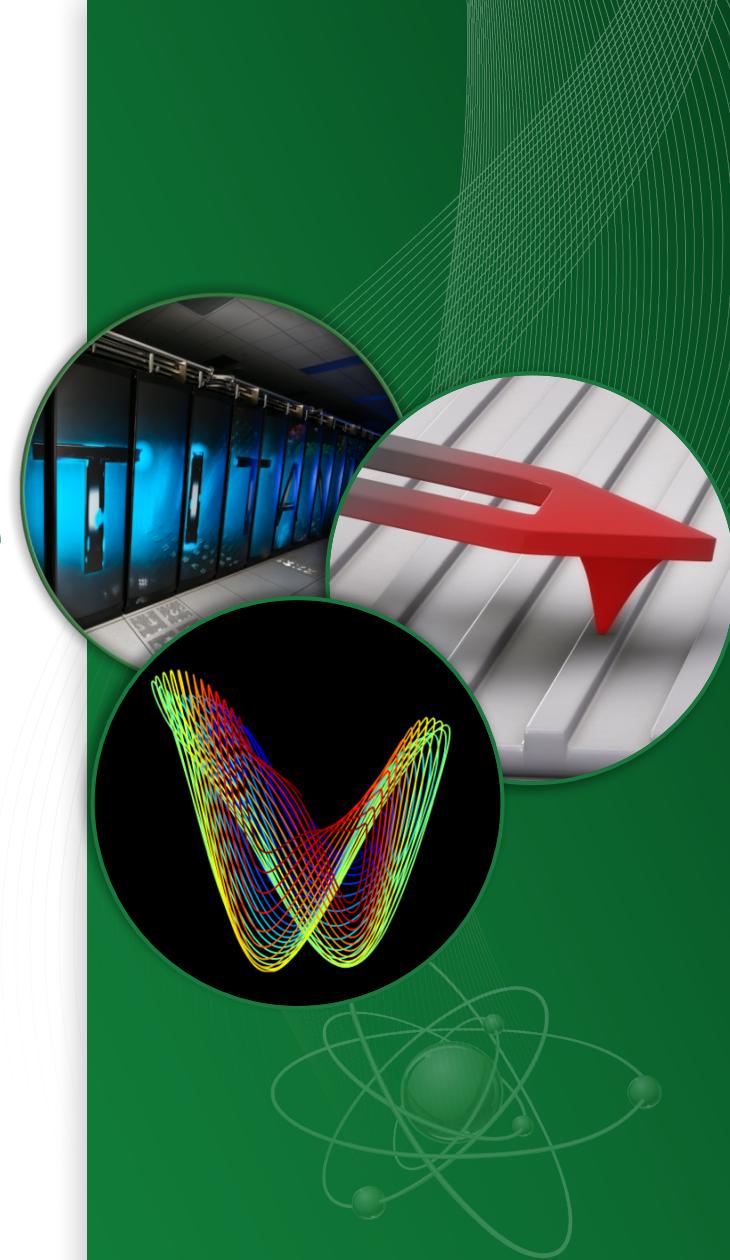
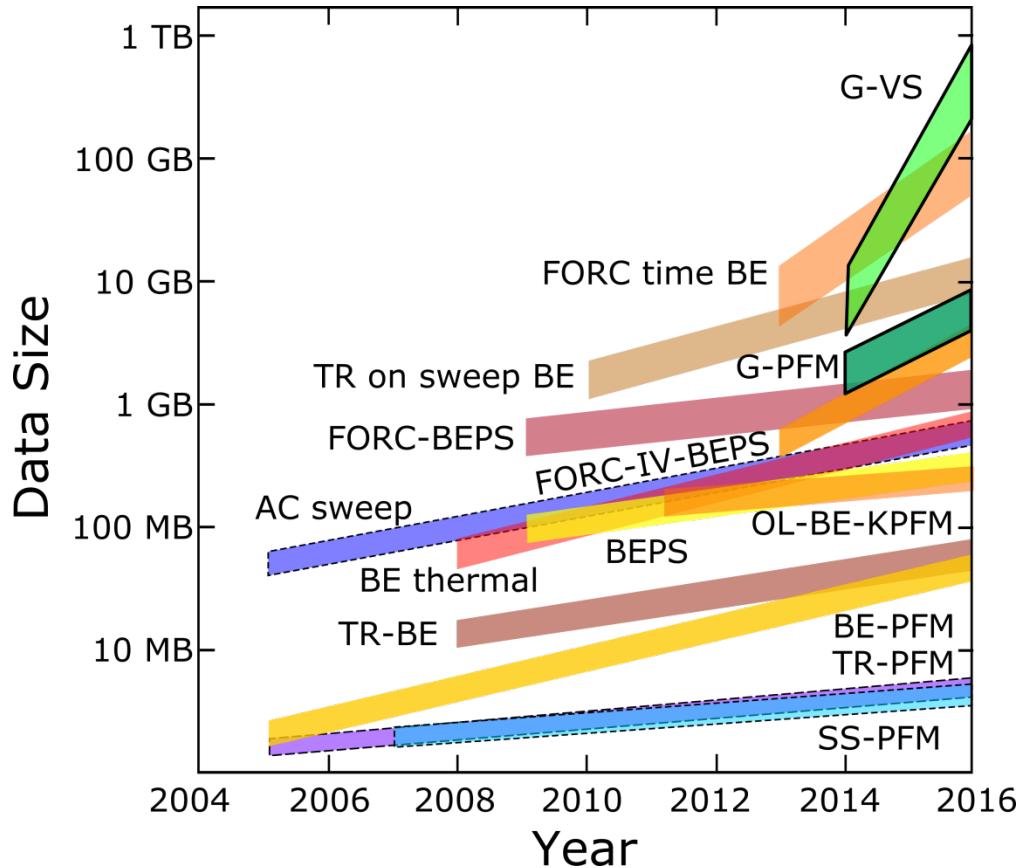


Pycroscopy – A community-driven software package for storage and analysis of imaging data

Suhas Somnath, Chris R.
Smith, Stephen Jesse,
Nouamane Laanait



Opportunities in Imaging / Microscopy



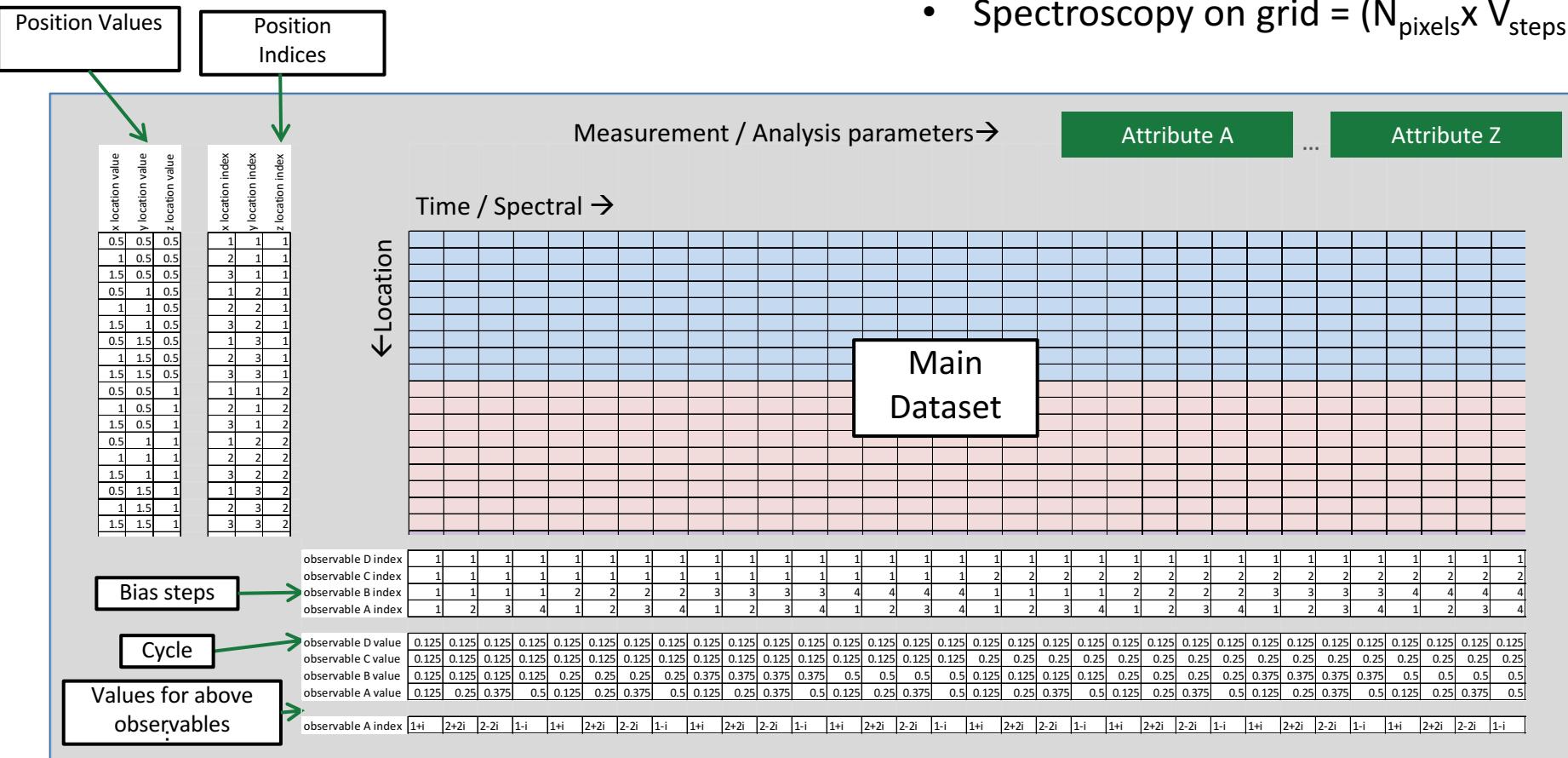
- **Growing data sizes & dimensionality**
 - Cannot use desktop computers for analysis
- **Multiple file formats**
 - Multiple data structures
 - Incompatible for correlation
- **Disjoint and unorganized communities**
 - Similar analysis but reinventing the wheel
 - Norm: emailing each other code, data
- **Expensive analysis software**

1. Universal Data Structure

- Data stored as 2D matrix of (position x spectral values) regardless of dimensionality
- Ancillary datasets explain the data

Example data types:

- 2D images = $(N_{\text{pixels}} \times 1)$
- Single spectra = $(1 \times Z_{\text{steps}})$
- Spectroscopy on grid = $(N_{\text{pixels}} \times V_{\text{steps}})$

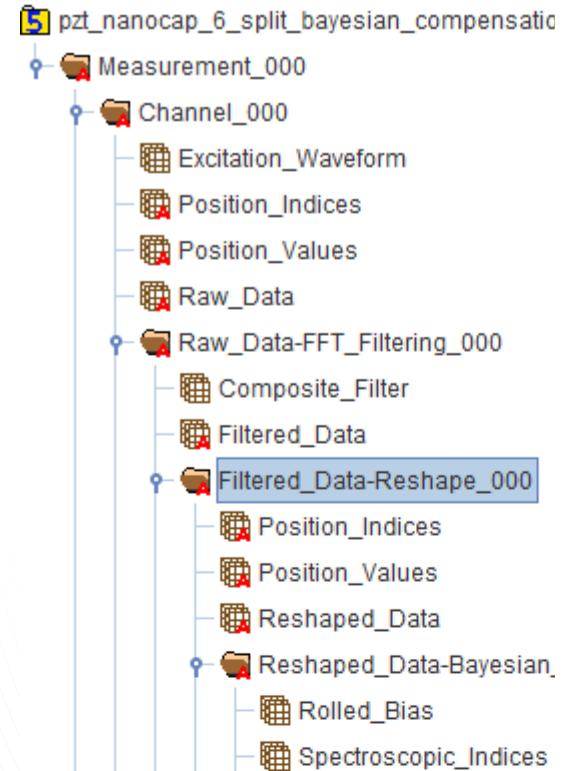
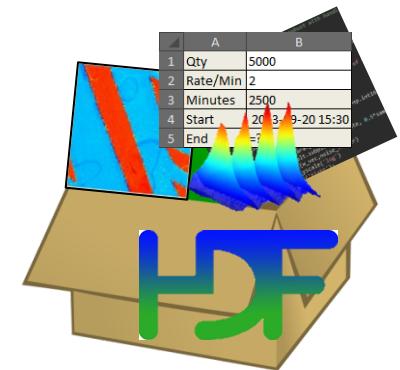


Additional information available at:

Suhas Somnath, somnaths@ornl.gov https://github.com/pycroscopy/pycroscopy/blob/master/docs/Pycroscopy_Data_Formatting.pdf

2. Hierarchical Data Format (HDF5)

- A HDF5 file is a smart container
 - Capable of storing multidimensional datasets, Images, text, measurement parameters
 - Contents organized like traditional folders and files
 - **Datagroup** - Analogous to folders in a file system
 - **Dataset** – contains 1 to N dimensional data
 - Integer, floating point, complex numbers etc
 - **Attributes** – {Key : value} pairs useful for describing data and experimental parameters, etc.
- Easily accessible – C, C++, python, Java....
- Parallel read / write, HPC compatible
- Tree structure + nomenclature +attributes are **records of workflow** applied to dataset



3. Python for Scientific Research

User-friendly / Easy to code

Numerous, **powerful** libraries for science



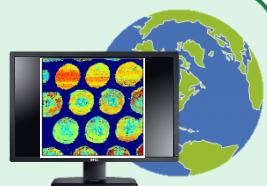
NumPy



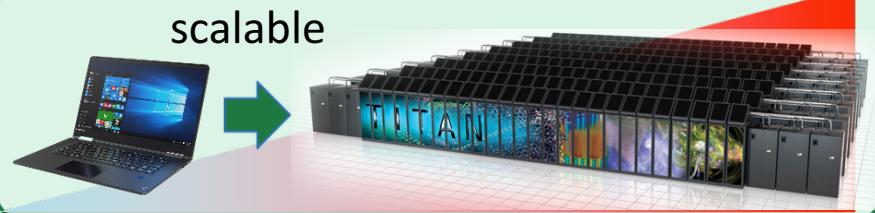
C/
LabVIEW



python



Java



Well
documented

Strong user
community



- Accelerates scientific progress

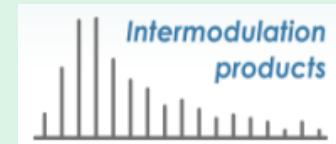


Microscopy, micro-biology, data science,
neutron science....

Many microscopy companies using python:



Electron Microscopy



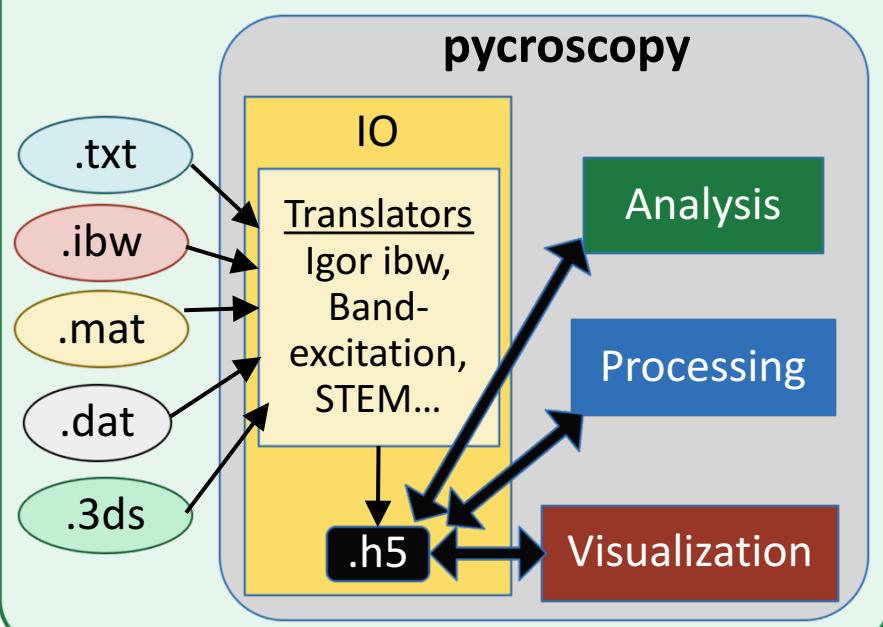
Multifrequency AFM

All for a princely sum of **\$0!**

4. Centralized, Open Software Repository

Pycroscopy is the free, open-source python package for analyzing and storing microscopy data
<https://github.com/pycroscopy/pycroscopy>

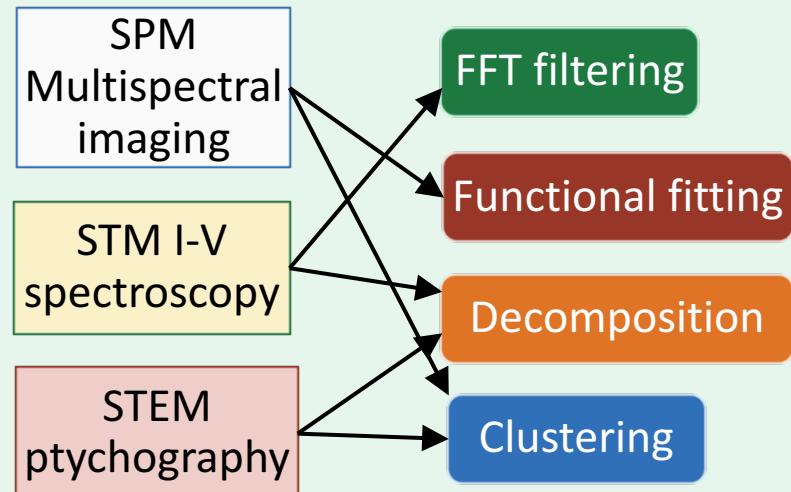
- H(DF)5 file is the hub for all operations
- Analysis, processing, visualization available after translation to .h5



- Freely available
- Easy to install



- Universal data format allows generalized code to be reused for any kind of data



- Well documented, easy to read or modify

```
class Cluster(h5_main, method_name, num_comps=None, *args, **kwargs)
```

Bases: `object`

Pycroscopy wrapper around the `sklearn.cluster` classes.

```
do_cluster(rearrange_clusters=True)
```

Clusters the hdf5 dataset, calculates mean response for each cluster, and writes the labels and mean response back to the h5 file

Parameters: `rearrange_clusters` : (Optional) Boolean. Default = True

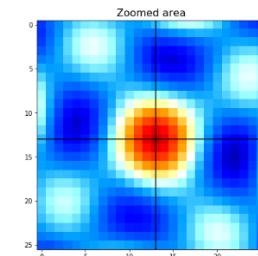
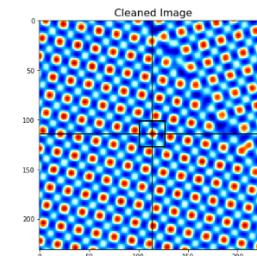
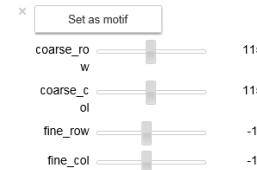
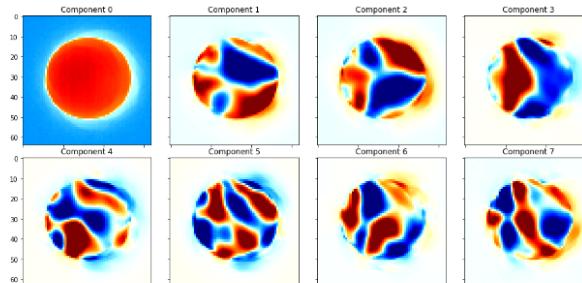
Scientific Workflows in pycroscopy

jupyter notebooks – Open source web application for creating, sharing documents with live code, equations, explanatory texts, and interactive visualization

Code snippets with visualization

```
In [11]: # Visualize the eigenvalues or loading maps from PCA:  
eigenvalues = np.reshape(h5_v[:num_comps], (-1, num_sensor_rows, num_senso  
r_cols))  
eigenvalues = np.transpose(eigenvalues, (1, 2, 0))  
fig, axes = px.plot_utils.plot_map_stack(eigenvalues, num_comps=num_comp  
s, heading='Eigenvectors',  
cmap=px.plot_utils.cmap_jet_white_
```

Eigenvectors



Interactive visualization using sliders, buttons, etc.

Notes and explanations

Image cleaning and atom finding using pycroscopy

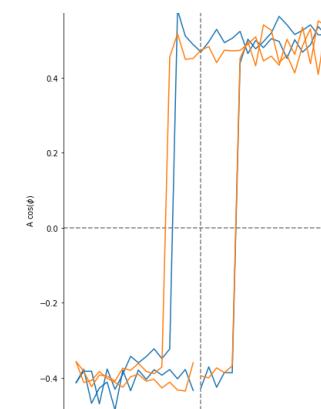
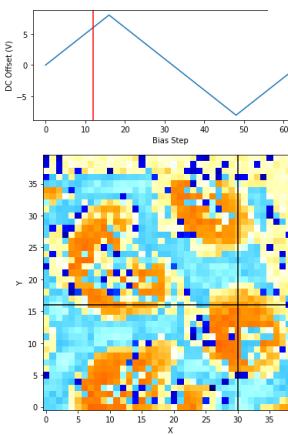
Suhas Somnath, Chris R. Smith, Stephen Jesse

The Center for Nanophase Materials Science and The Institute for Functional Imaging for Materials
Oak Ridge National Laboratory
1/19/2017

Configure the notebook first

```
In [1]: # set up notebook to show plots within the notebook  
%matplotlib inline
```

Suhas Somnath, somnaths@ornl.gov

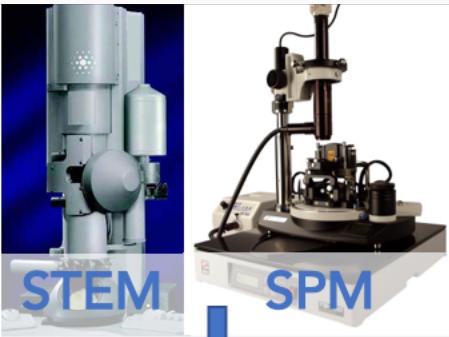


Is a report by itself - Shows all code, documentation, results...

Integrating with Supercomputing



Bellerophon Environment for Analysis of Materials



Lightweight
Desktop
Application

1. Large multi-dimensional data sets from advanced microscopes are uploaded through BEAM's lightweight, user-friendly desktop application.

Popular pycroscopy analysis routines scaled up to supercomputers

Web Server & Data Storage System



2. Data is uploaded over a high-speed network connection to BEAM's cloud-based, petabyte-scale filesystem and web server.



Oak Ridge Leadership
Computing Facility

3. Advanced analysis algorithms are executed on the uploaded data using America's leadership-class supercomputers like Titan at the OLCF and CADES compute clusters with the push of a button.

ORNL Compute and Data
Environment for Science

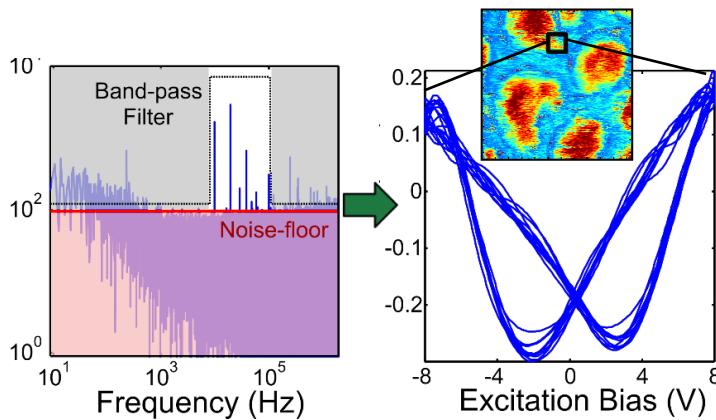


BEAM allows users to access their uploaded data, on-demand supercomputing analysis, and powerful visualization from anywhere at anytime with just an internet connection.

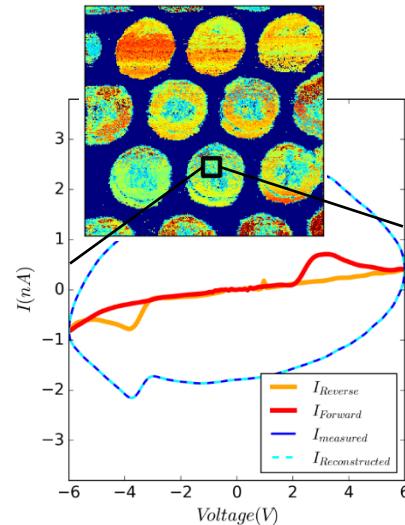
Lingerfelt et al., *Procedia Computer Science* **80**, 2276-2280 (2016).

Scientific Advancements with Pycroscopy

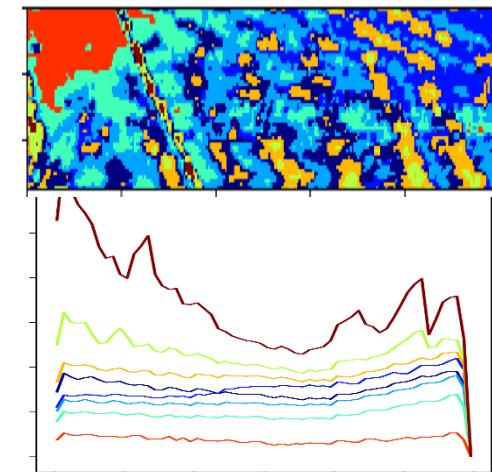
3,500x faster imaging via adaptive signal filtering, linear unmixing of signals



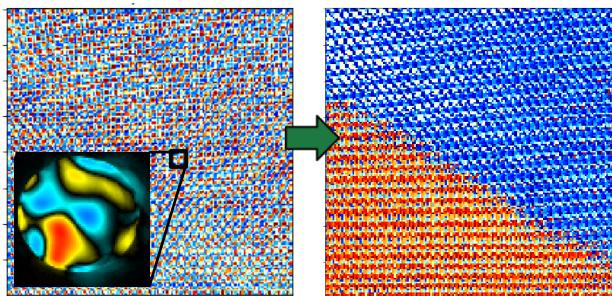
200x faster I-V spectroscopy via Bayesian inference



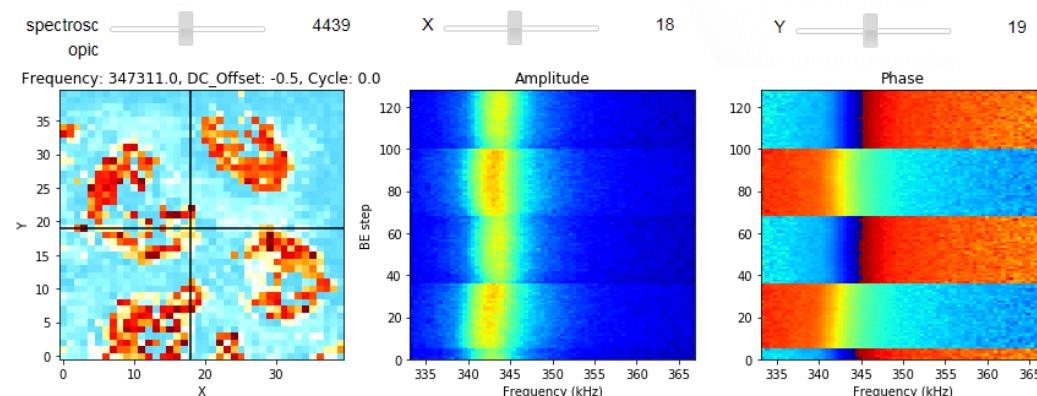
Denoising and clustering to identify superconductivity at the nanoscale



Identifying invisible patterns using multivariate analysis



Navigating multidimensional data



Collaborate with us

Industry partners we are actively collaborating with to integrate pycroscopy:



State-of-art atomic force
microscope manufacturer



Scanning transmission
electron microscope
manufacturer



Scientific data visualization
& analysis software
provider

- We would be happy to collaborate to add functionality and other features of interest into pycroscopy
- Join our slack project at <https://pycroscopy.slack.com> to discuss about pycroscopy
- Feel free to get in touch with us at pycroscopy (at) gmail [dot] com
- More information available at: <https://github.com/pycroscopy/pycroscopy>