Applied Mathematics For Experimental Science

J.A. Sethian

Head, Mathematics Department LBNL
Simons Chair in Mathematics, Department of Mathematics, UC Berkeley

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Want to do two things:

Provide an overview of CAMERA

Discuss detectors, data, mathematics, and models.

The Center for Applied Mathematics for Energy Research Applications (CAMERA)

Mission: Build the applied mathematics that can accelerate scientific discovery at DOE experimental facilities

Execution: Coordinated team of applied mathematicians, beam scientists, computational chemists, computer scientists, materials scientists, statisticians, image and signal processors, …

Initial set of partners:

Advanced Light Source

Molecular Foundry

NCEM

Initial Support: LBNL LDRD

(P. Alivasatos, H. Simon, D. DePaolo, R. Falcone, D. Brown, J. Neaton, H.A. Padmore, M.J. Banda, K.A. Yelick)

Now: Joint ASCR-BES Pilot Project

U.S. DEPARTMENT OF ENERGY | Office of Science

(Steve Lee, P. Lee, B. Harrod, S. Binkley, H. Kung)

BERKELEY LAB

Overview of CAMERA

Build the advanced mathematics that can:

Extract information from murky data, and help interpret experimental results
Provide on-demand analysis as results are being generated
Steer experiment and suggest optimal solutions
Decrease turn-around time/save money: More experiments and more users
Extend the capabilities of existing and future experimental facilities

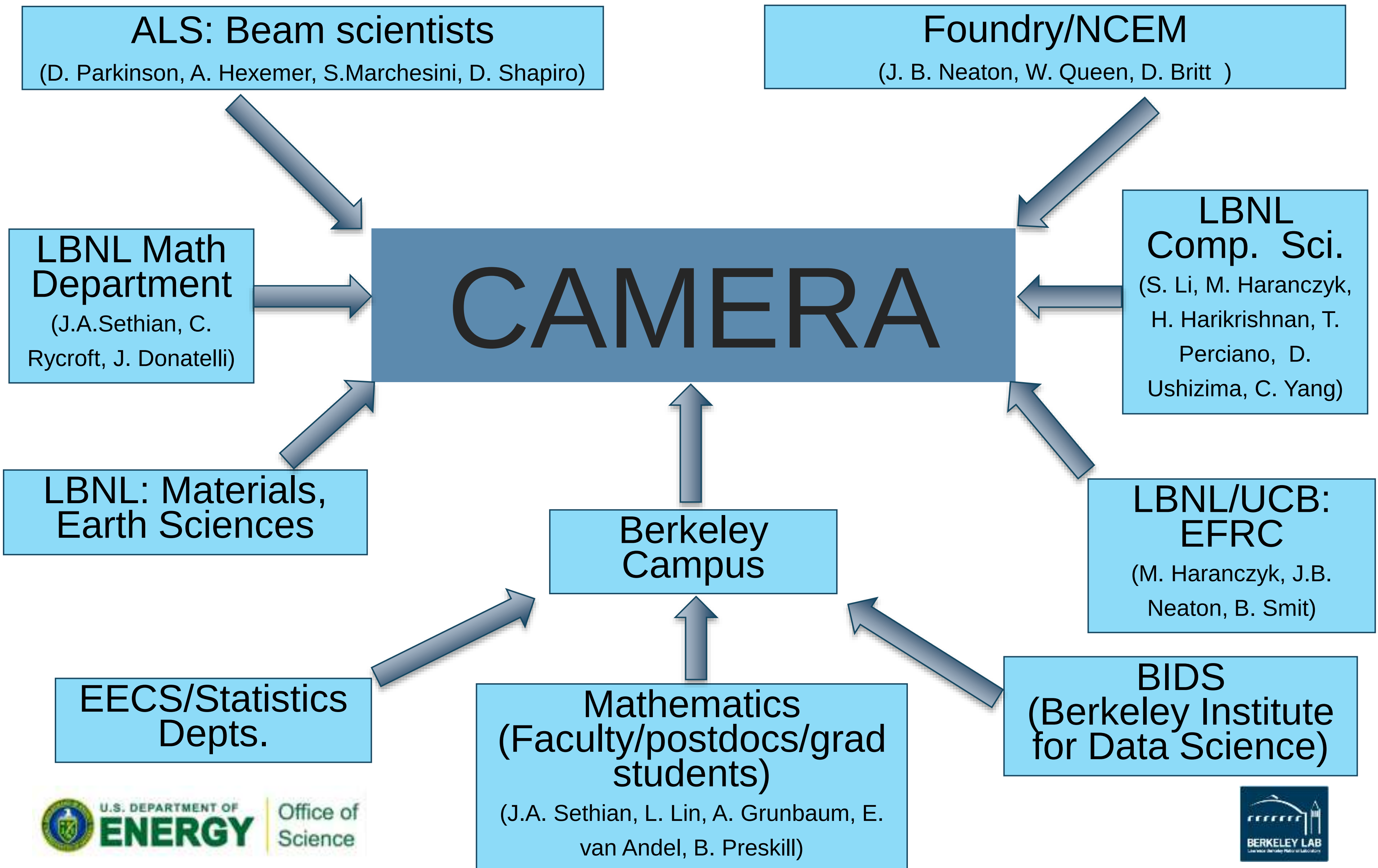
To do so, we need to:

Have experimental scientists/applied mathematicians work together
Develop common language
Build new mathematical models, invent algorithms, build prototype codes
Test on "shop floor", iterate until codes are solid and useful

Goal: Deliverables users can use (without becoming mathematicians):

Advanced mathematics embedded in useable software tools

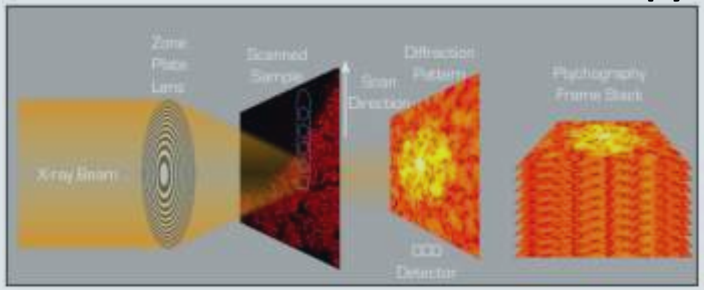
CAMERA: Organization

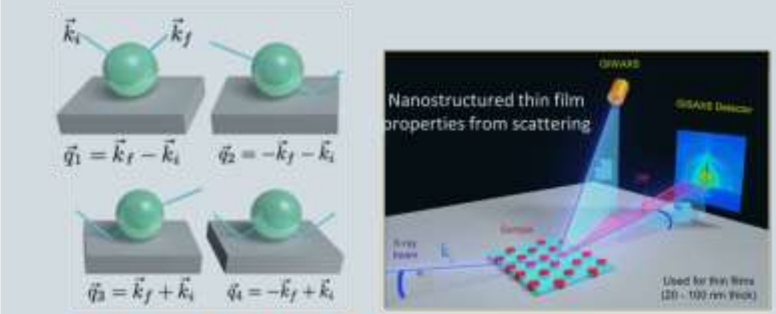


CAMERA: Original/Current Projects

Pytchography (ALS/MATH)

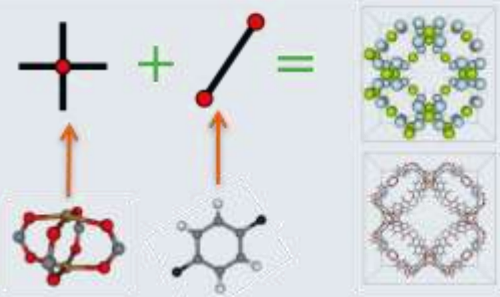
Coherent Diffraction with Microscopy



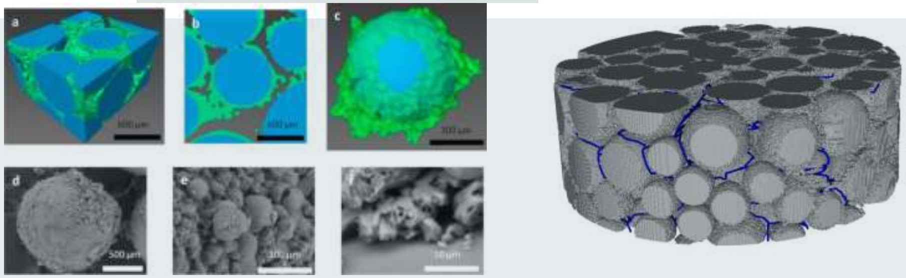


GISAXS (ALS/MATH)

Grazing incidence small angle x-ray scattering

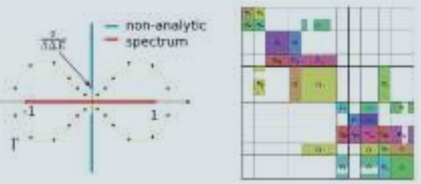


Automatic Image Analysis (ALS/MATH)



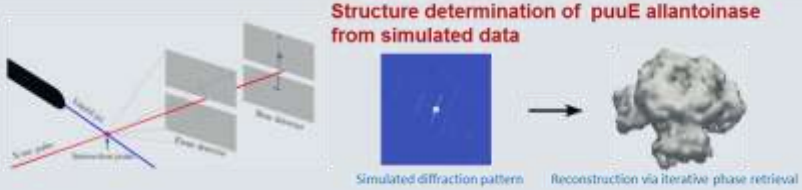
Designer Materials (Molecular Foundry/Math)

New methods for Density Functional Theory (MF/MATH)



$$H[\rho]\psi_i(x) = \left(-\frac{1}{2}\Delta + \int dx' \frac{m(x') + \rho(x')}{|x - x'|} + V_{xc}[\rho]\right)\psi_i(x) = \varepsilon_i \psi_i(x)$$

$$\rho(x) = 2\sum_{i=1}^{N/2} |\psi_i(x)|^2, \qquad \int dx\, \psi_i^*(x)\psi_j(x) = \delta_{ij}, \qquad \varepsilon_1 \le \varepsilon_2 \le \cdots$$



X-Ray nanocrystallography (ALS/MATH)

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

CAMERA: End Goal: Released Software: (camera.lbl.gov)

HipGISAXS: (http://camera.lbl.gov/software/hipgisaxs_software)

Flexible grazing-incidence small-angle X-ray scattering (GISAXS)
Distorted wave Born approximation
Speed: graphics processors and multicore processors.

PEXSI (http://camera.lbl.gov/software/pexsi)

Fast method for electronic structure calculation: Kohn-Sham DFT.
Can regularly handle systems with 10,000 to 100,000 electrons.
Achieves scalability on more than 10,000 processors..

SHARP-CAMERA (http://camera.lbl.gov/software/sharp_camera_download)

Multi-GPU Accelerated Ptychography Software .
Combines diffraction + microscopy + high performance GPUS
Advanced acceleration algorithms for convergence and analysis
Freely available, open environment for collaboration/customization

---

QUANT-CT (http://camera.lbl.gov/software/)

Image enhancement, filtering, segmentation and feature extraction
Currently on ALS beamline 8.3.2, multi GPU. Open source: FiJi plugin

---

Zeo++ (http://camera.lbl.gov/software/zeo/)

Analysis/assembly of crystalline porous materials.
Geometry-based analysis of structure/topology of material void space
Current users: EFRC Nanoporous Materials Genome Center, EFRC Materials Project, Bosch, SABIC and Samsung.

Outline of Talk

How did this start?

History and Motivation

Why mathematics?

Math/Data/Computing

What is being delivered?

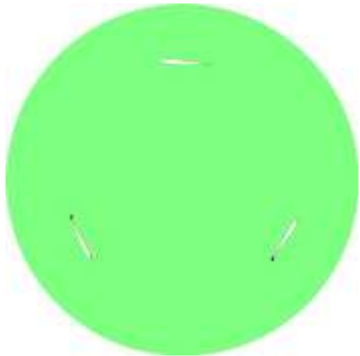
Four models for delivery

Where is it going?

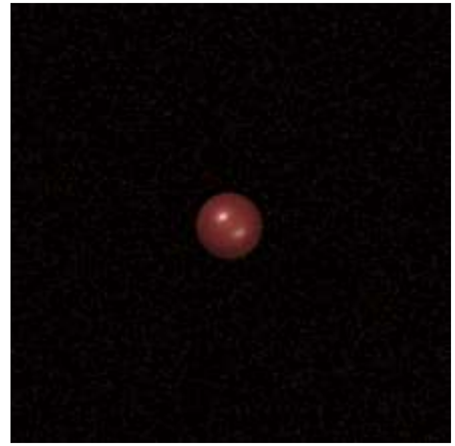
A DOE Resource

Background: LBNL/UCB Mathematics:

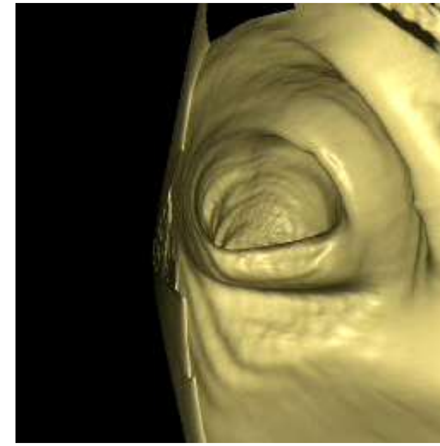
Long Standing DOE Program: (LBNL+UC Berkeley)



Vertical Axis Wind Turbines



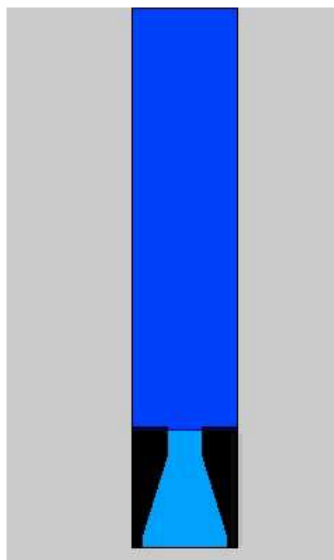
Cell Cluster Growth

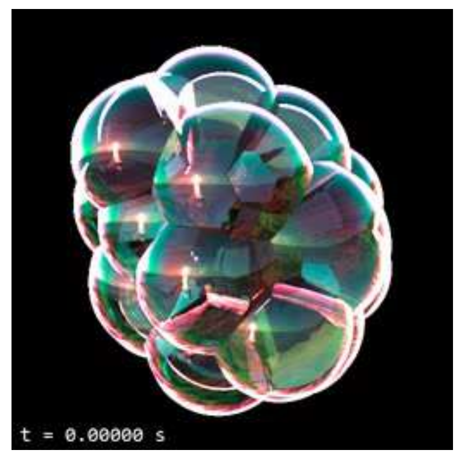


Virtual Colonoscopy

Draining in Coal Hoppers

Industrial Inkjet Printing





Industrial Foams

Example: Semiconductor Algorithms: Samsung, Intel, Motorola, Infineon, Synopsis…

A New Challenge from LBNL/Director

Look broadly at mathematical needs of Office of Science facilities, starting with the ALS, Molecular Foundry, NCEM, Joint BioEnergy Institute (JBEI), and future facilities

Question: How can applied mathematics help facilities do

More science

More efficiently

(users, materials, turn-around time…)?

DOE Facilities in 2025:
More Data, More Users, More Discovery

Experimental facilities will be transformed by high-resolution detectors, advanced mathematical analysis techniques, robotics, software automation, and programmable networks.

Detectors capable of generating terabit data streams.

Integration of experimental and computational facilities in real time, using programmable networks.

Computational tools for analysis, data reduction & feature extraction *in situ*, using advanced algorithms and special-purpose hardware.

Data management and sharing, with federated identity management and flexible access control.

Post-processing: reconstruction, inter-comparison, simulation, visualization.

Increase scientific throughput from robotics and automation software.

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Mathematics for accelerating the analysis of experimental data

Now → **Later**

Computational tools for analysis, data reduction & feature extraction *in situ*, using advanced algorithms and special-purpose hardware.

Post-processing: reconstruction, inter-comparison, simulation, visualization.

Mathematics for each can be quite different:

What is the minimum/fastest computational model/algorithm that gives (at least some) useful information?

What is the maximal amount of information you can get out of the data?

Can you quickly determine if data is useful, not useful, or in between?

Can data be measured, processed, organized and displayed to help understand/suggest further experiment?

Can you quickly do analysis and steer experiment to more optimal configurations or output?

Can data be transformed to initialize computational models, and output framed to complement experiment?

Why is this so interesting (and challenging?)

(a) Problems have not yet been “mathematicized”.

(b) No “equations of motion”

(c) Deep connections between the science and math

To tackle these problems requires new mathematics that bridges across mathematical disciplines.

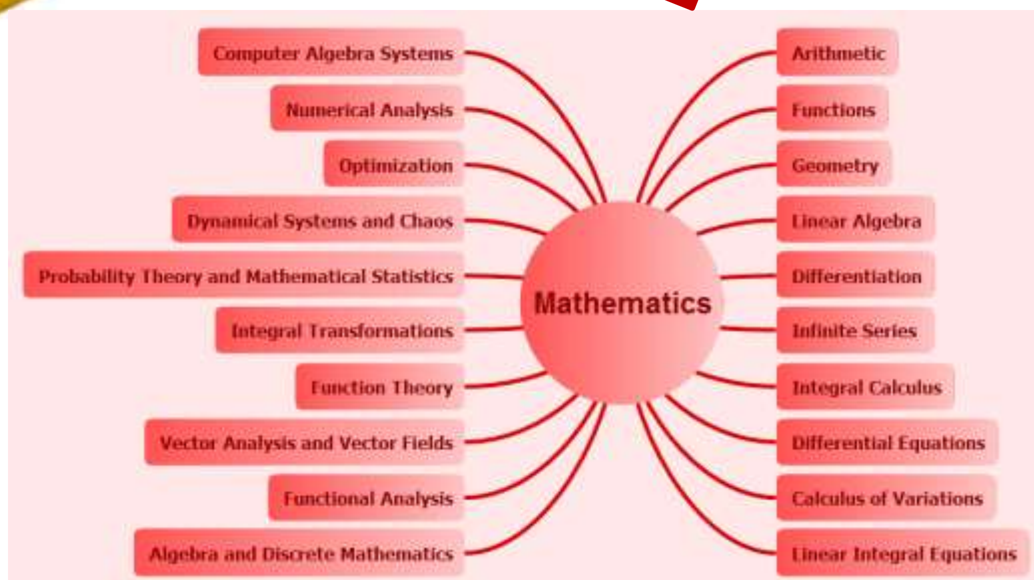
Fortunately, Applied Mathematics is Undergoing a Profound Transformation

Traditional walls between continuous math, discrete math, analysis, probability and statistics, topology, algebra, geometry …. **are all breaking down.**

Mathematics and (or versus!) "Big Data"

Data

Computers



Mathematics

Mathematics is what changes data into information

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Challenges are growing:

More data, more resolution
More complexity
Less obvious relational linking
More noise
More false signals

…

Mathematics is what changes data into information

Going to need mathematics more than ever…

CAMERA: Center for Applied Mathematics for Energy Research Applications

CAMERA

BERKELEY LAB

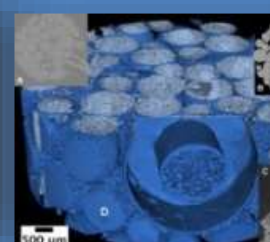
Today:	Tomorrow:	Critical need:	LBNL approach:	New math to:
Facilities data is time-consuming	More data. More quickly. High resolution.	algorithms and analysis for *understanding*	Focused teams of mathematicians/ domain scientists	Guide and optimize experiments

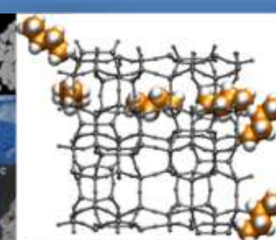
Goal: Build the applied mathematics that helps *transform experimental data into understanding*

Pilot Partners

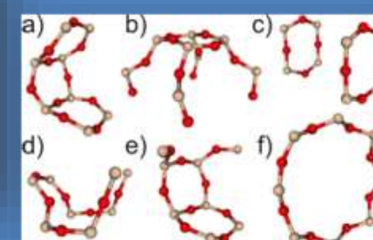
computing *structure* from imaging

analyzing samples and proposed new materials





designing new materials



Key: Leverage state-of-the-art mathematics

Spectral clustering · Maximum likelihood estimators · Graph theory · Machine learning · Mori-Zwanzig theory · Computational harmonic analysis · Discrete Galerkin methods · Clique analysis · PDE-based image segmentation · Statistical sampling · Discrete/continuous shape descriptors · Hamilton-Jacobi solvers · Voronoi methods · Representation theory · Bayesian analysis · Optimization methods

CAMERA

CAMERA: Personnel
Who is working on this?

Advanced Light Source (ALS):

A. Hexemer (Beam Scientist/GISAXS)
S. Marchesini (Ptychography)
D. Parkinson (Beamline Scientist, Hard X-ray tomography)
D. Shapiro (Beamline scientist)

Molecular Foundry

D. Britt (Organic and Macromolecular Synthesis)
J. Neaton (Electronic Structure)
W. Queen (Inorganic Nanostructures)

National Center for Electron Microscopy (NCEM)

P. Ercius (Scanning transmission electron microscope)

Computational Research Division (CRD)

M. Haranczyk (Materials Design)
X. Li (GISAXS/)
L. Lin (Electronic Structure)
R. Martin (Materials Design)
C. Yang (Electronic Structure)
D. Ushizima (Image Analysis)
T. Perciano (Image Analysis)
H. Krishnan (Image Analysis/HPC)

CRD Mathematics Department:

J. Donatelli (X-Ray Nanocrystallography)
C. Rycroft (Optimal Chemical Design)
J.A. Sethian (Director)

- **Opportunity: Steady stream of new Berkeley faculty/postdocs/grad students**

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

What Does CAMERA Deliver?

CAMERA

(2) Remote browsers executing code locally running at facilities.

(1) Codes that run locally on computers embedded at facilities.

(4) Downloaded and run remotely.

(3) Codes remotely run on data downloaded from facilities to supercomputer centers

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

"Long-distance Delivery": SPOT-Suite

Led by C. Tull, LBNL

CAMERA

Towards an End-to-End Solution for Light Source Data, Analysis, & Simulation

- **Combining…**
  - scalable software systems.
  - HPC/HTC/network resources.
  - advanced algorithms & analysis.
  - advanced simulation.
  - realtime feedback.
- **Multi-division team: CRD, ALS, Math, CAMERA, ESNet, MSD, & NERSC.**
- **Extending to include SAXS, µDiff, and Ptychography beamlines. Focus on in-situ, time-resolved experiments, new algorithms, data sharing & collaboration.**
- **SC14 Demos include LCLS, APS, NSLS datasets.**

Research into generalizable real time workflows and metadata already yielding valuable insight for photon scientists.

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Who is using CAMERA's deliverables?

CAMERA

Ptychography: CXRO/SEMETEC, LLNL/NASA, UI Chicago, UC San Diego, UC Davis, UCB, McMaster, Stanford. ALS, BNL, F. Maia, Uppsala, BYU, multiple workshops, tutorials

QuantCT: Advanced Light Source Users, available on "shop floor". Downloadable FiJi Plugin (world-wide user base).

PEXSI: Accelerated Kohn-Sham Density Functional Algorithms Embedded in SIESTA (**Spanish Initiative for Electronic Simulations with Thousands of Atoms): Next:** CP2K

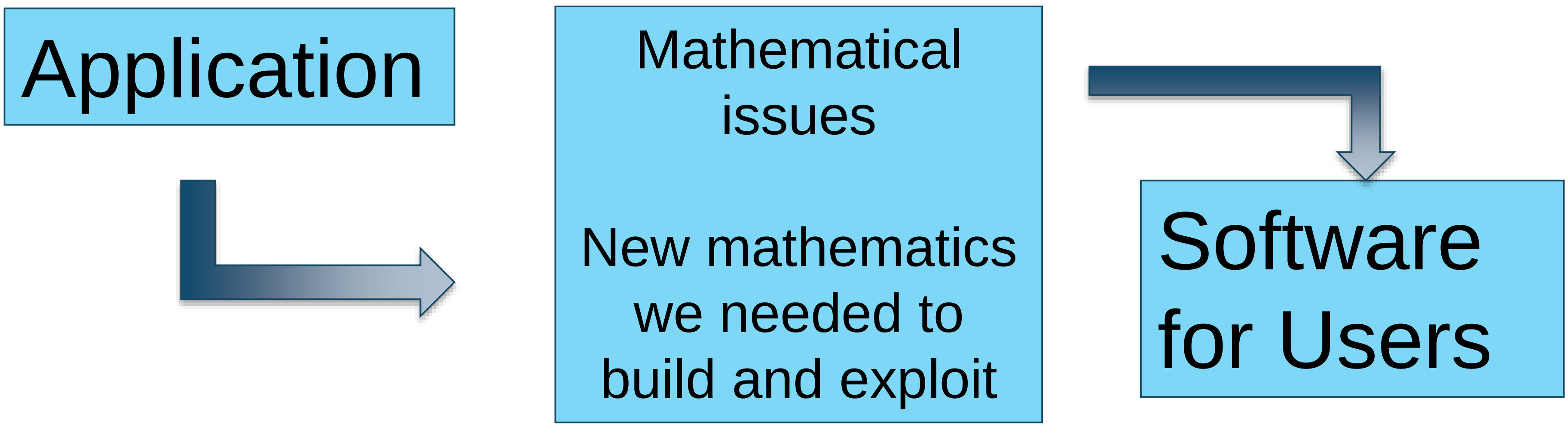
Zeo++: Open-source package - www.zeoplusplus.org. Zeo++ : a default tool for two BES Materials Genome Centers (Nanoporous Materials Genome Center (Minnesota) and Center for Functional Electronic Materials (LBNL)) and LBNL EFRC for Gas Separations Roughly 200 registered users world-wide in both academia/industry (e.g. Bosch, Samsung)

HipGISAXS: ORNL, ANL, ALS, Molecular Foundry, numerous universities, …

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

AN OVERVIEW OF SOME OF THE WORK UNDERWAY



(Describe problem, emphasize new mathematics, describe deliverables)

SHARP

(Scalable Heterogeneous Adaptive Robust Ptychography)

Fast scalable methods for ptychographic reconstructions

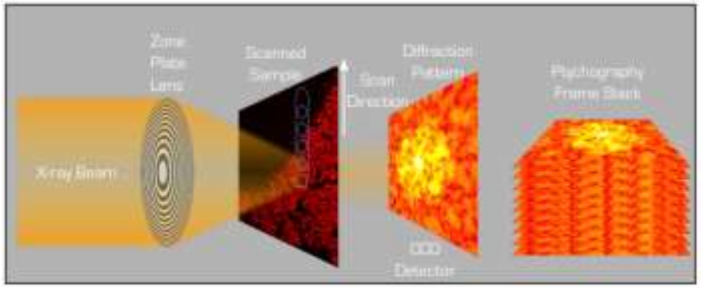
S. Marchesini, D. Shapiro (Advanced Light Source)
H. Krishnan (LBNL Computing Sciences)
F. Maia (LBL/Uppsala)
H-T Wu (LBNL/Stanford, now Toronto)

CAMERA: Ptychographic Imaging

CAMERA

S. Marchesini, D. Shapiro, H. H. Krishnan (LBL) F. Maia (LBL/Uppsala), H-T. Wu (Stanford))

Combine Coherent Diffraction with Microscopy

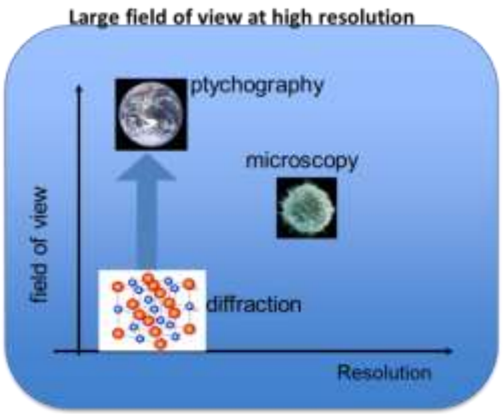


Fundamental idea: combine:

- High precision scanning microscope with
- High resolution diffraction measurements.
- Replace single detector with 2D CCD array.
- Measure intensity distribution at many scattering angles

Each recorded diffraction pattern:

- contains short-spatial Fourier frequency information
- only intensity is measured: need phase for reconstruction.
- phase retrieval comes from recording multiple diffraction patterns from same region of object.



Pytchography:

- uses a small step size relative to illumination geometry to scan sample.
- diffraction measurements from neighboring regions related through this geometry
- Thus, phase-less information is replaced with a redundant set of measurements.

Lots of ptychographic equipment/codes throughout DOE, universities, world-wide

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

CAMERA: Ptychography Mathematical and Algorithmic Issues

When does it (not) work?

(no convergence proof yet available for method)

Existing algorithms may have trouble converging on large data sets:

(iterative methods intrinsically operate by interchanging information between nearest neighbor frames (diffraction patterns) at each step, so it might take many iterations for frames far apart to communicate.)

Effects of noise and physical uncertainties:

(how do reconstruction algorithms perform with uncertainties in photon statistics, lens perturbations, illumination positions, incoherent measurements, detector response and discretization, time fluctuations, etc.)

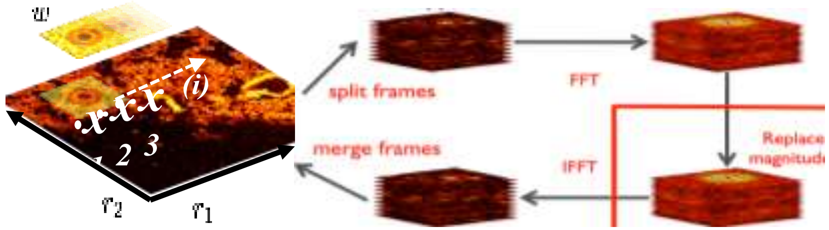
What is the best lens and illumination scheme for arbitrary specimens?

(given a detector, with a limited rate, dynamic range and response function, what is the best scheme to encode and extract more information per detector channel?)

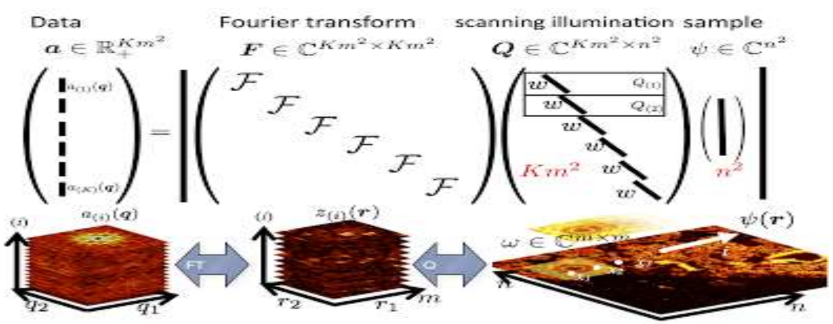
Phase retrieval in high dimensional space

Tackling Efficiency: Challenges with basic alternating projection algorithm:

Poor scaling:
long range interactions among frames decay exponentially with distance.

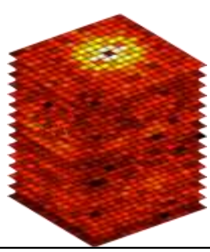


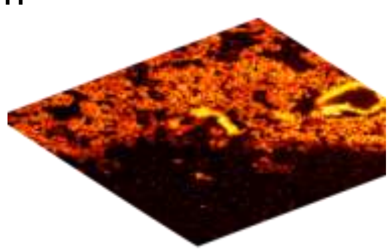
Poor initial guess:
can significantly delay convergence.



Ultimately, an overdetermined problem
in high dimensional space.

Short-time Fourier Transform





Large dimensional data

Low dimensional space

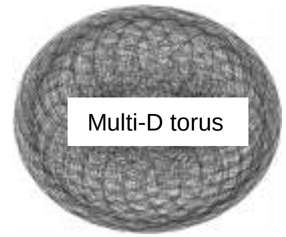
How can we speed this up?

Graph Theory/Graph Laplacians

Building a better starting guess:

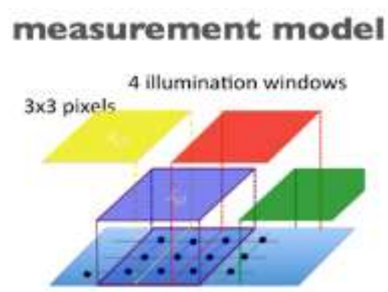
(1) View every pixel of every frame as a dimension. Each data point lives on a torus (complex plane)

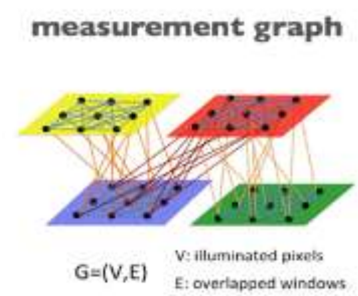
Diffraction data manifold



Approximate torus with ball

(2) Build “relationship network RN: a graph (V,E) that relates each frame to its neighbors.





(3) Construct Graph Laplacian of RN: defined as difference between the degree matrix D and the adjacency matrix A: GL = D - A

(4) The largest eigenvector of the Connection graph provides the most aligned phases encoding the (approximate) data topology.

This provides a strong starting guess.

Released Code: SHARP: Scalable Ptychography Solver

Code: Open source, downloadable package

☒release ☑prototype ☐under way/testing

•Scalable code, (☒source package, ☒remote interface, ☐ web interface, ☐ API).
- ☐ real time feedback by reducing latency
- ☑ 80x speedup with algorithms
- ☒ 30x speedup with GPUs
- ☒ >16x speedup with distributed GPU

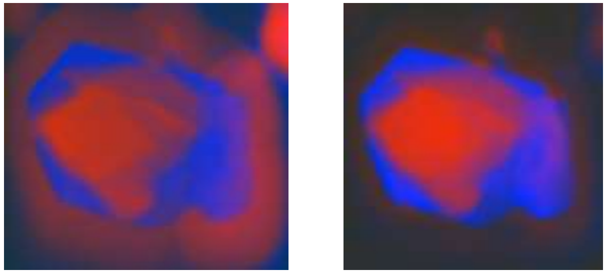
- ☒ Optimal Network fabric design for throughput
- ☒ Optimal lens design for SNR

- ☑ Iterative tomography (☐ network/bandwidth optimized)
- ☑ Chemical mapping (robust PCA/SVD)
- ☑ Dynamics

"Compute design"

SHARP real time specs:
- 3D torus p2p fabric
- CCD/RDMA streaming
- instrument calibration

Intercalation Battery Research: Mechanisms in Lithium Ion Phosphate ALS BL 5.3.2 (Nat. Phot. /in press)



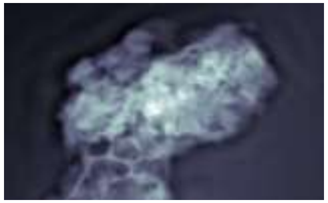
Partners:

CXRO/SEMETEC, LLNL/NASA, UI Chicago, UC San Diego, UC Davis, UCB, McMaster, Stanford. ALS, BNL, F. Maia, Uppsala, BYU

Software presentations: Ptycho 2013, FIO/LS, SIAM IM14, MSPPR, XRM, Coherence 14

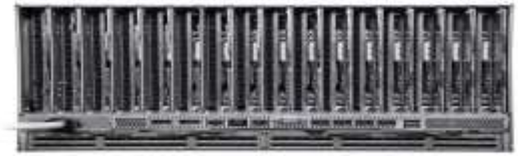
Software tutorials: Coming: SSRL/CAMERA xx/2014, CAMERA/ALS/BNL AUG 2014 CAMERA/ALS/APS Sep 10/14, COHERENCE, XRM , SIAMIM, FIO/LS, RACIR summer school, ALS Users workshop

Toward real-time feedback



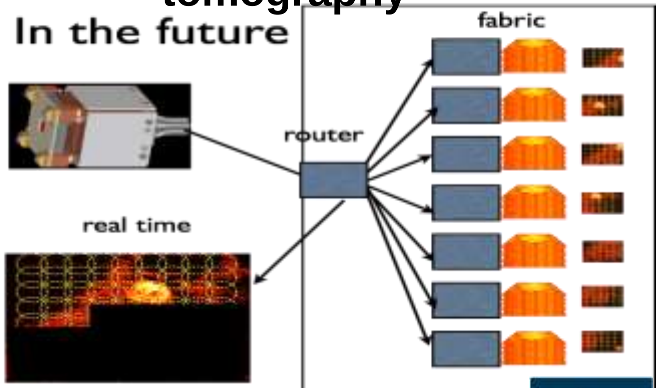
Scan: 10 micron^2, 10 nm resolution, 60x60 (1024) frames/minute.
Processing: 60x60 (1024^2) frames/minute

Currently
the user interface starts processing at the end of a full scan. (1 minute each)



High bandwidth 3D torus p2p For hyperspectral ptycho tomography

In the future
low Latency (<5 ms) feedback by streaming detector frames on distributed direct memory access fabric.



Real time enables smart self-calibrating, auto-tuning feedback of the microscope control system.

Two Weeks Ago: Finished a prototype "Real Time" version-code directly off of CCD

QuantCT

Automatic image analysis tools for micro-CT

D. Ushizima, D. Morozov, H. Krishnan, T. Perciano (LBNL Computing Sciences)
D. Parkinson (Advanced Light Source)

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

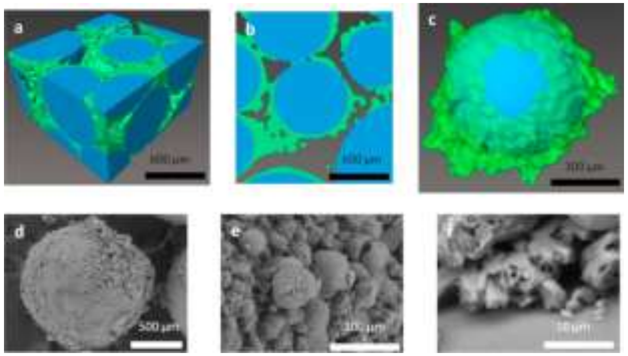
CAMERA: Quantitative Image Analysis of Micro-CT Samples

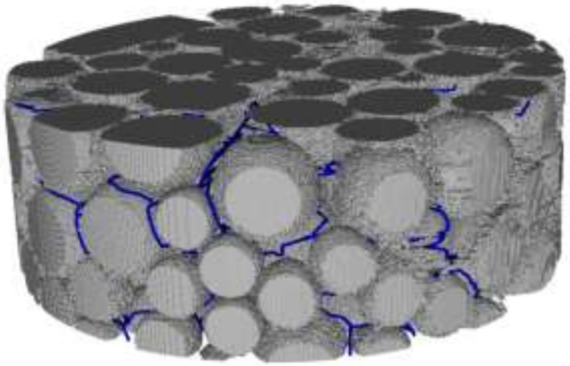
CAMERA

Goal: Develop algorithms for 3D/4D quantitative analysis of experiments, addressing challenges posed by noise, artifacts, sheer size, and heterogeneous materials.

Analyze structure: porosity, pathways, interior voids, ...

- Application: High-resolution synchrotron-based X-ray absorption microtomography.
- Suitability of materials and biomineralization processes for carbon sequestration.
- Acquire projection views at equi-spaced angles: produce 2D cross-sections.
- Gray level value of image voxels reflects x-ray attenuation and density.
- Compute pathways through materials:





Imaging Pipeline Requires:

- Filtering: remove noise, sharpen contrasts **(bi-lateral and non-linear filters)**
- Segmentation to isolate, and extract shapes from images **(PDE-VIIM methods)**
- Feature detection/analysis **(Reeb graphs, topological analysis, channel detection)**

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

CAMERA

QuantCT: Timeline of Mathematics/ Algorithm Development

BERKELEY LAB

2011 — 2014

Filtering of microCT

Gaussian

Median

Bilateral

Anisotropic diffusion

Non-linear tensor PDE

Segmentation of (near) homogeneous regions

Thresholding (local/global)

Variational Level Set Methods

Fast Marching Methods

Statistical Region Merging

Voronoi Implicit Interface

Analysis of microstructures

Porosity

Intensity descriptors

Topological descriptors

- Pore network
- Max Flow curves
- Slope of max flow
- Persistent pockets

Get rid of artifacts

Get structure from clutter

Get parameters from grains

DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

QuantCT: B. PDE-based Automatic Segmentation and Extraction

CAMERA

BERKELEY LAB

(1) **Mumford-Shah functional** for image segmentation of two phases

(index i indicates separate phases, Find interface **Γ** to minimize E)

$$E(\Gamma, I_1, I_2) = \int_A (I(x,y) - I_1)^2 d\mathbf{x} + \int_B (I(x,y) - I_2)^2 d\mathbf{x} + \mu \int_\Gamma g(\Gamma(s)) ds$$

(2) **Becomes PDE transport method using level set methodology:**

$$\phi_t + F|\nabla \phi| = 0, \text{ where } F = [((I - I_1)^2 + ((I - I_2)^2) - \mu \nabla \cdot (g \nabla \phi / |\nabla \phi|)]$$

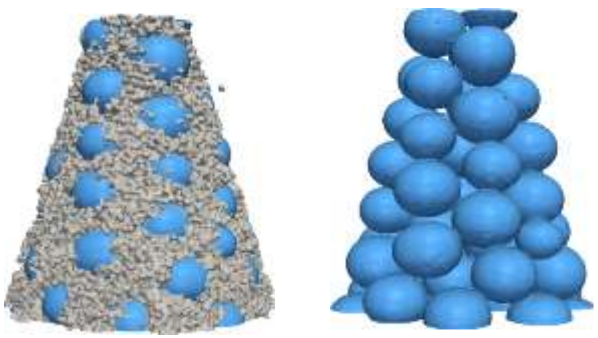
(3) **New approach: Extend the Mumford-Shah energy**

functional to multi-phase multi-interface

Voronoi Implicit Interface Method (VIIM)

$$F_i = [((I - I_i)^2 - \mu \nabla \cdot (g \nabla \phi / |\nabla \phi|)]$$

(combination implicit embedding plus dual Eikonal Voronoi reconstruction)



(4) **Allows simultaneous extraction of multiple structures in 3D.**

U.S. DEPARTMENT OF ENERGY | Office of Science

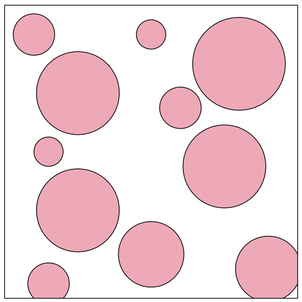
Calcite precipitation: ”pore clogging”

BERKELEY LAB

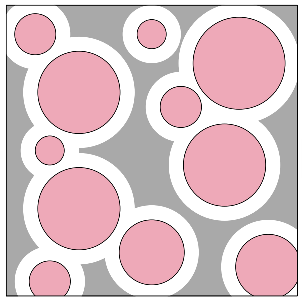
QuantCT: C. Determination of Connectivity and Channel Pathways

CAMERA

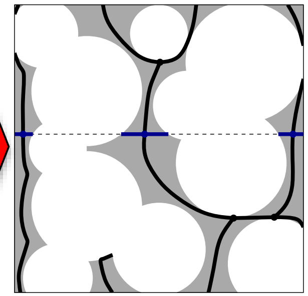
Augmented Topological Descriptors: Max Flow Graphs and Persistence Diagrams



Void Space

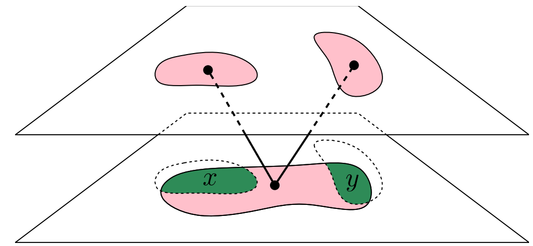


Wide pathways

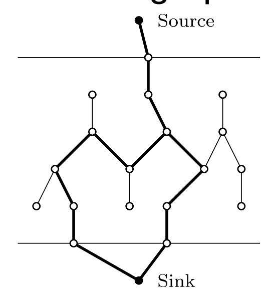


Reeb graph

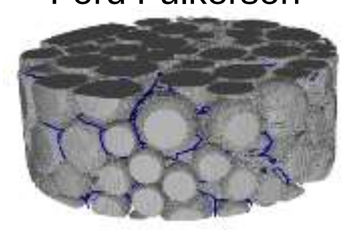
Edge capacities



Flow graph



Ford Fulkerson

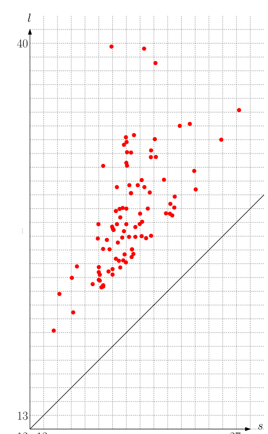


Max-Flow:

- Reeb graph: Evolution of level sets of function on manifold.
- Use to detect pathways for particle of size a
- Edge capacities = Intersection area between slices
- Flow between source/sink without exceeding capacities
- Family of graphs: Vary α

Persistence Diagram:

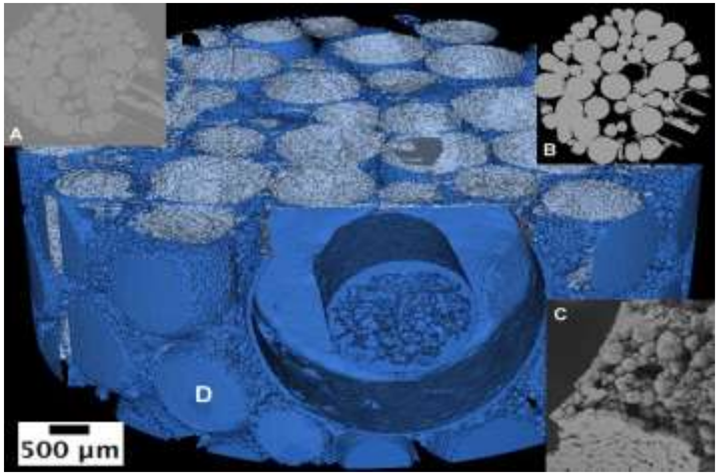
- Track components in superlevel set of distance function
- When component merge: “younger” component merges into “older” component



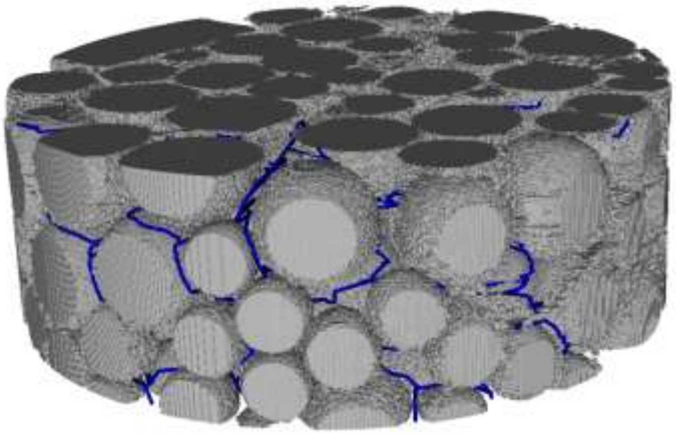
Pocket distribution from persistence diagram

U.S. DEPARTMENT OF ENERGY | Office of Science

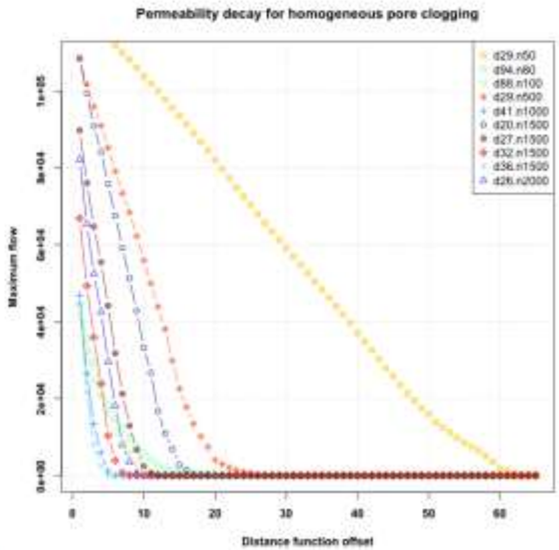
QuantCT: Results

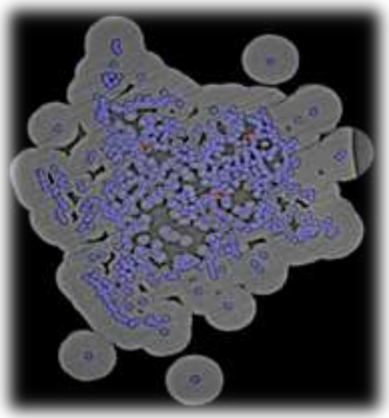


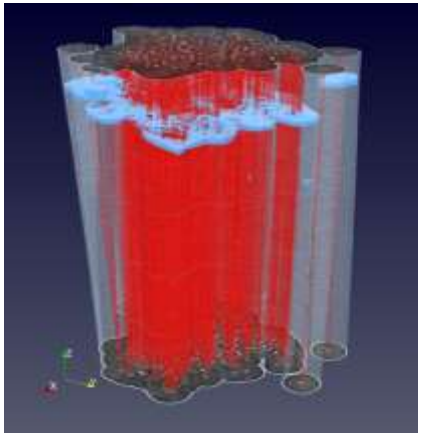
QuantCT software for microCT analysis (0.33 images/s)



Pore network through porous material







Automatic detection of 3D fibers and matrix cracking from assembled 2D slices

U.S. DEPARTMENT OF ENERGY | Office of Science

QuantCT: Workflow

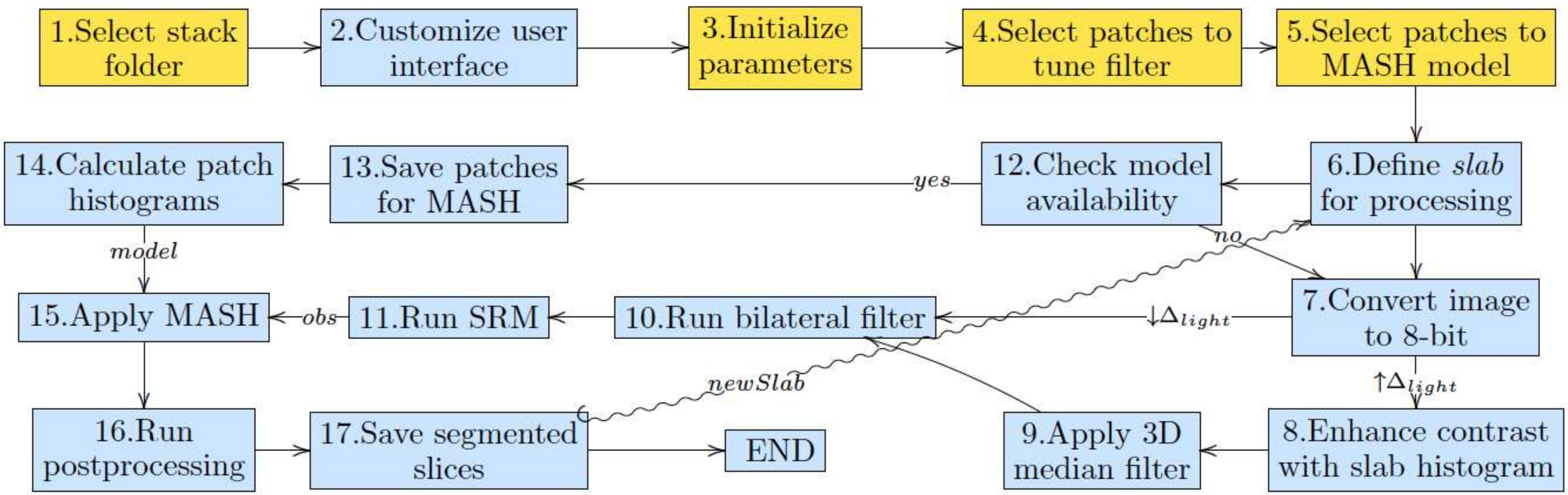


Figure 1: Flow diagram of Quant-CT segmentation workflow: yellow indicates user-interaction event and blue indicates a program action.

Ref. *Ushizima, D.M., Bianchi, A.G.C, deBianchi, C., Bethel, W., "Material science image analysis using quant-CT in ImageJ", in: ImageJ User and Developer Conference, 2012.*

QuantCT: Delivery Mechanisms

Delivery mechanisms:

Current:

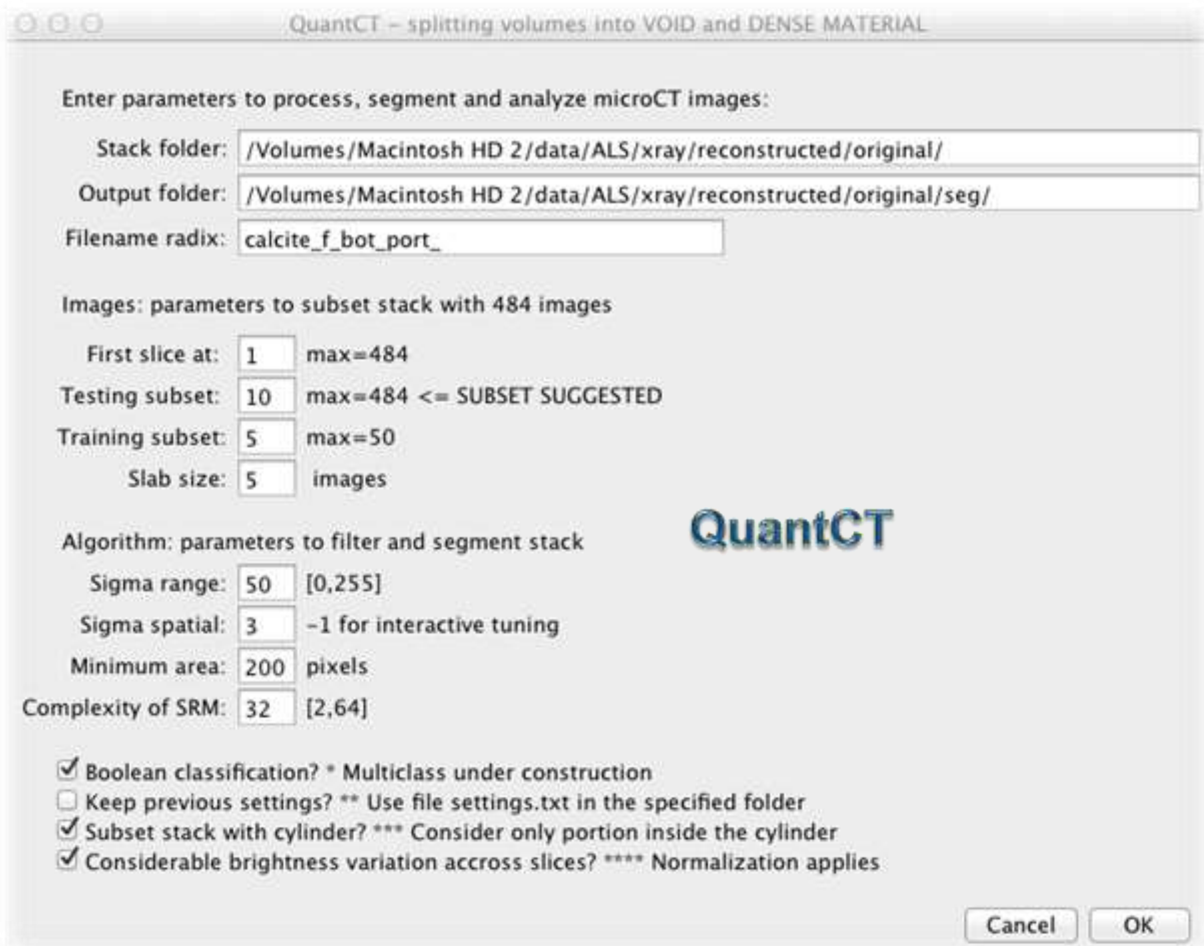
(1) Browser/computer at ALS

(2) Available as FiJi plugin

(3) Prototype source downloadable.

Code Specifics:

- Implemented in Java.
- Part of *Fiji* framework.
- Implemented in *OpenCL.*
- Called from Java code through *JOCL*.
- Dedicated thread assigned to each OpenCL device to handle multiple accelerators on any given node.
- Each thread requests unprocessed slices up to the maximum allowed by the hardware.



*Ref. **Ushizima**, Parkinson, Nico, Ajo-Franklin, Macdowell, Kocar, Bethel and Sethian, Statistical segmentation and porosity quantification of 3D X-ray micro-tomography. Applications of Digital Signal processing XXXIV, Vol. 8135, pp.1-14 (2011).*

Fluctuation X-Ray Scattering

Mathematics for structure reconstruction

J. Donatelli, J.A. Sethian (LBNL Computing Sciences)
P. Zwart (LBNL Advanced Light Source and Physical Biosciences)

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Fluctuation X-Ray Scattering
(Brand-new work)

CAMERA

(1) Fluctuation X-Ray Scattering:
- Extension of Small- and Wide- angle X-Ray Scattering
- X-Ray snapshots taken below rotational diffusion times
- Significantly more experimental information than traditional techniques
- Powerful technique for modern synchrotrons and free electron lasers (FEL)

(2) Going from real space structure to fluctuation scattering data is straightforward.

(3) But the reverse "inverse" problem is tough.

(4) A joint CAMERA collaboration between LBNL Physical Biosciences, ALS, and Computing Sciences has produced a new technique: **"M-TIP"**, which exploits multi-tiered iterative phasing and solves this inverse problem.

(5) The new method figures out structure of objects that cannot be crystallized, at a far higher resolution than previously available (to appear: PNAS in few weeks)

(6) Example: Using **M-TIP**, was able to reconstruct 3D profile of pentameric ligand-gated ion channel (pLGIC), from Protein Data Bank entry 4NPP

CAMERA: Where is this going? Reaching out:

Positive response from the community

BESAC Committee Presentation (June 2014)

Oak Ridge/SNS Joint NSRC Workshop (June 2015)

"Big, Deep, Smart Data Analytics in Materials Imaging"

Brookhaven and NSLS (May 2015)

Invited Seminar Talk

Argonne/APS Joint Workshop (March 2015)

"Frontiers in Data, Modeling, and Simulation

LBNL/Advanced Light Source (Oct. 2014)

"CAMERA Workshop on Real-Time Robust Ptychography"

Many software requests:

Download, use CAMERA codes

Requests for CAMERA to house, curate, and host algorithms and software across the light sources

New Joint Projects starting up:

GISAXS and extensions to neutron sources (ORNL/SNS)

Ptychography (BNL, ANL)

Fluctuation scattering (LCLS)

CAMERA: Take Home Messages

Knowing what to build, how to build it, and how to use it requires close-knit, coordinated teams with many different skills.

With careful attention to mathematics and algorithms, we can build codes and software tools that can transform data into the information that users really want.

camera.lbl.gov