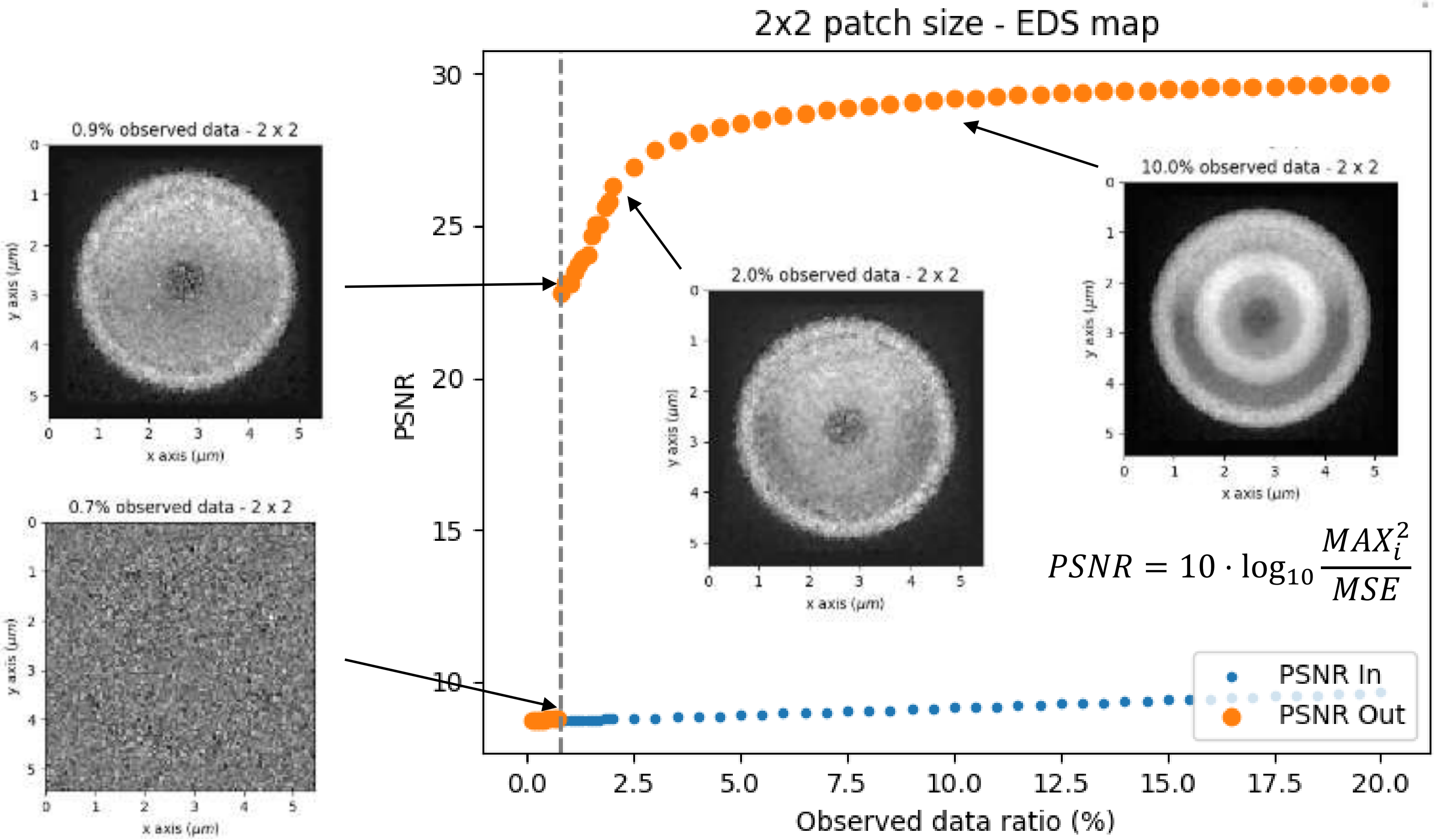


$$PSNR = 10 \cdot \log_{10} \frac{MAX_i^2}{MSE}$$

Reconstruction “phase transition”



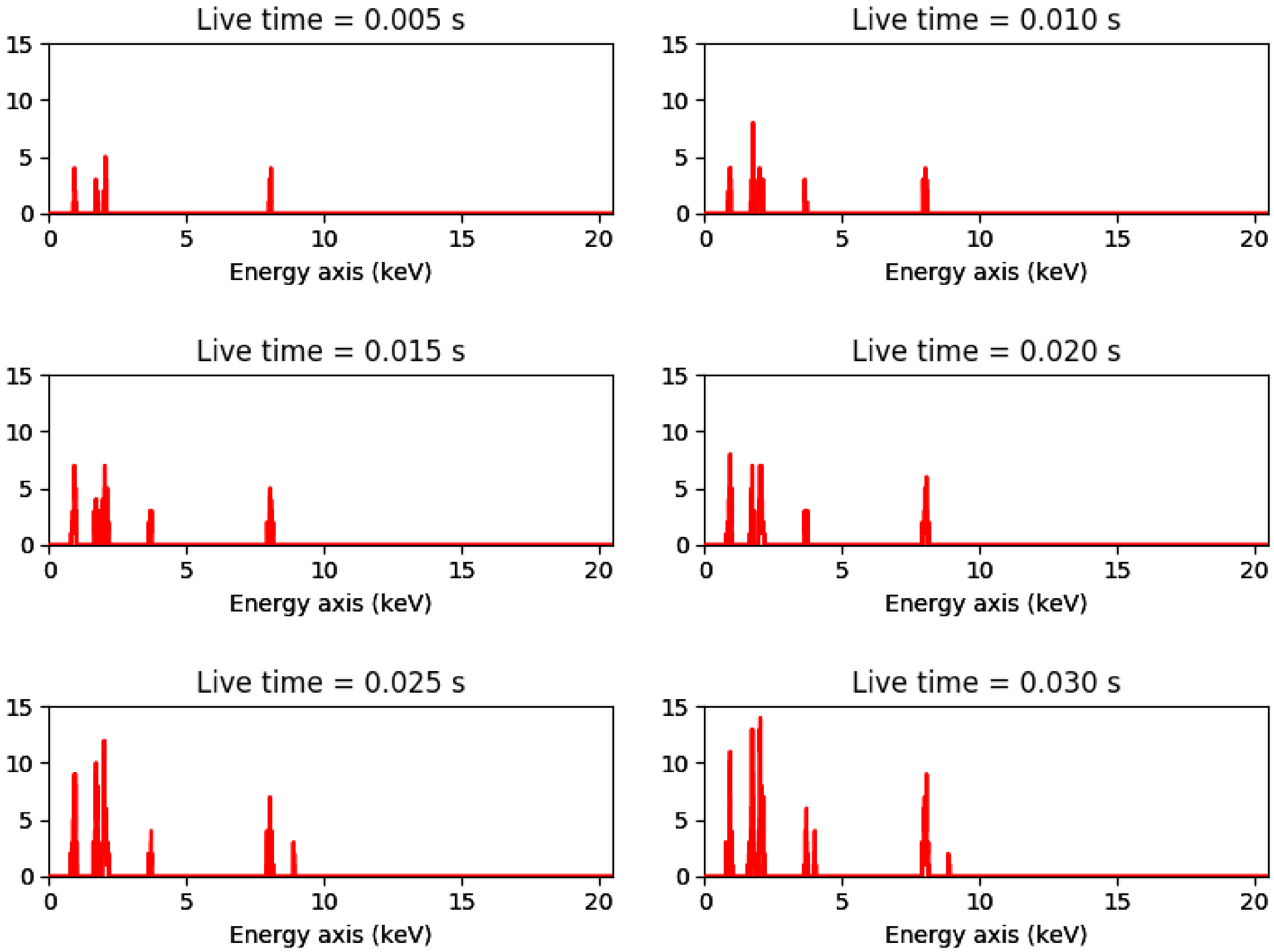
- **Donoho-Tanner transition**
    - Reconstruction either fails completely, or does pretty well
    - Abrupt onset of success at a certain signal level
    - Quality of reconstruction continues to improve with more data



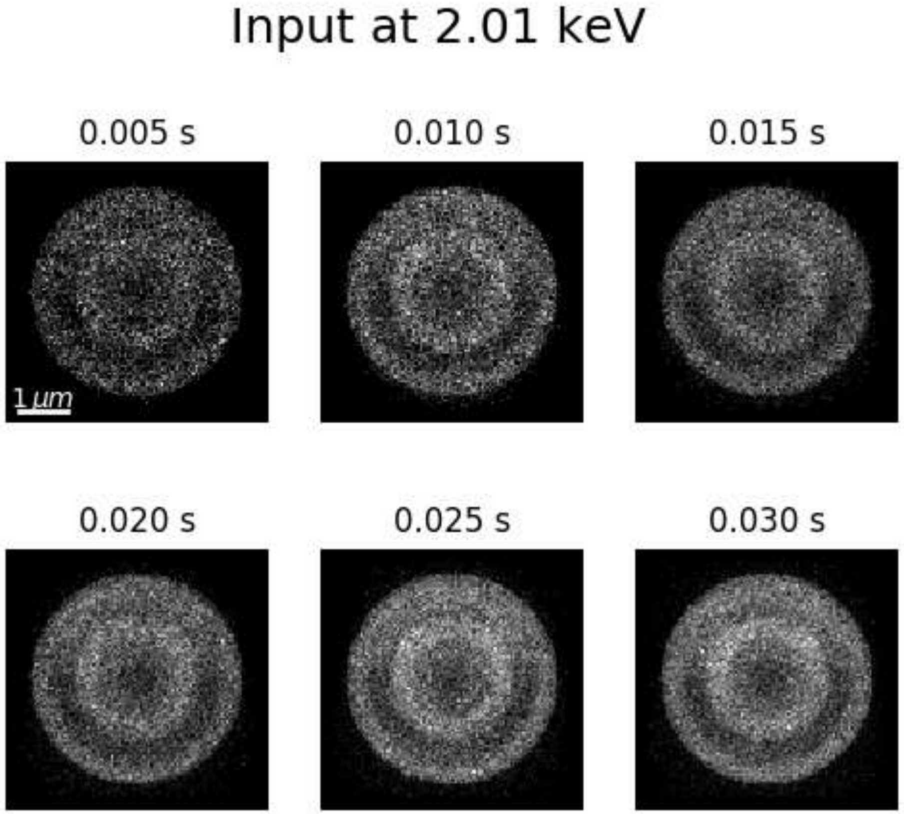
$$PSNR = 10 \cdot \log_{10} \frac{MAX_i^2}{MSE}$$

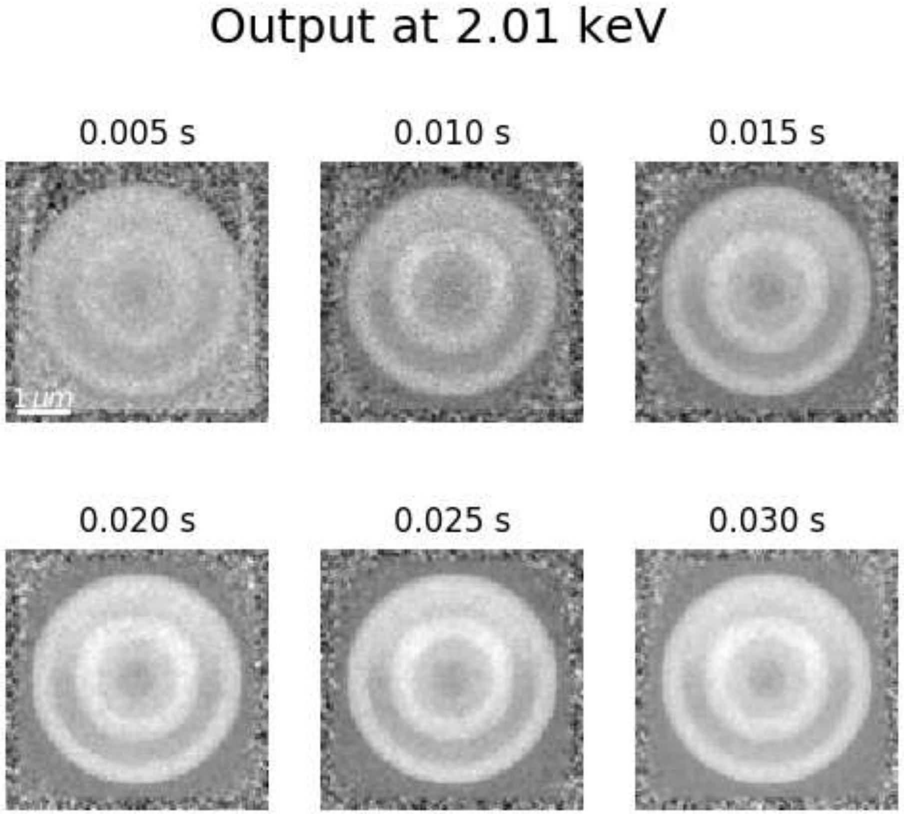
More realistic test

- **Random subsampling of entire datacube is not particularly realistic**
- **Instead, simulate EDS collections with varying dwell times**
  - Simulates faster experimental maps, without subsampling locations (or energies)



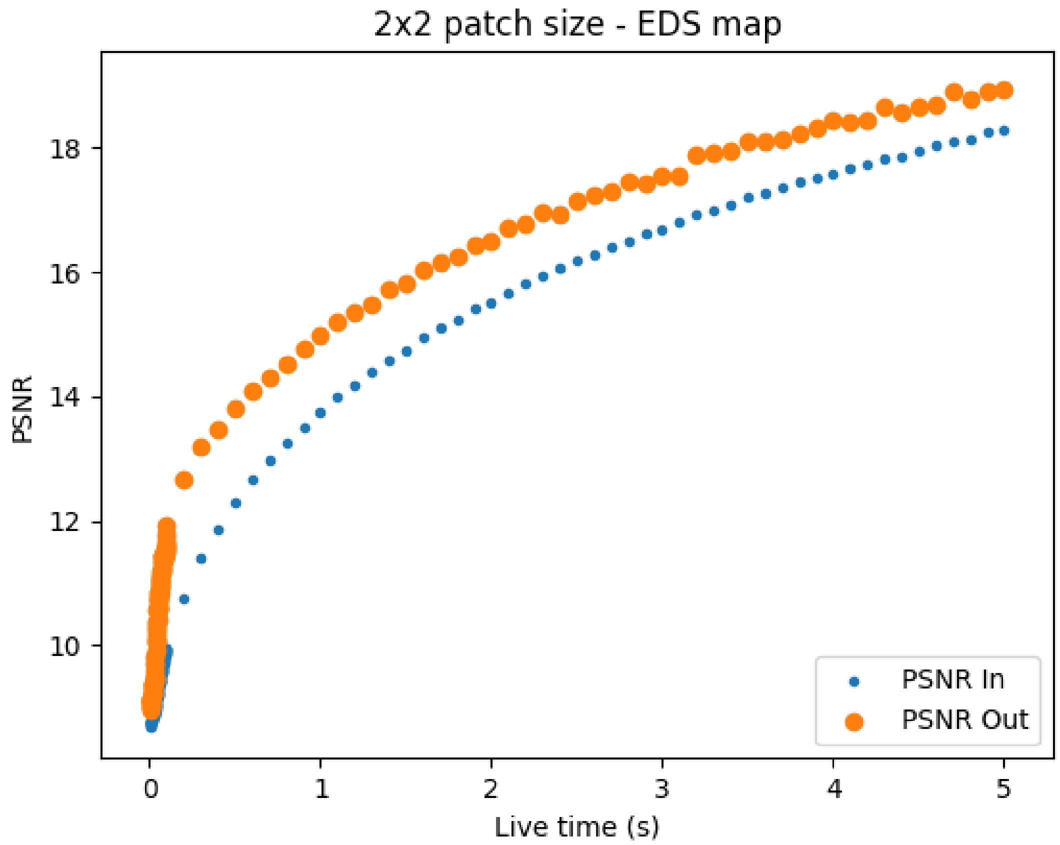
Results of live time variation


Input at 2.01 keV


Output at 2.01 keV

Results of live time variation

- **Improvement in PSNR statistics not as dramatic as true random sampling**
  - Never reaches high 20s value of previous example
- **Reasons?**
  - Suspect not sparse enough in energy axis
  - Thresholding of continuum X-rays could help
  - Still investigating these results



Future directions

Ongoing work

- **Improving results of reduced dwell time reconstructions**
    - X-ray continuum makes signal non-sparse, leading to bad performance
    - Implement some sort of thresholding, or artificially subsample energy dimension?
    - Eventually need to demonstrate effectiveness on experimental data
- **Extension of algorithm to 3D**
    - Should be relatively simple, and could enable even lower electron doses
- **Extend algorithm to allow incoming information**
    - Could make interactive EDS map collections more immediately informative
- **Make code more performant**
    - Currently single-threaded in Matlab and has not been optimized at all

Thank you!

Questions/comments?

joshua.taillon@nist.gov
x2913