



BERKELEY LAB

LAWRENCE BERKELEY NATIONAL LABORATORY

From face detection to the faces of scientific images: Scaling Analytics for Image Data from Experiments



Dani Ushizima, Harinarayan Krishnan, Talita Perciano, Dula Parkinson,
Peter Ercius, Wes Bethel and James Sethian

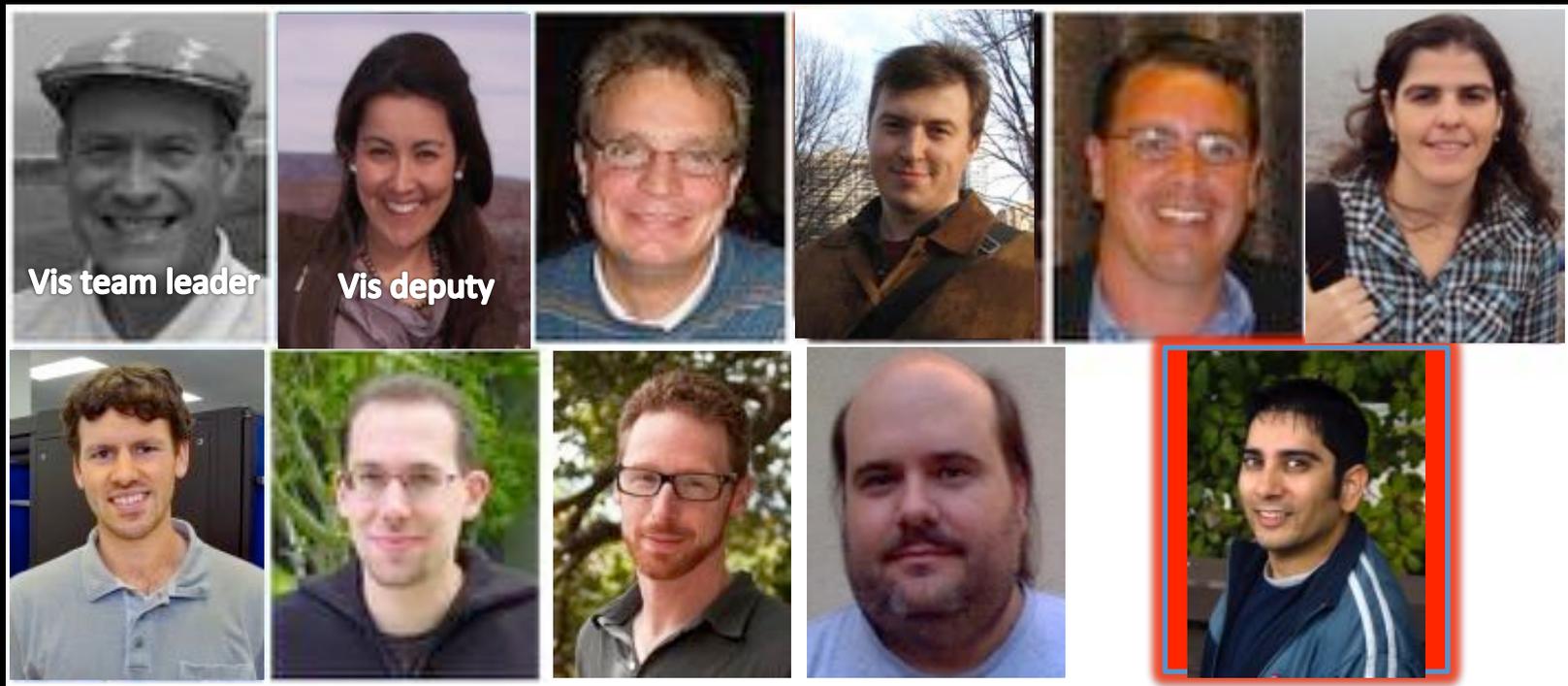


Lawrence Berkeley National Laboratory, Berkeley, CA, USA



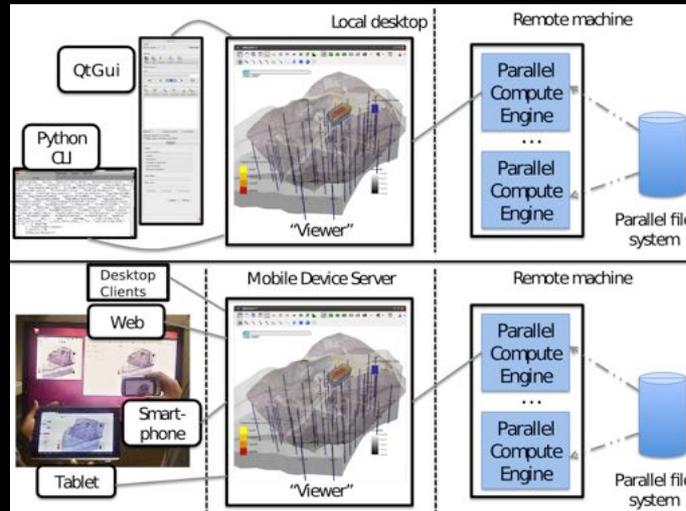
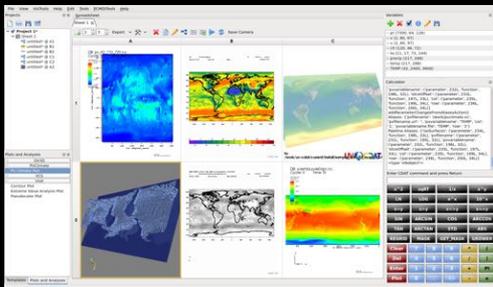
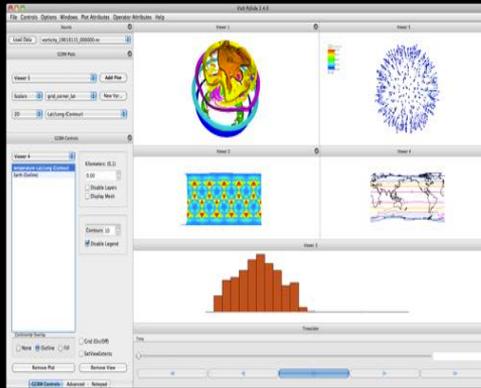
Data Analytics & Visualization

DAV Group



Wes Bethel, Daniela Ushizima, Gunther Weber, Dmitriy Morozov, Hank Childs, Talita Perciano, Mark Howison, Oliver Ruebel, Burlen Loring, David Camp, Hari Krishnan

Custom UI



Domain Processing

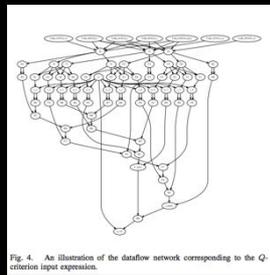
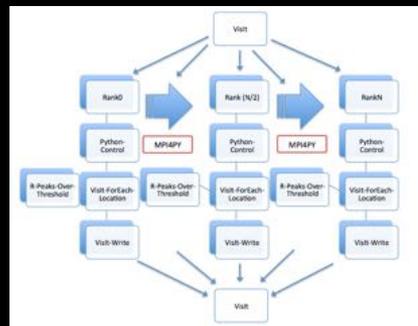
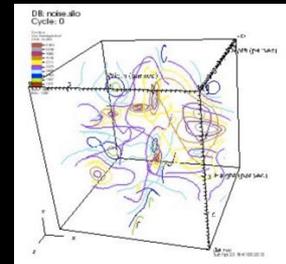
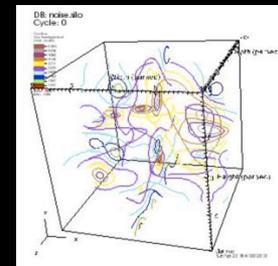
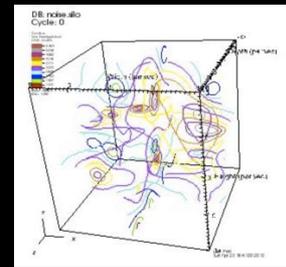
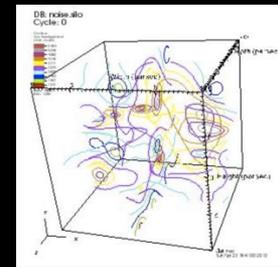
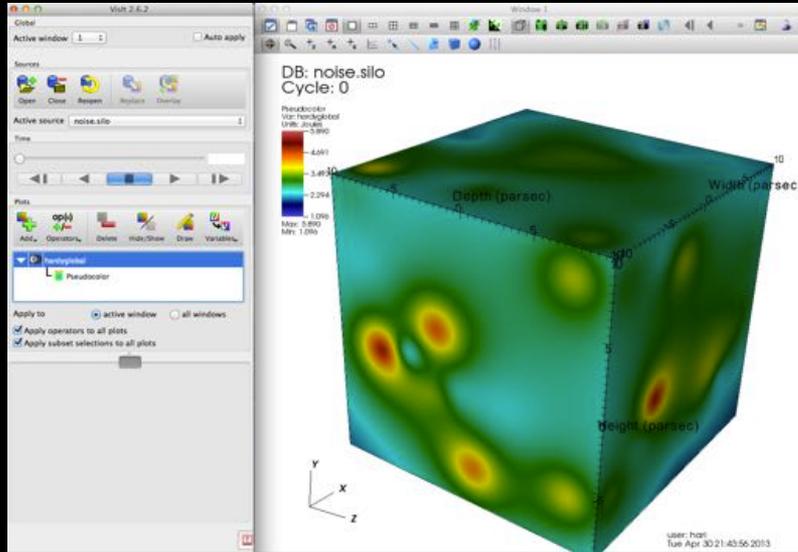


Fig. 4. An illustration of the dataflow network corresponding to the Q-criterion input expression.

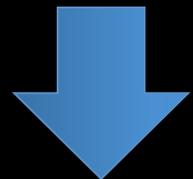
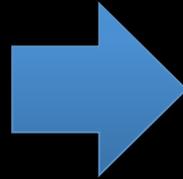
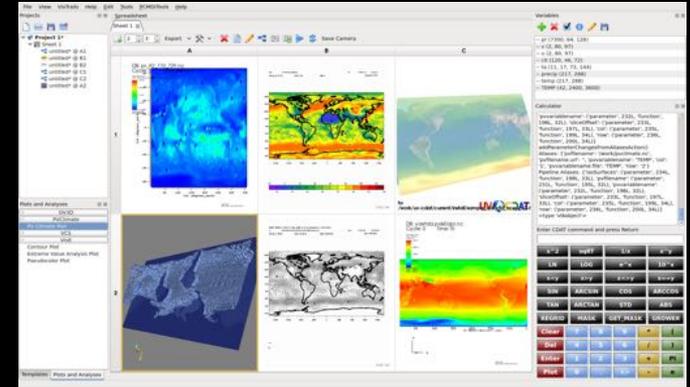


Collaborators



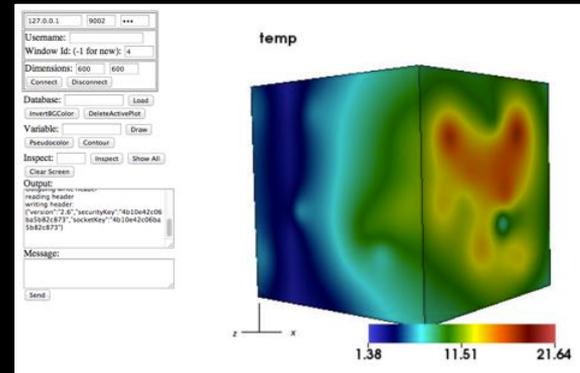
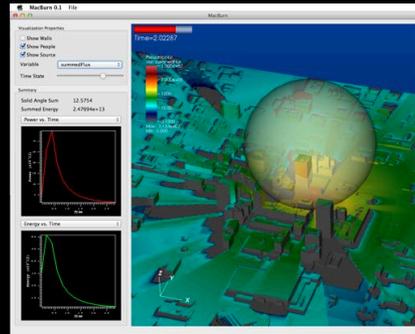
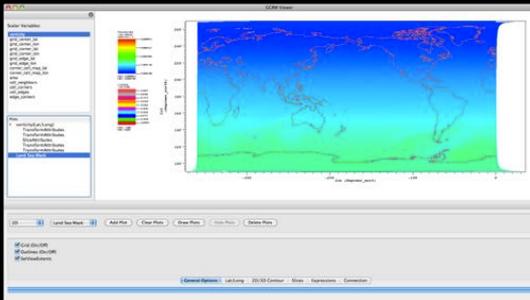


Embedded



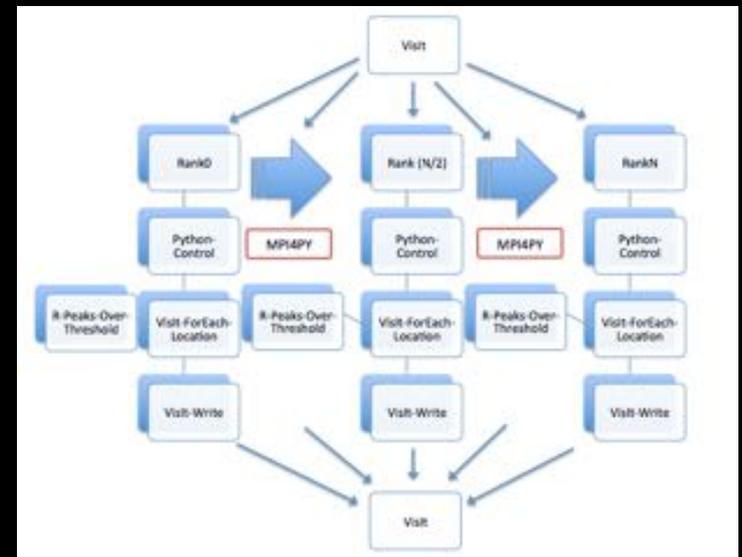
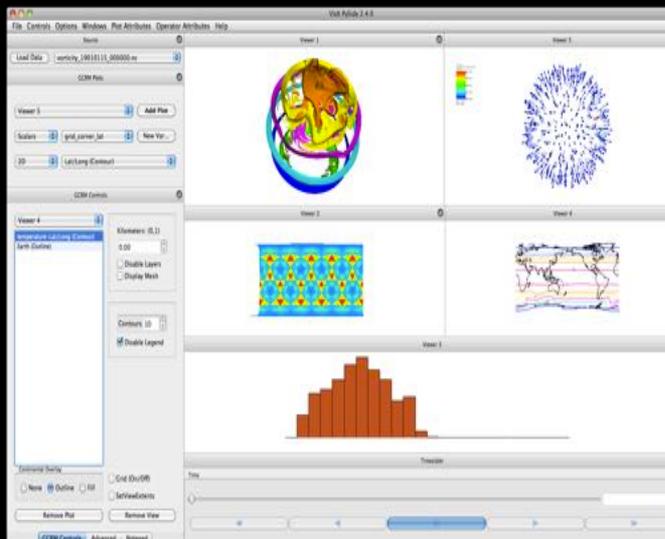
Lightweight, Collaboration

Tailored Vis



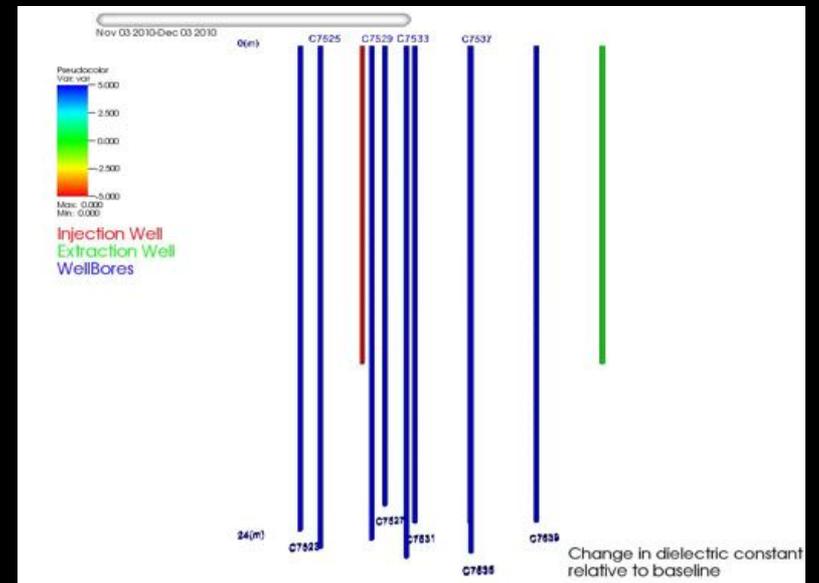
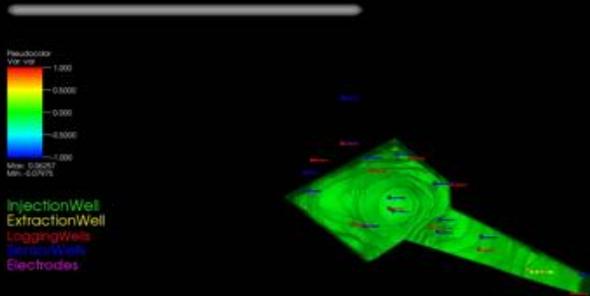
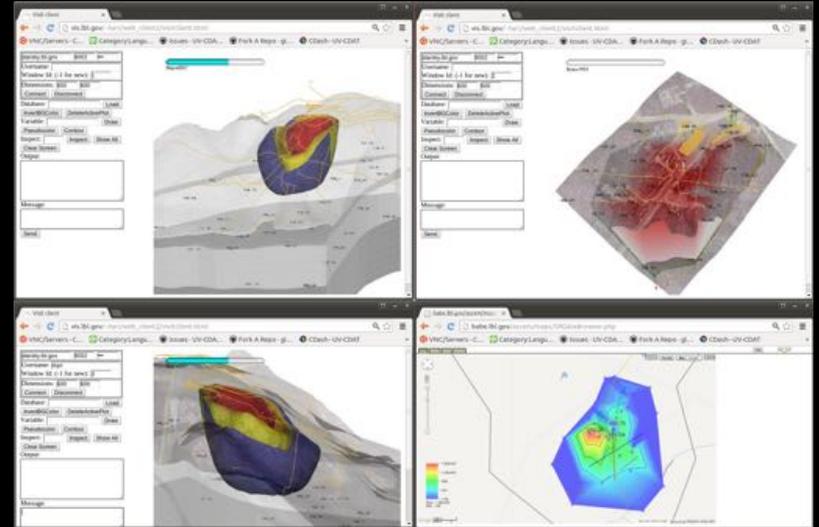
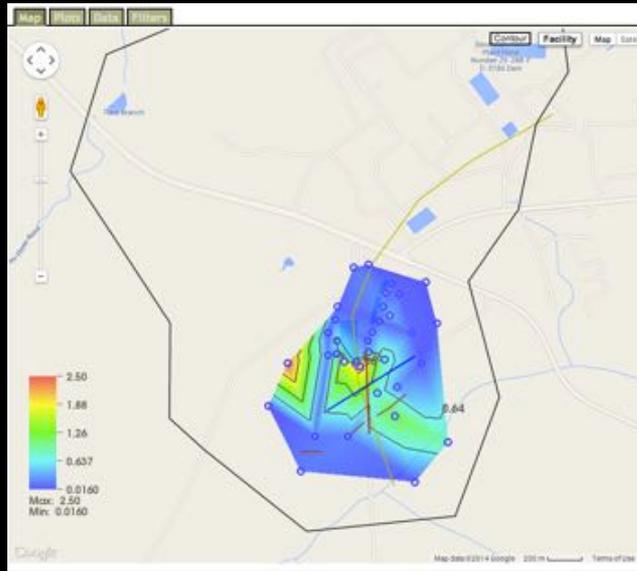
Climate Science

- Customize user interfaces:
 - Interface – Lat/Long 2D Grid, 3D globe, Continental Overlays.
 - Options – Zonal Mean Averages, Extreme Values, Peaks Over Threshold, etc..
- Collaboration and Provenance:
 - Collaborate, Control, & Communicate result with peers Record and recreate workflows.
- Extend Capabilities:
 - Extend Extreme Value Analysis or Peaks-Over-Threshold algorithm or write custom analysis routines to explore data.



Environmental Science

Data Browser (SDM, ACS), Deep Vadose Zone (PNNL), John Peterson, Susan Hubbard



Astrophysics

Astrophysics (YT),
Climate science (R),
VTK (python)

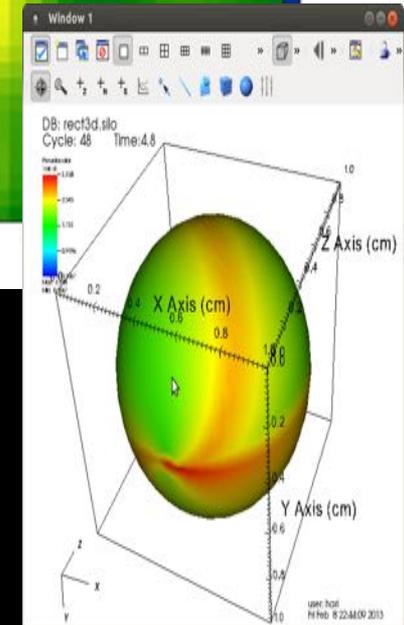
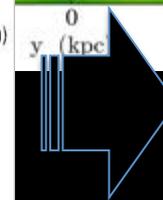
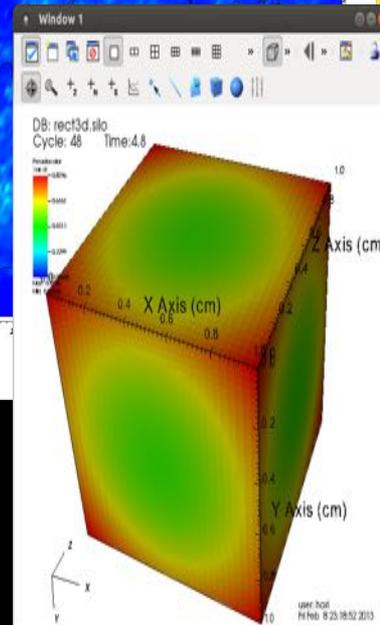
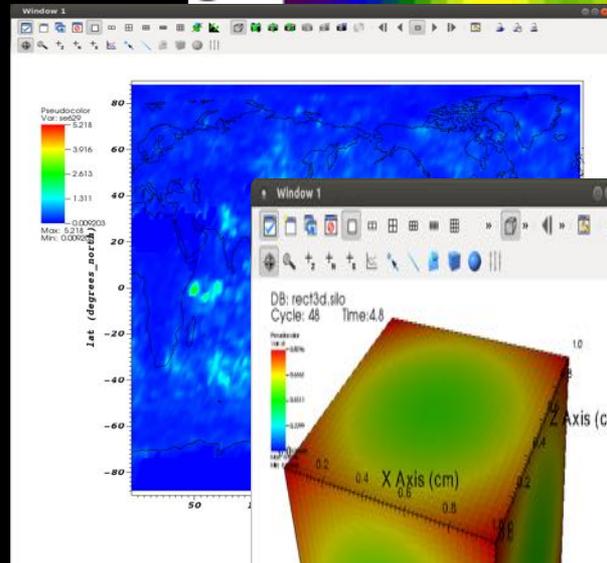
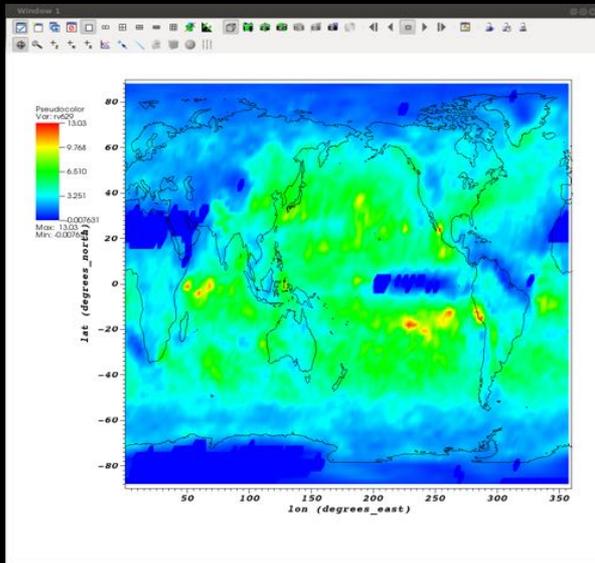
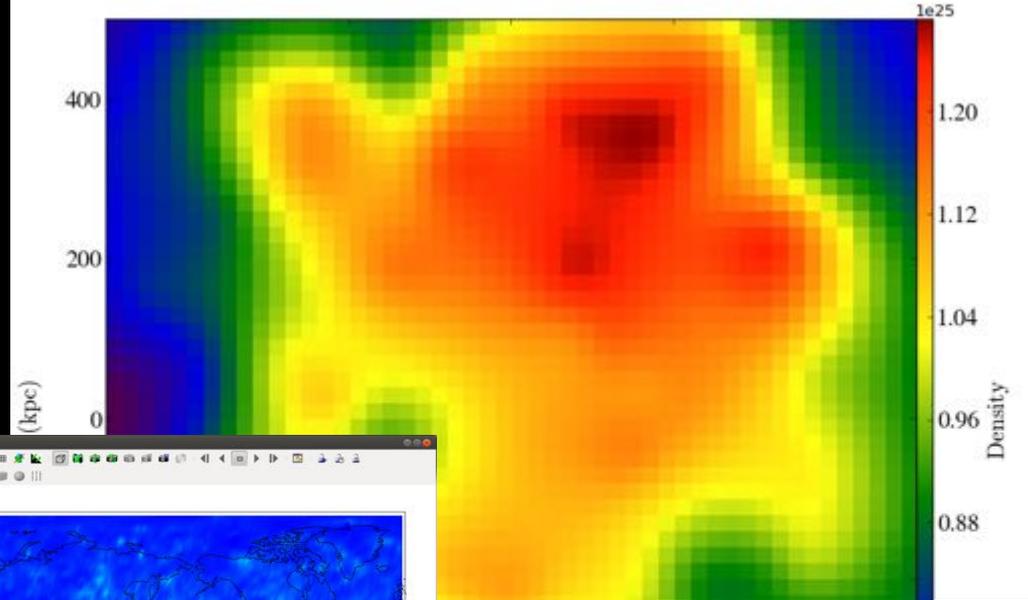
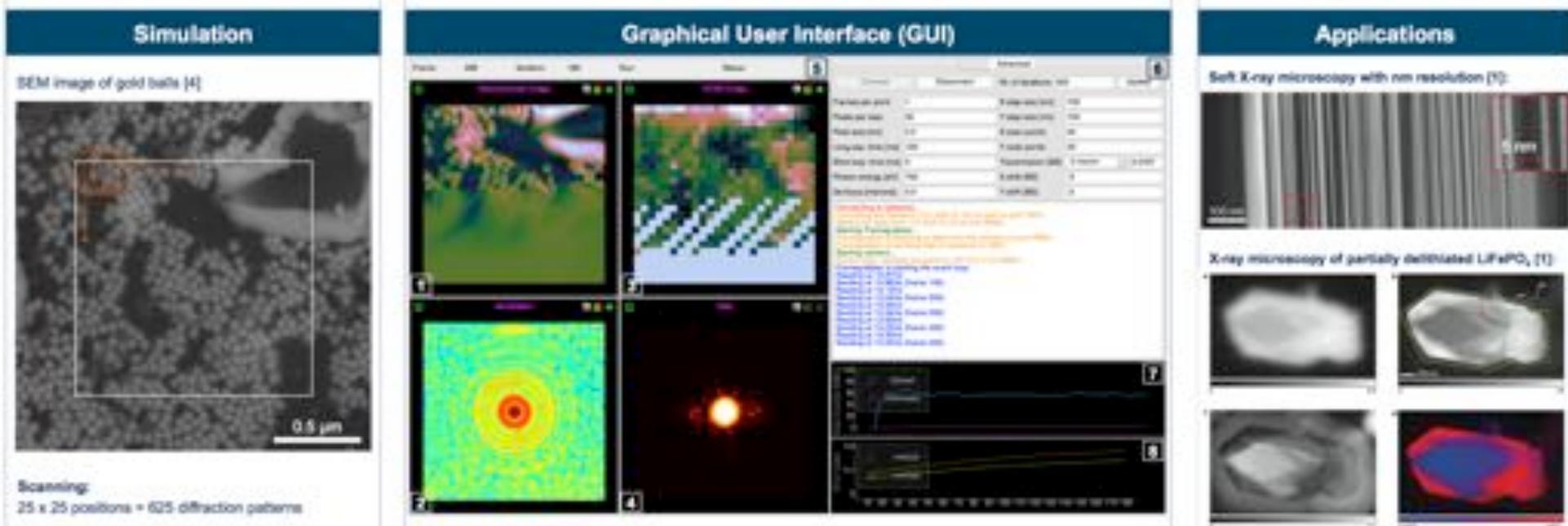


Image Processing, Reconstruction, Segmentation, and Analysis



Rest of the talk...

Overview

1. Investigating image-based experiments:
 - a. Material Science-focused image analysis;
 - b. Health-focused image analysis;
2. Computer methods and results;
3. Scaling through partnerships;
4. Image in the exascale landscape

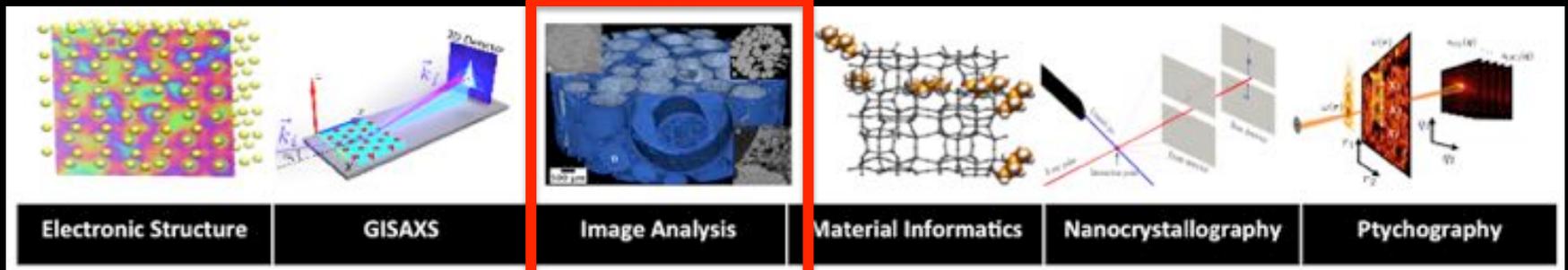
OUR TOOLS





CAMERA

Center for Applied Mathematics for Energy Research Applications





6/16/16

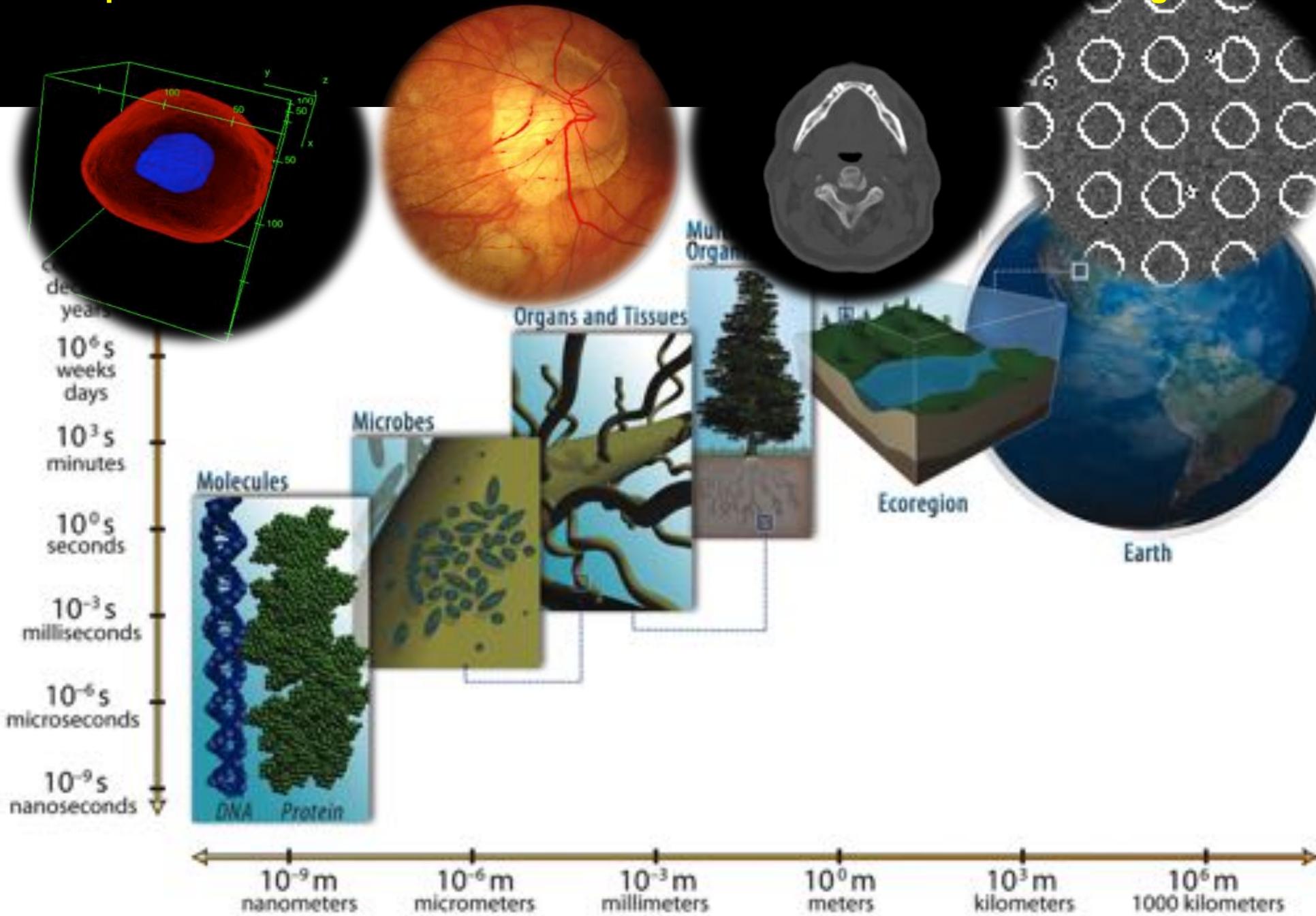


Nanoparticle

Ocular fundus

Head CT

Radar image



SAIDE projects – from nano to meter scale

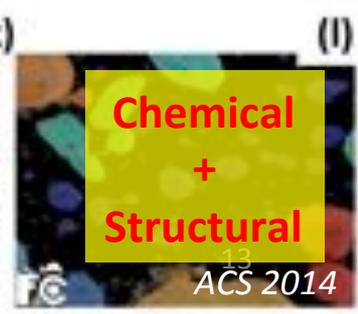
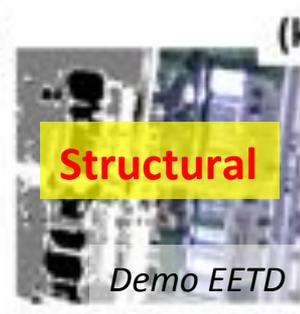
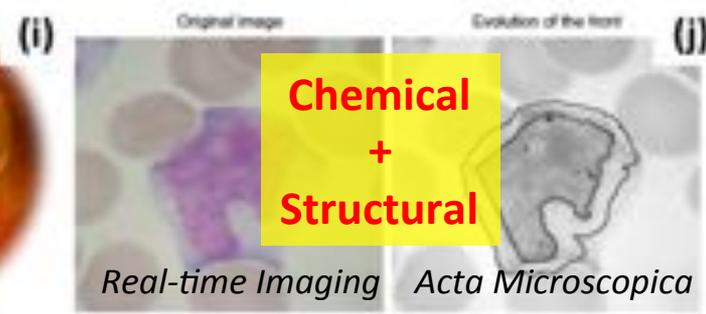
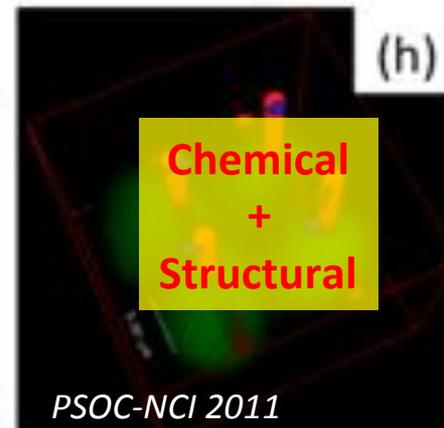
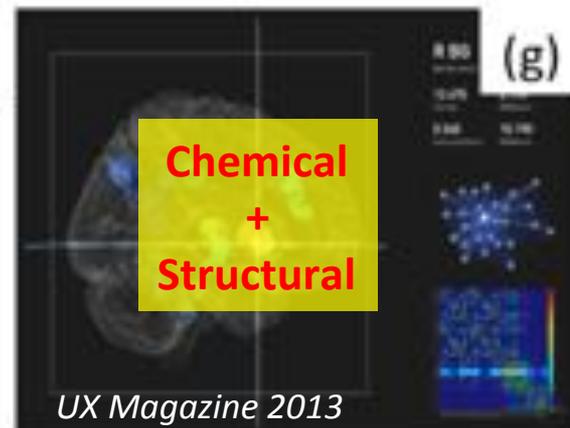
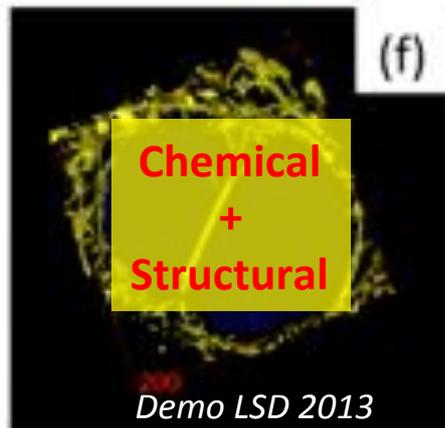
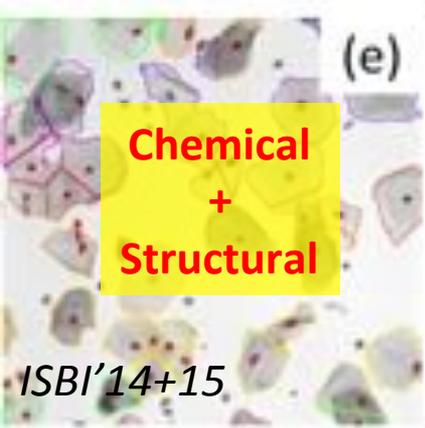
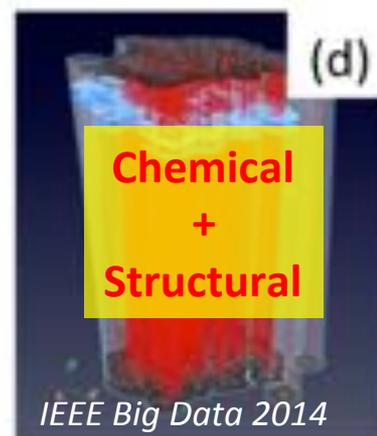
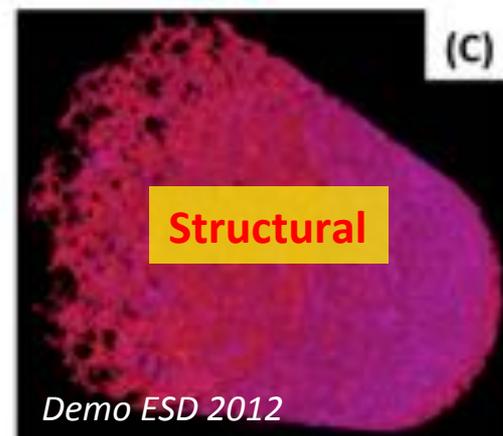
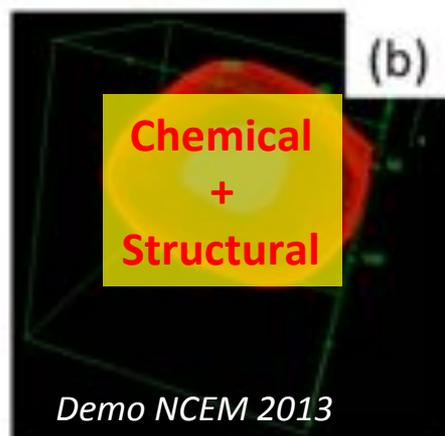
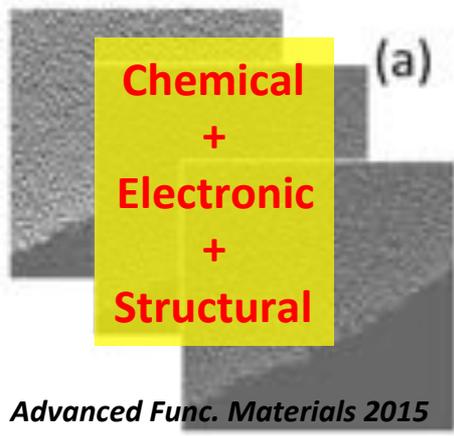
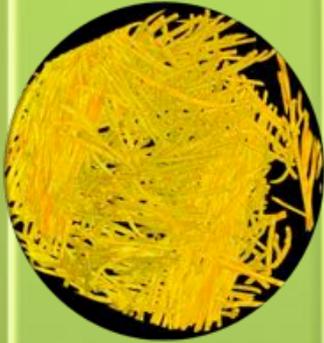


Image Across Domains



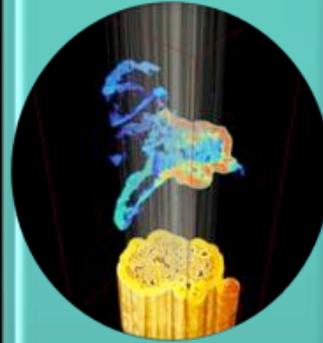
Specimens

- Materials, composites, compounds and biological samples.



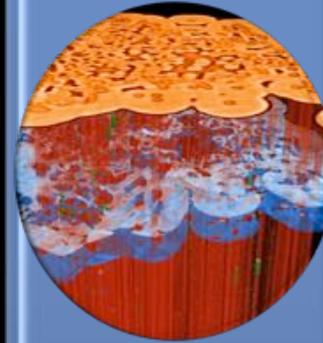
Formats

- Tiff, jpeg, hdf5, feature vectors, multi-resolution pyramids, binaries.



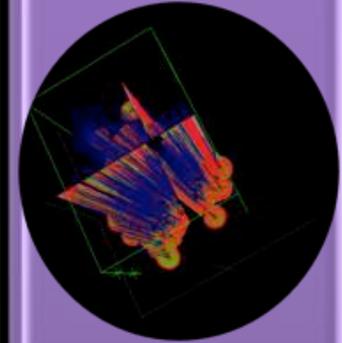
Data Analysis

- Morphometry;
- Spectral content;
- Multimodal;
- Templates.



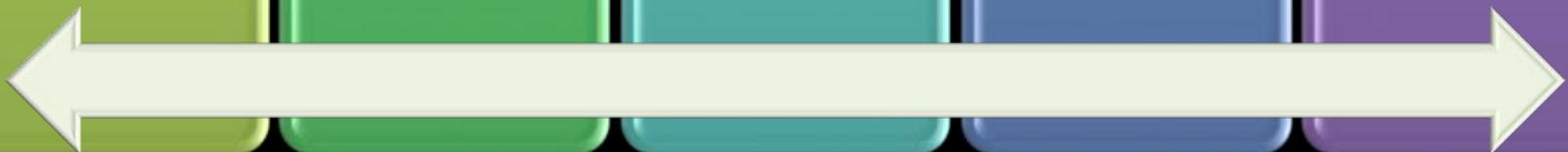
Data Understanding

- Clustering;
- Classification;
- Randomized schemes;
- Visualization.

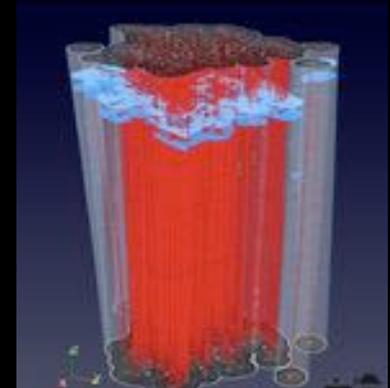
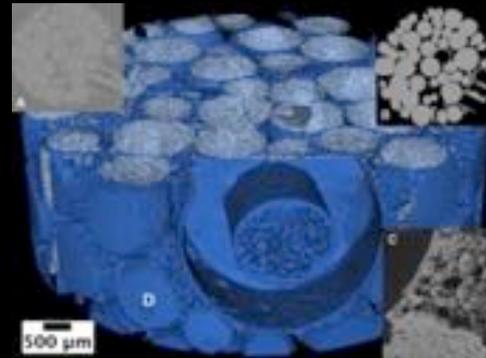
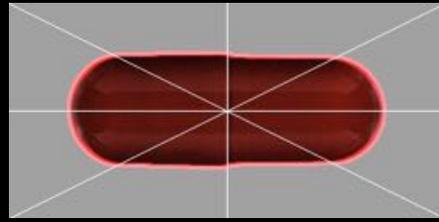
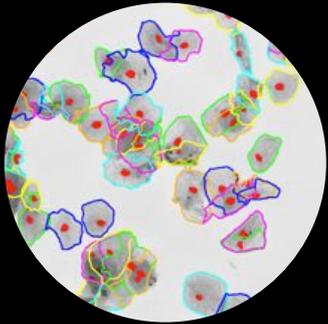


Reproducible research

- Data repositories;
- Software repositories;
- Collaboration.



1. Image analysis @ UCB-BIDS/LBL



Cervical
cells

Pill
identifier

Geological
samples

Resistant
composites

Health

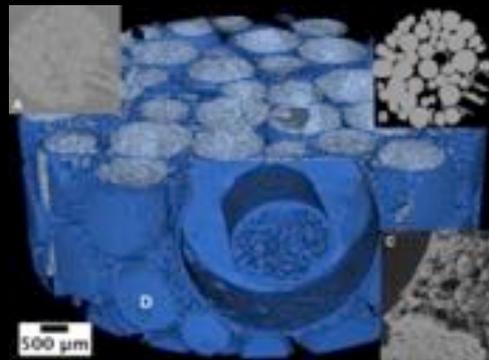


Material Science

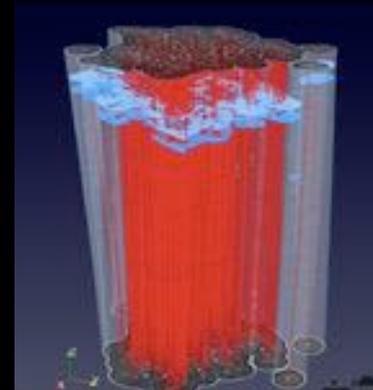


Free-software, open-source, git, reproducible

1.a. Image analysis @ LBL/UCB-BIDS



Geological samples



Resistant composites

Material Science

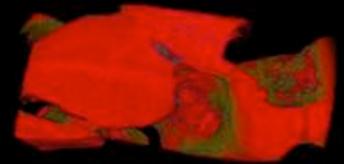
ECRP-2015

ECRP, CAMERA and DAV

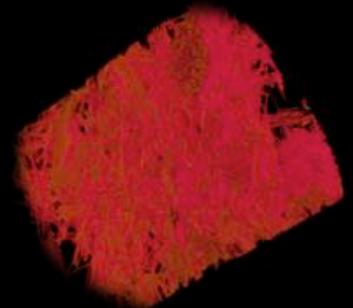
- Scidac 2012
 - Geological samples
 - Carbon sequestration
- Math Foundry 2013
 - MicroCT-imaged samples
 - Confocal and PS-OC
- CAMERA 2014
 - ASCR + BES



rock



bone

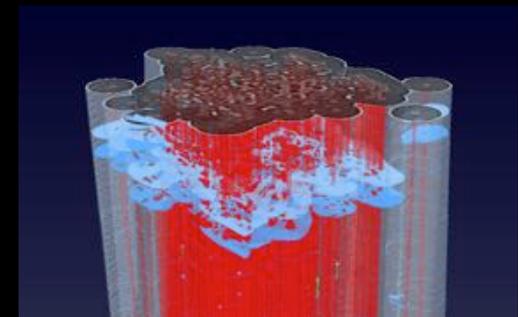
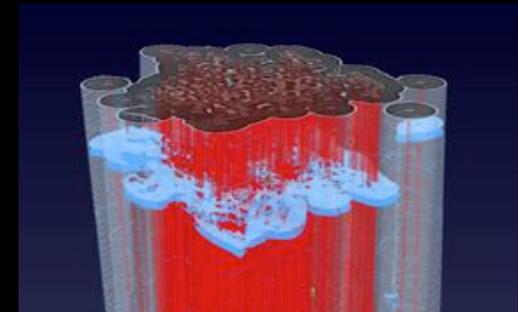
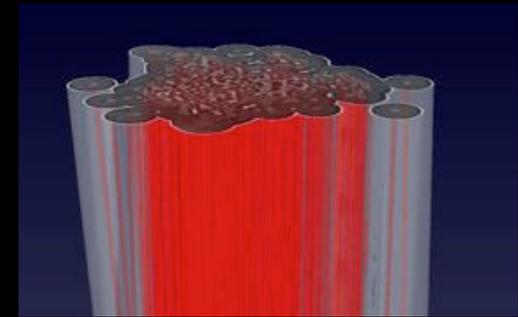
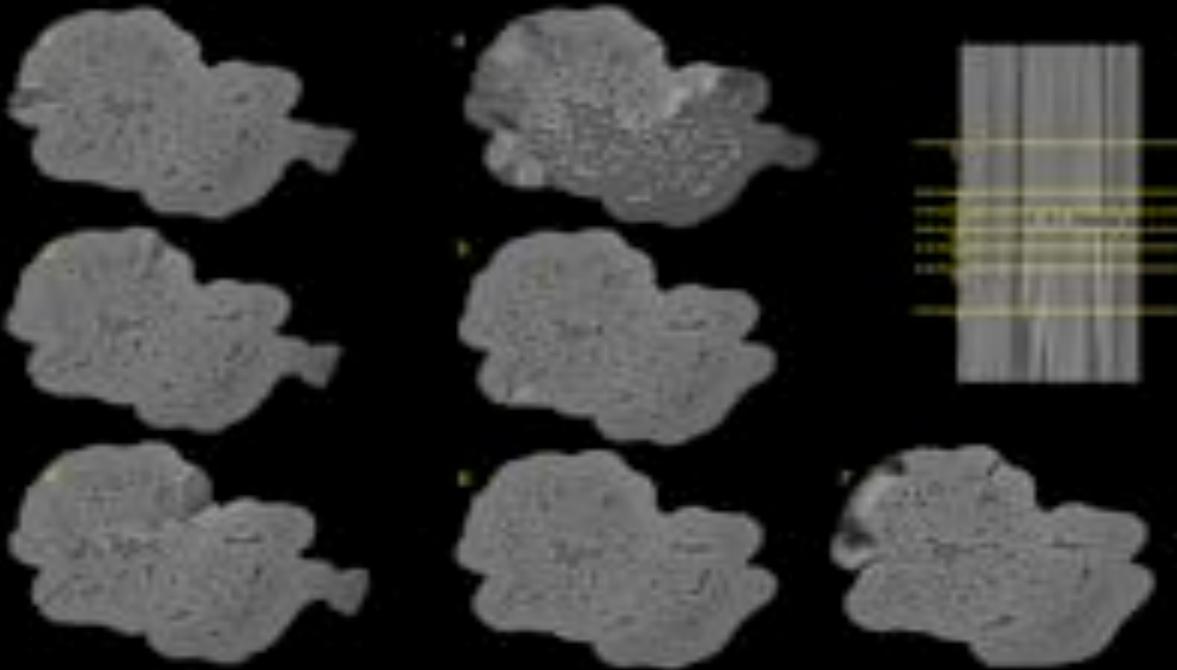


composite

The science question: material resilience sample (CMC) and instrument (microCT)

Pressure & temperature

- Detect cracks and fiber breaks from microCT images from ALS to **quantify** the robustness and **resilience** of new materials: no automated methods exist for this type of analysis;
- Constraints: (1) existing software tools incapable of meeting **throughput** requirements and scale to **full-resolution** of the experiment (raw~60GB) (2) unable to provide real-time feedback.



t

Micro-CT Pattern Recognition

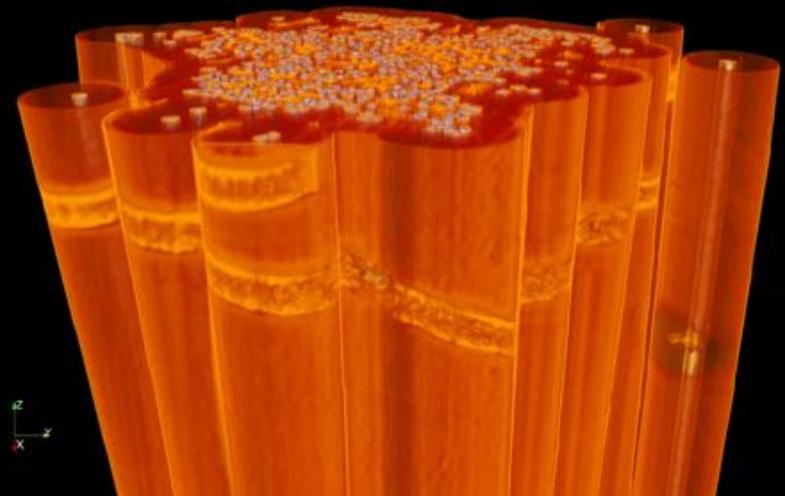
Problem: quantify **micro-structural damage** of ceramic matrix composites using time-resolved data for full exploration of the micro-tomography content;

Goal:

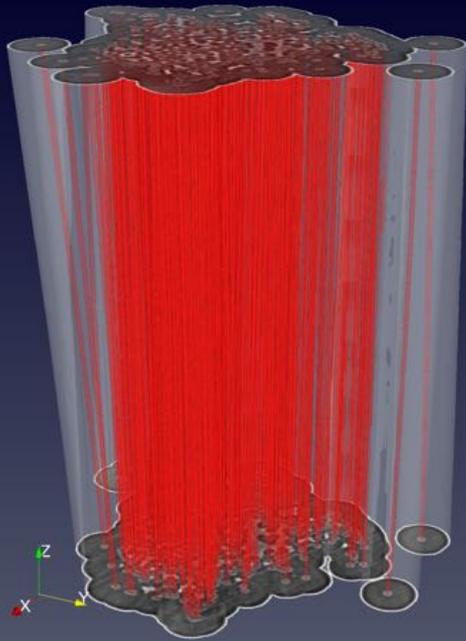
- Identify **material failure** and deformities from micro-CT, for example, to inspect fiber reinforced CMC, and dendrites permeating batteries;
- Real-time feedback about data collection and sample condition;

Approach:

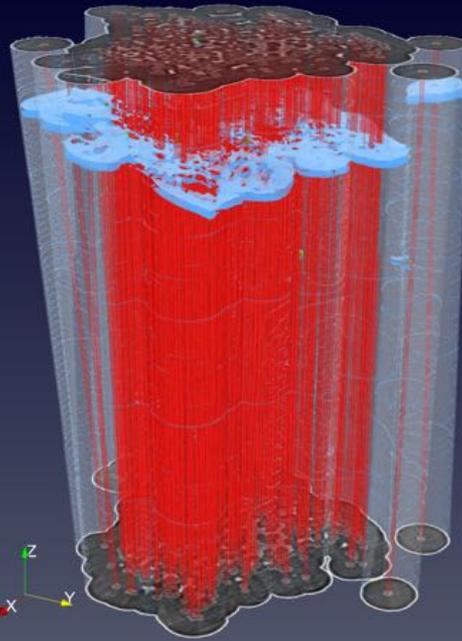
- Develop scalable pattern **recognition algorithms** to find defects from 3D images;
- Create **software tools** to better interface humans to instruments with high resolution high-throughput.



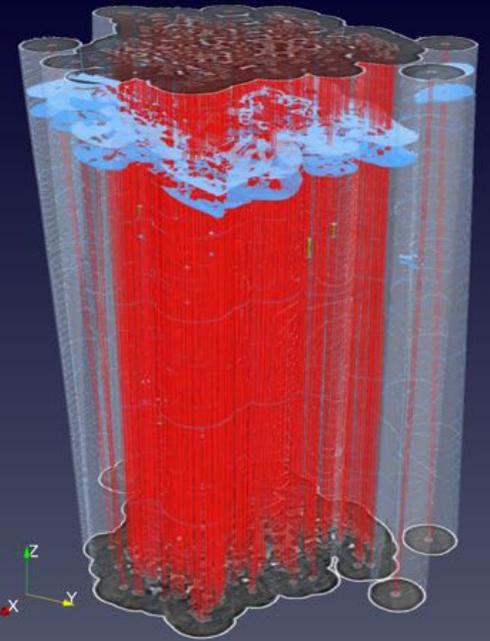
Deformation evolution



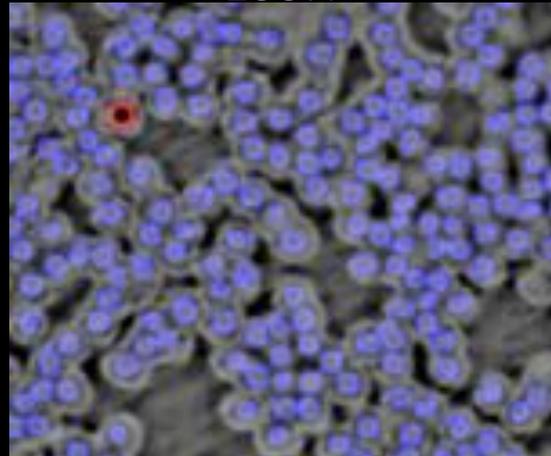
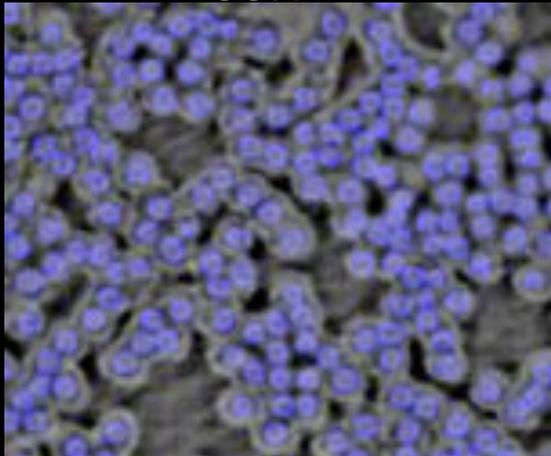
93N



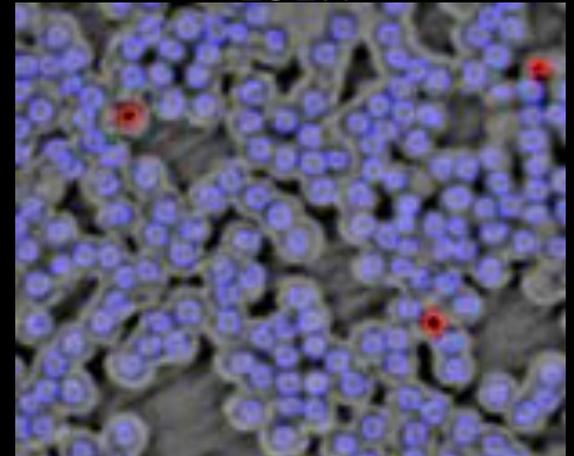
133N



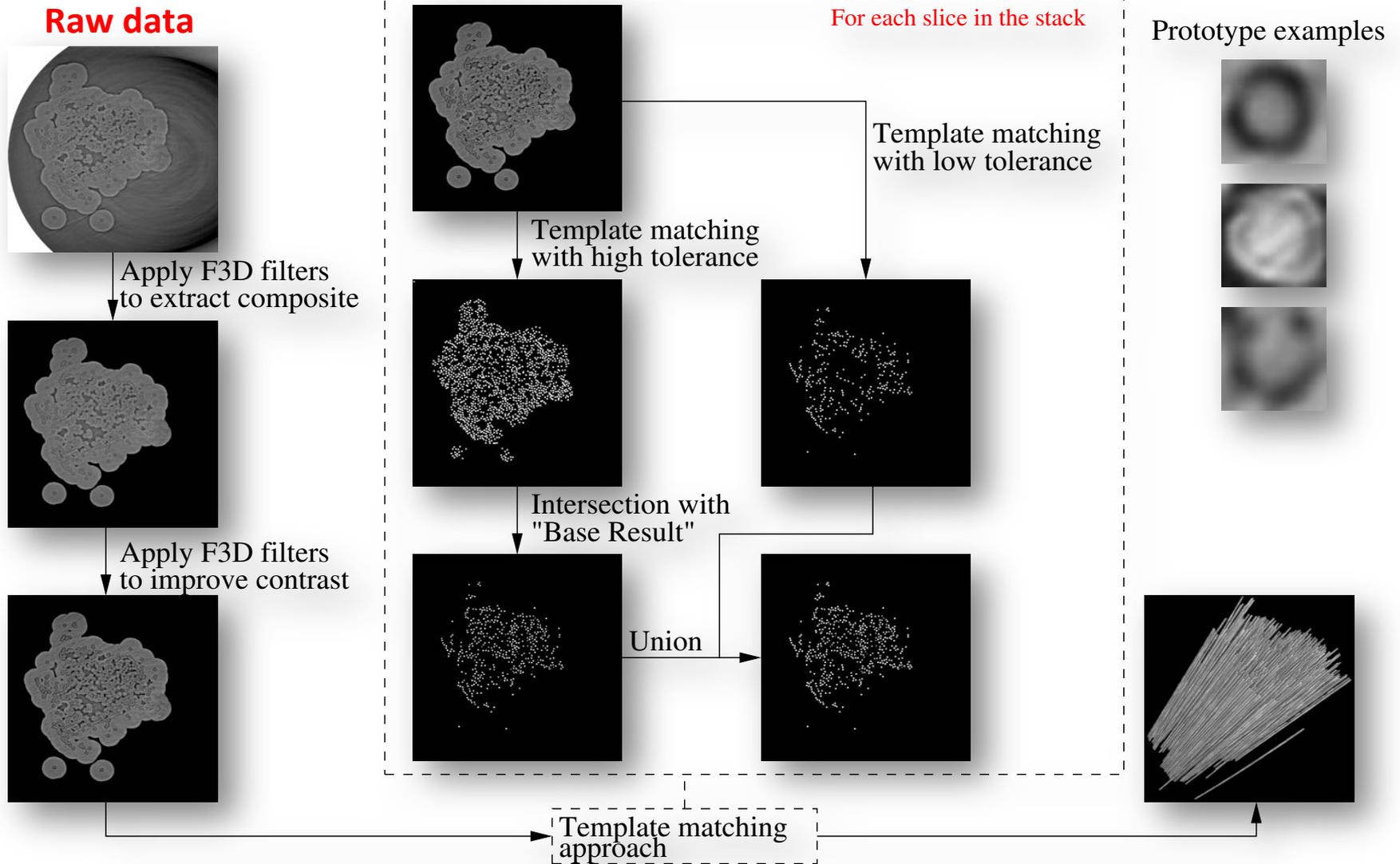
151N



Fracture Analysis of High-res Images



Identification of structures

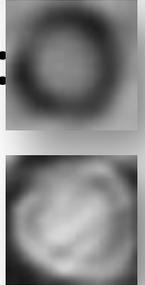


Template matching

Prototype examples

1) Similarity between prototypes and local regions:

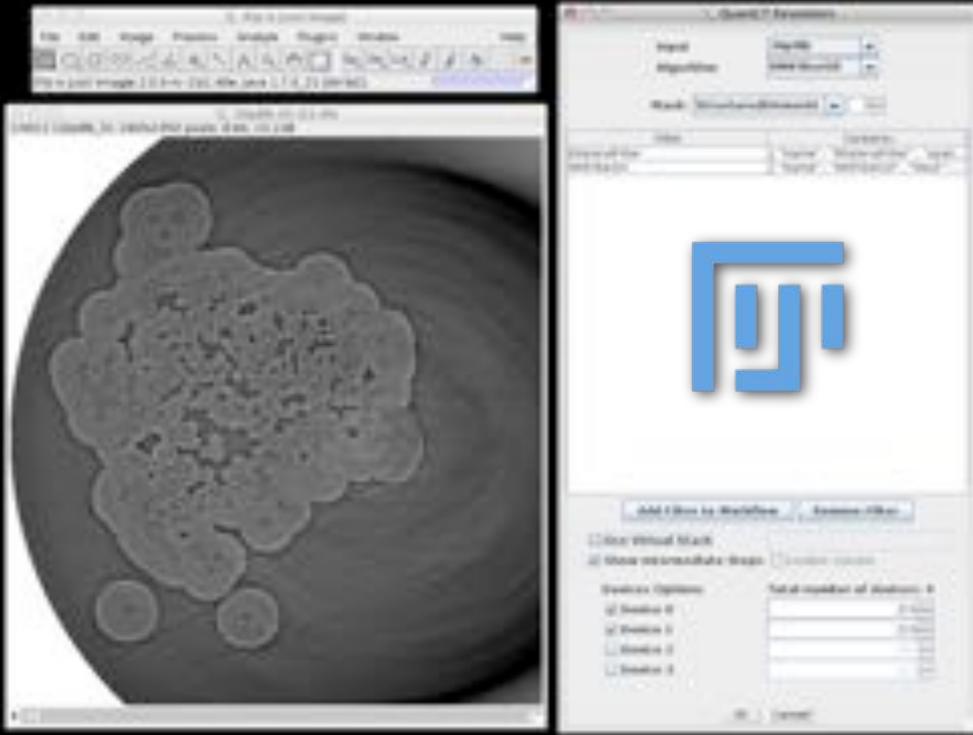
$$MSE(x, y) = \frac{1}{n} \sum_{i,j} [p(i, j) - f(x + i, y + j)]^2$$



2) Determine the best matches:

$$NCCC(x, y) = \frac{\sum_{i,j} p(i, j) - \bar{p}(i, j) \sum_{i,j} f(i, j) - \bar{f}(i, j)}{\left[\left(\sum_{i,j} p(i, j) - \bar{p}(i, j) \sum_{i,j} f(i, j) - \bar{f}(i, j) \right)^2 \right]^{1/2}}$$

Image processing at high-resolution



- Non-linear edge preserving filters
- Morphological operators with varying *strel*

F3D plugin

- **Accelerate** key image processing algorithms
- Enable segmentation and analysis of **high resolution image** datasets
- Requirement: **parallel-capable algorithms** to accommodate large data sizes and to allow real-time feedback

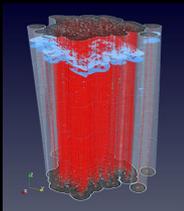
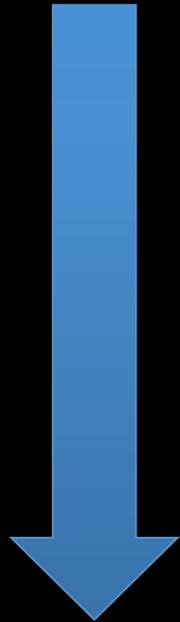
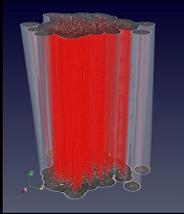
<https://github.com/CameraIA/F3D>

ECRP-2015

DOE Early Career Research Program

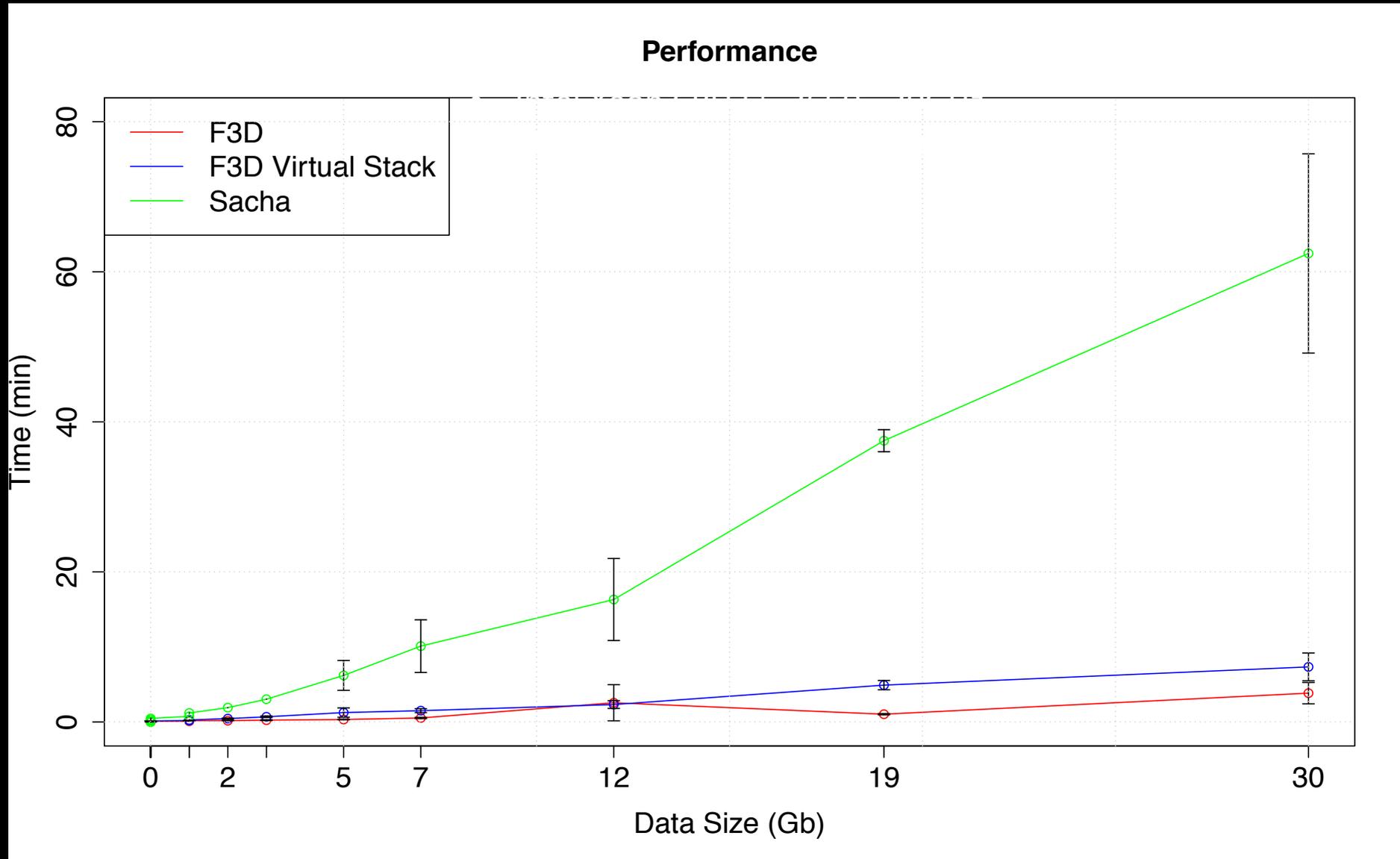


FC Quantitative results



Data	Precision	Recall
16N	0.9215	0.9586
93N	0.95	0.98
111N	0.9139	0.9625
120N	0.9379	0.9418
133N	0.9454	0.9753
144N	0.9483	0.984
151N	0.9454	0.9726
122N	0.9468	0.9746
87N	0.9442	0.971

Performance evaluation: comparison between proposed filter and only tool previously available in Fiji

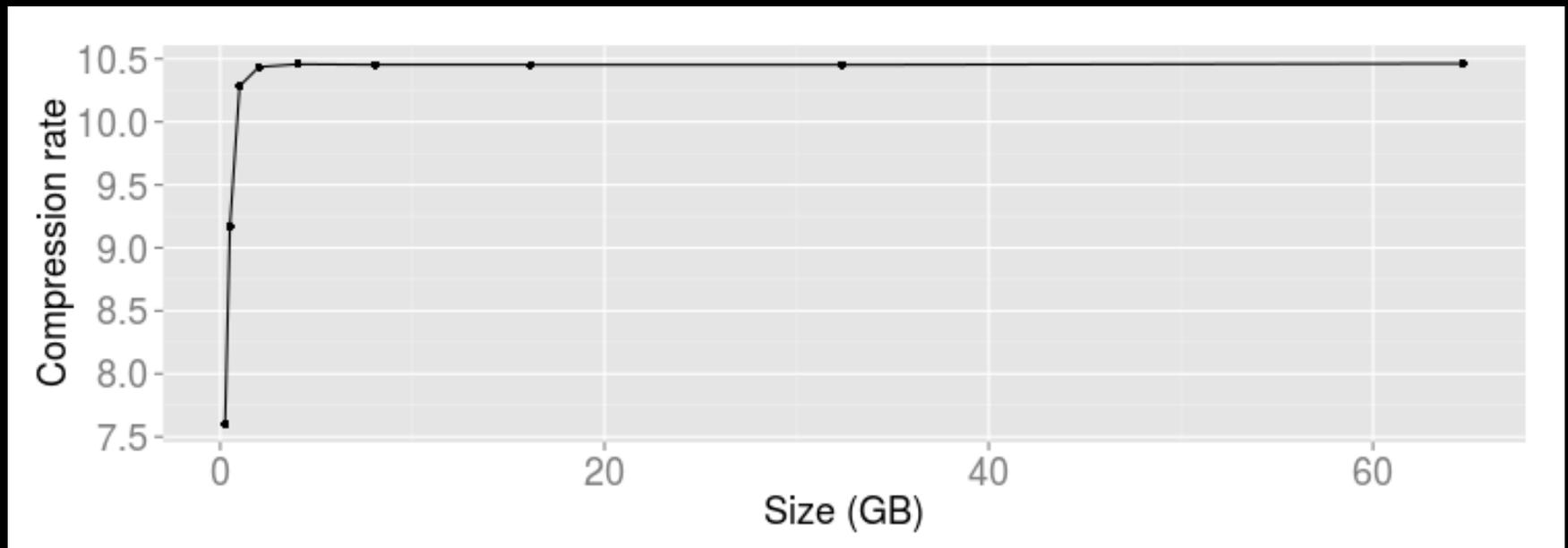


Terabyte-size image representation

- Problem:
 - Large datasets (originally 16GB per frame)
- Solution:
 - Multiresolution pyramids at four different scales stored as HDF5 chunked multi-dimensional arrays through Big-DataViewer;
 - Plugin originally offers interactive arbitrary virtual reslicing of multi-terabyte recordings, so that the user can inspect the experimental data efficiently;
 - Compress files and allow encapsulation of terabyte-size image datasets, including metadata, and optimized access to multiple scales of the data, both for visualization as well as for processing.
 - Other advantages of BigDataViewer formatting: a) increased computing performance, b) decreased cluttering of the experimental archives, and c) potential for parallel I/O.

Ref: T. Pietzsch, S. Saalfeld, S. Preibisch, and P. Tomancak. Bigdataviewer: visualization and processing for large image data sets. Nature Methods, 2015.

Testing files with different sizes

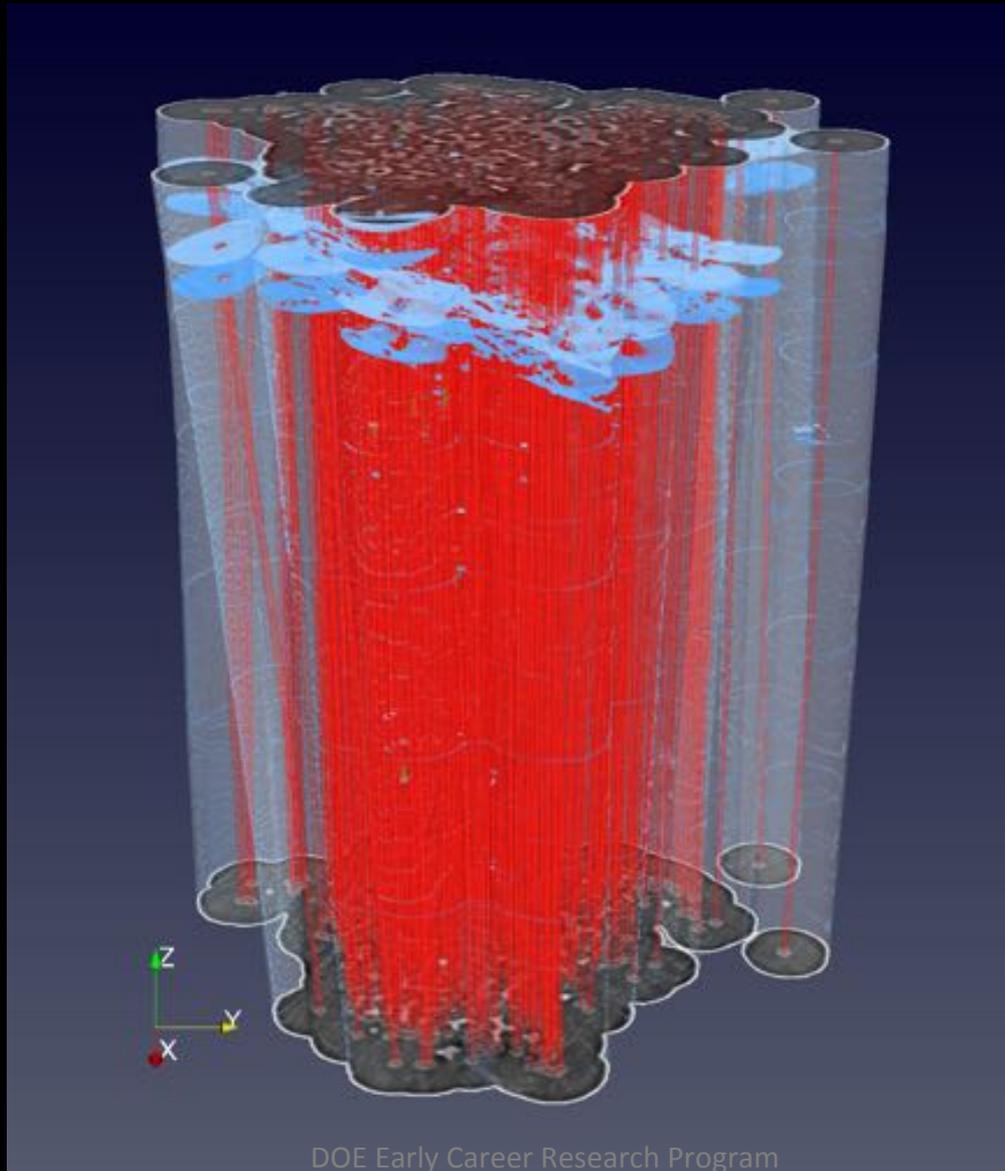


Scalability of the multi-dimensional representation using HDF5 with increasing data size.

Advanced technique: team work

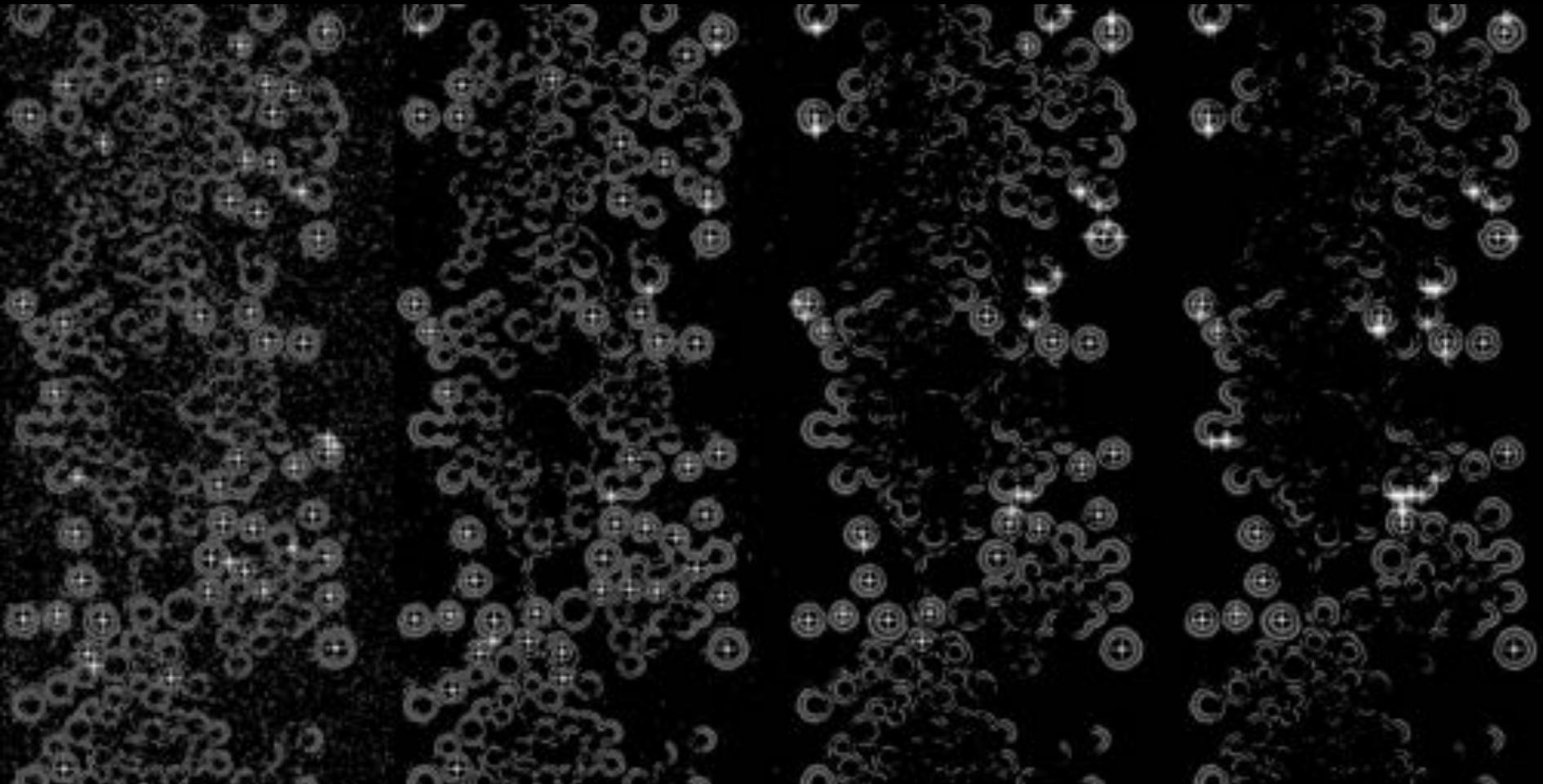


ECRP-2015



DOE Early Career Research Program

Inventing new codes for characterization of reinforced composites



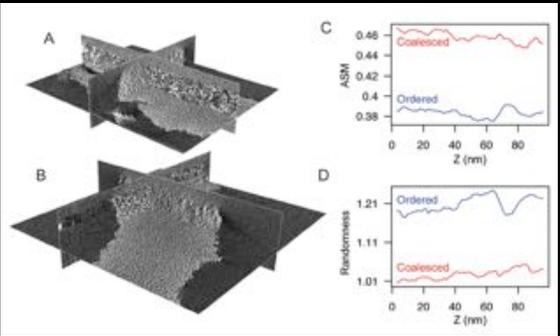
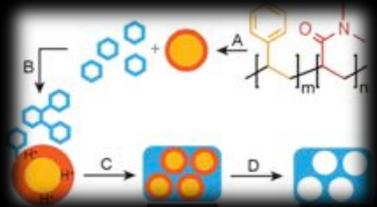
Scientific Achievement

- Analysis of thin films by using scanning transmission electron microscopy (STEM) tomography images in support of material architecture enhancement;
- Quantify pore structure evolution in order to control quality of fabricated films.

Significance and impact

- Results using porosimetry from STEM images corroborated in identification of fabrication conditions that led to the lowest ever dielectric constants for the needed films.
- Collaboration with Intel, LBL NCEM and Organic and Macromolecular Synthesis at the Molecular Foundry, and SLAC,

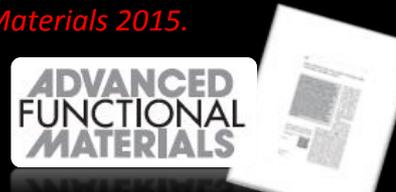
Image analysis for quality control of material architecture



Research details*

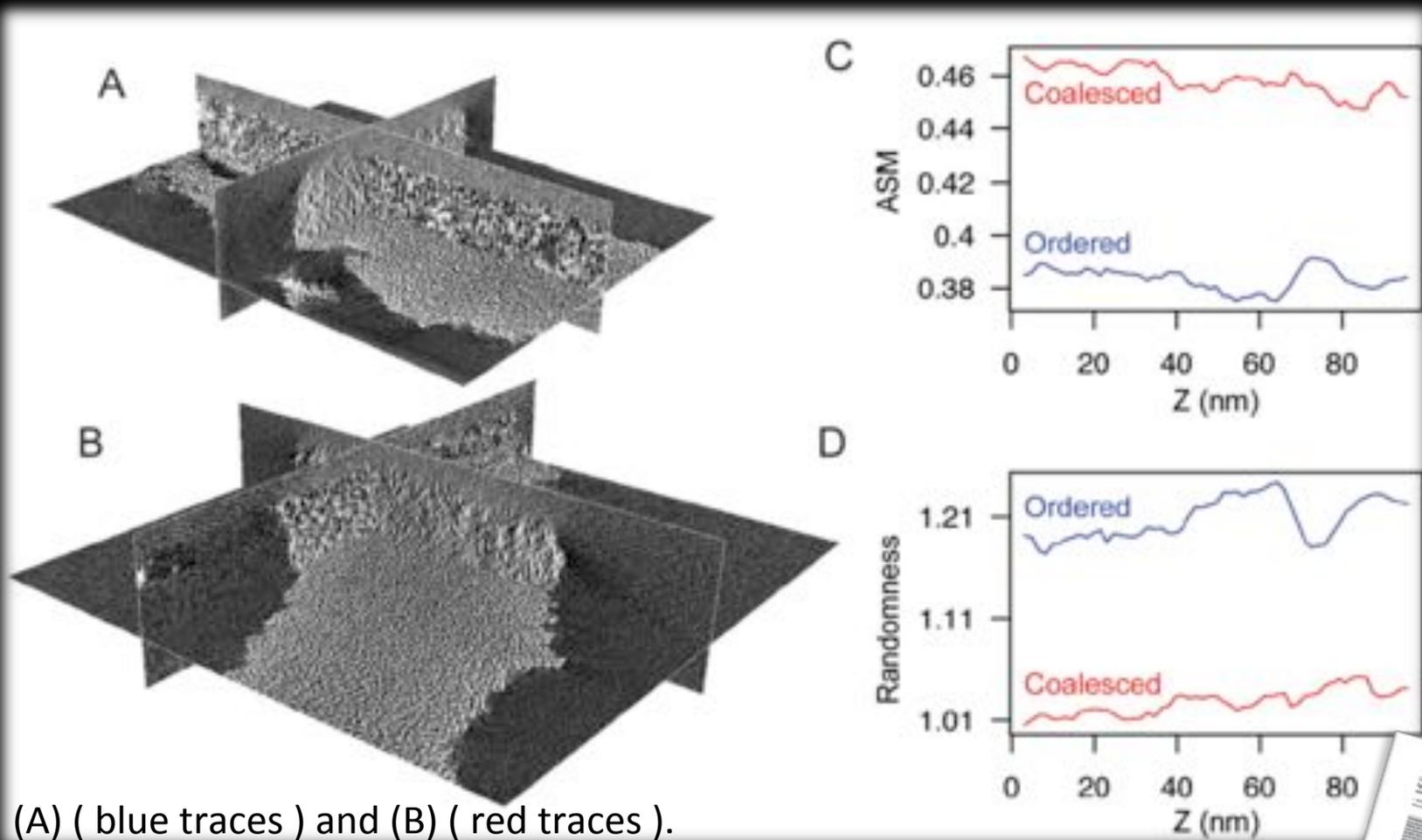
- Reported lowest ever dielectric constants for PMO matrix material, used in microelectronics;
- Texture analysis using second-order statistics of image intensity variations to measure film roughness;
- New tools adapted to 3D stacks for NCEM instruments;
- **New developments:** porosity analysis using new material architecture drivers (with T. Williams and B. Helms) and spectral analysis of catalytic processes (with K. Bustillo and P. Ercius).

Ref: *Wills et al, "Block Copolymer Packing Limits and Interfacial Reconfigurability in the Assembly of Periodic Mesoporous Organosilicas", Functional Materials 2015.*



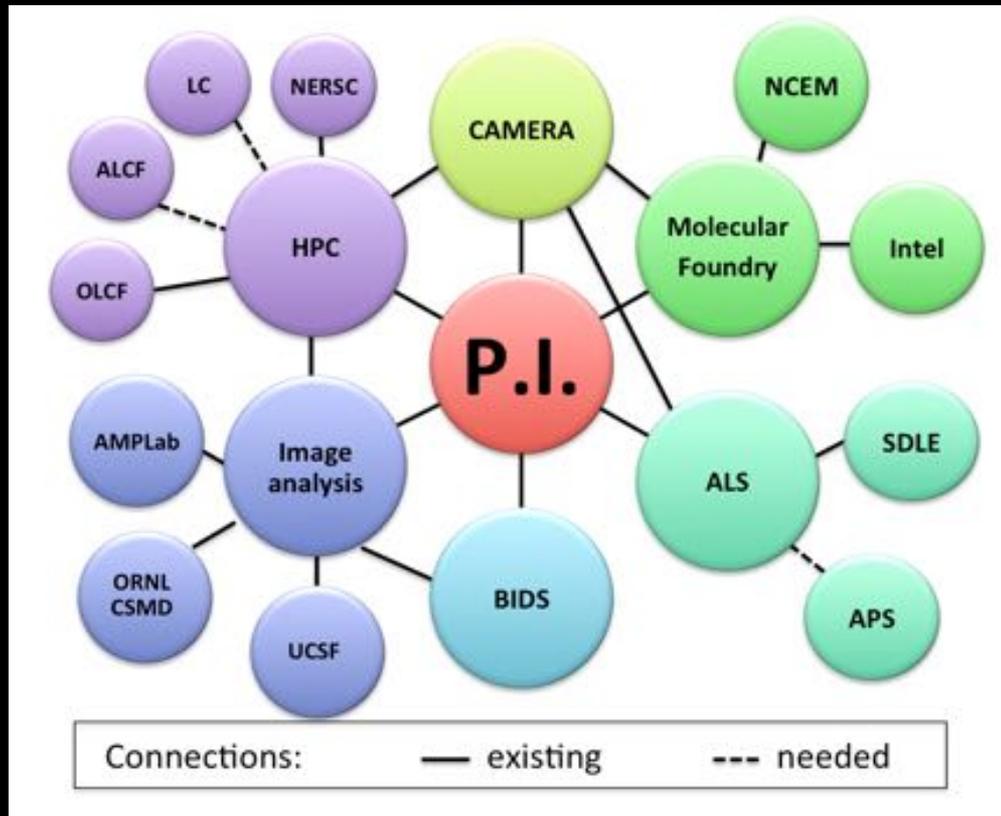
*Work was performed at LBNL by the CRD DAV and CAMERA. DAV is supported by ASCR and CAMERA jointly by ASCR and BES.

Feature design for STEM image data



Final remarks

- Scaling through partnerships:



Final remarks

- Algorithms in an exascale landscape
 - I/O awareness
 - Data reduction and in-situ analysis
 - Machine learning
 - Experimental/observational datasets
 - Digital twin