Applied Mathematics For Experimental Science

J.A. Sethian

Mathematics Department LBNL and Department of Mathematics, UC Berkeley

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

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

Support: LBNL LDRD, now Joint ASCR-BES Pilot Project

(Steve Lee, Program Manager)

U.S. DEPARTMENT OF ENERGY | Office of Science

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

ALS: Beam scientists
(D. Parkinson, A. Hexemer, S.Marchesini, D. Shapiro)

Foundry/NCEM
(J. B. Neaton, W. Queen, D. Britt)

LBNL Math Department
(J.A.Sethian, C. Rycroft, J. Donatelli)

CAMERA

LBNL Comp. Sci.
(S. Li, M. Haranczyk, H. Harikrishnan, T. Perciano, D. Ushizima, C. Yang)

LBNL: Materials, Earth Sciences

Berkeley Campus

LBNL/UCB: EFRC
(M. Haranczyk, J.B. Neaton, B. Smit)

EECS/Statistics Depts.

Mathematics (Faculty/postdocs/grad students)
(J.A. Sethian, L. Lin, A. Grunbaum, E. van Andel, B. Preskill)

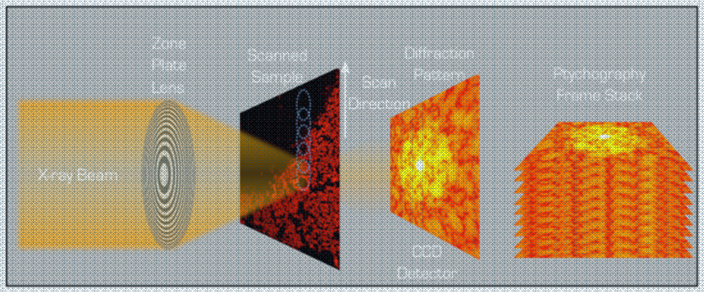
BIDS (Berkeley Institute for Data Science)

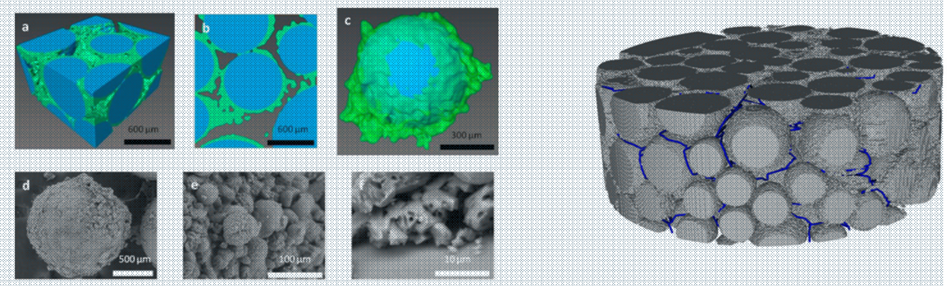
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

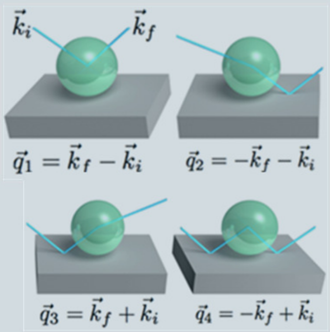
Pytchography (ALS/MATH)

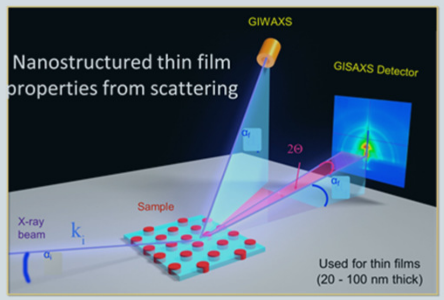
Coherent Diffraction with Microscopy





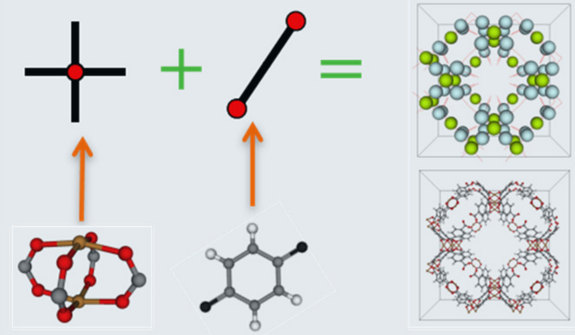
Automatic Image Analysis (ALS/MATH)





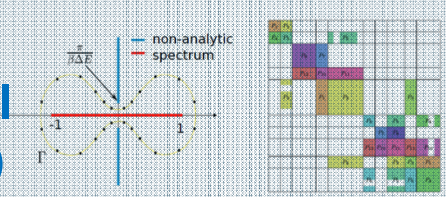
GISAXS (ALS/MATH)

Grazing incidence small angle x-ray scattering



Designer Materials (Molecular Foundry/Math)

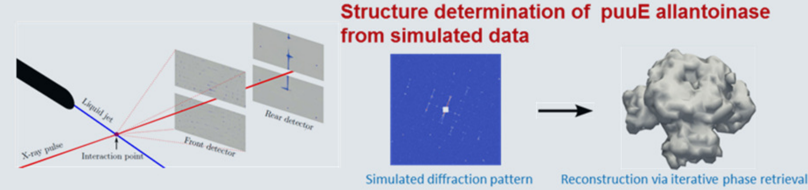
New methods for Density Functional Theory (MF/MATH)



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

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

Structure determination of puuE allantoinase from simulated data



X-Ray nanocrystallography (ALS/MATH)

Key Points

(1) For a relatively small investment, applied mathematics can help advance science, decrease turnaround, reduce cost at facilities.

(2) Mathematics provides critical tools: reduces “barrier to entry”.

(3) This is an opportune time: traditional mathematical boundaries are breaking down, especially around experimental data.

(4) Biggest opportunity: when scientists, experimentalists, and mathematicians work together, in close proximity, at facilities.

(5) Ultimately, underlying mathematics and algorithms should be invisible/automatic— users “push a button and get information”.

Outline of Talk

How Did This Start?

History and Motivation

How are we doing this?

Organization. The challenges. Opportunities.

What are the Efforts?

Six Different Projects

Where is it going?

(Opportunities for model to tackle other DOE problems)

Background: LBNL/UCB Mathematics:

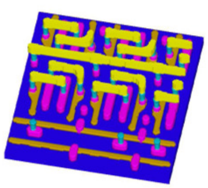
Long Standing Program:

Develop the mathematics, algorithms, and implementations for problems of interest to DOE's Advanced Scientific Computing Research.

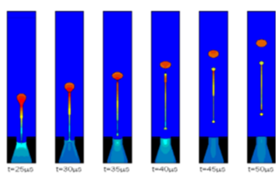
Combination of:

- A lab/academic/university/high performance computing environment (intimate connection to UC Berkeley: Math, Engin., Bio., Chem., Materials...)

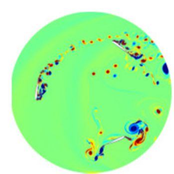
- Scientific/engineering problems across traditional academic boundaries



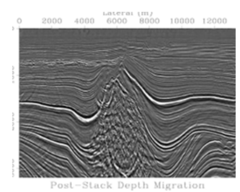
Semiconductors



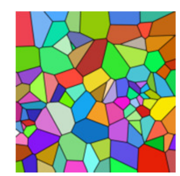
Industrial printing



Wind turbines



Seismic imaging



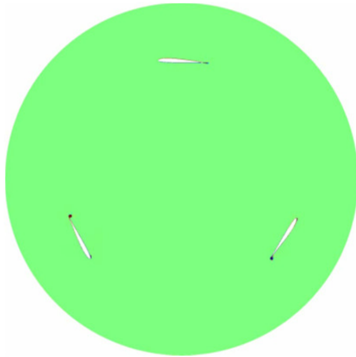
Foams in manufacturing

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

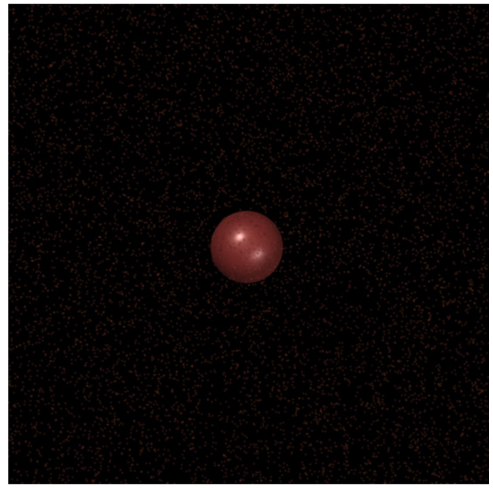
People: Campus Faculty, Postdocs, Graduate students, Visitors (up at LBNL)

Long-Running Program (40+ years)

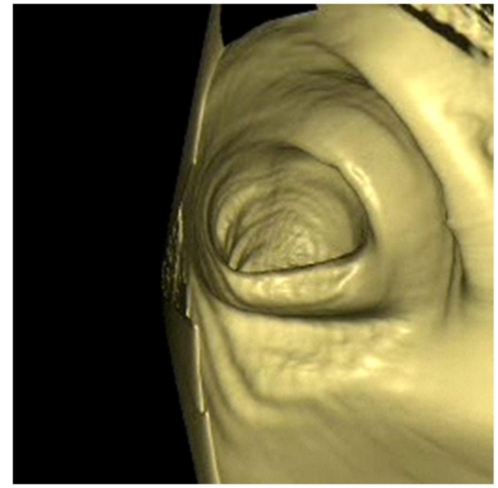
Examples:



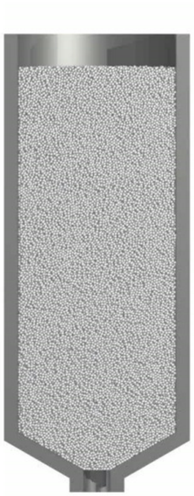
Vertical Axis Wind Turbines



Cell Cluster Growth

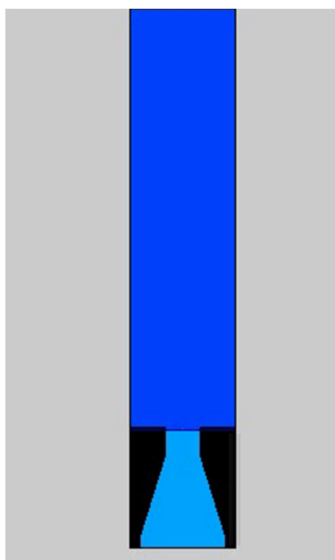


Virtual Colonoscopy



Draining in Coal Hoppers

Industrial Inkjet Printing



Industrial Foams

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

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…)?

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

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
Lawrence Berkeley National Laboratory

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?

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

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

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

Can data be measured, processed, organized and displayed in a way that helps understand and help shed light on further experiment?

Can this data be transformed so that it can be used to initialize computational models, with output framed to complement experiment?

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

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

(d) Many problems are similar, but not the same—requires customization and tailoring....

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.**

An explosion of work in new, hybrid fields:

Computational harmonic analysis in image reconstructions (ALS)

Stochastic analysis and uncertainty quantification (ALS)

Topological data reduction and classification (Foundry)

Discrete/Continuous fast PDE solvers (Foundry, ALS)

Group theoretic methods in machine learning (ALS)

Differential geometric methods for interface evolutions (MF)

For our six projects:

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

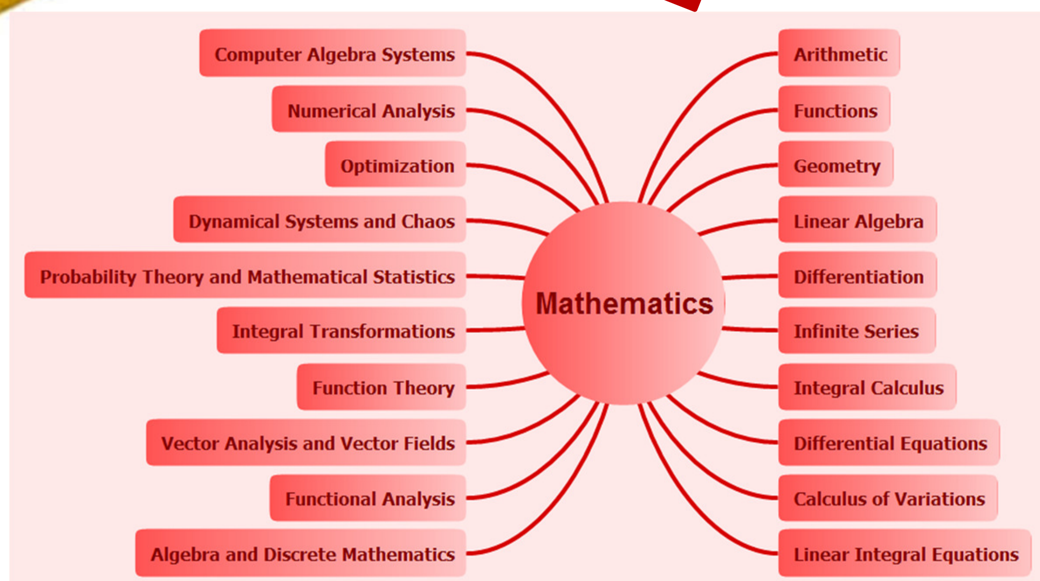
U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Mathematics and (or versus!) "Big Data"

Data

Computers



Mathematics

Mathematics is what changes data into information

Problems are only going to get worse

More data

More complexity

Less obvious relational linking

More noise

More false signals

Going to need mathematics more than ever…

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

CAMERA: Center for Applied Mathematics for Energy Research Applications

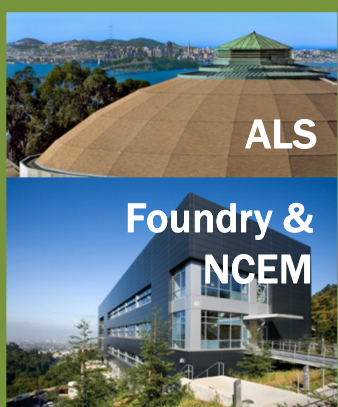
CAMERA

BERKELEY LAB

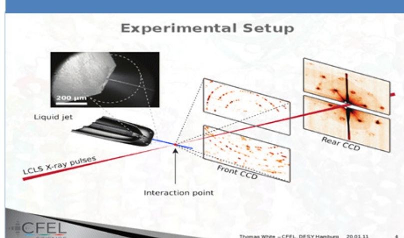
Today: Facilities data is time-consuming	**Tomorrow:** More data. More quickly. High resolution.	**Critical need:** algorithms and analysis for *understanding*	**LBNL approach:** Focused teams of mathematicians/ domain scientists	**New math to:** Guide and optimize experiments

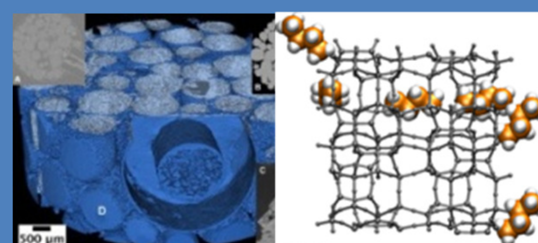
Goal: Build the applied mathematics that helps *transform experimental data into understanding*

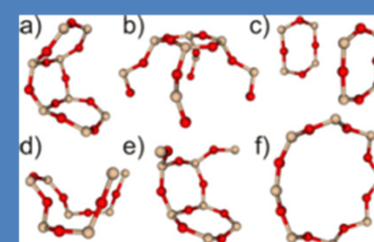
Pilot Partners



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: 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?

(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

Additionally: providing codes/browsers/portals to SpotSuite

BERKELEY LAB

AN OVERVIEW OF SOME OF THE WORK UNDERWAY

Application

Mathematical issues

New mathematics we needed to build and exploit

Software for Users

(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)

U.S. DEPARTMENT OF ENERGY | Office of Science

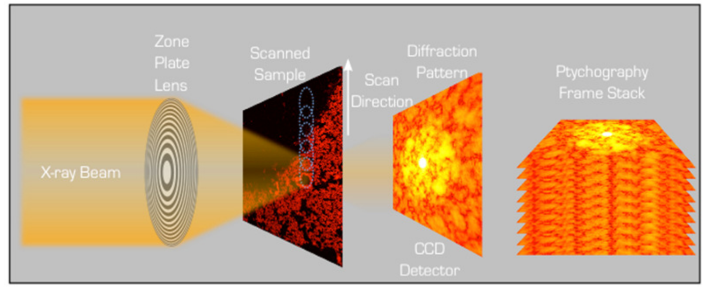
BERKELEY LAB
Lawrence Berkeley National Laboratory

CAMERA: Ptychographic Imaging

CAMERA

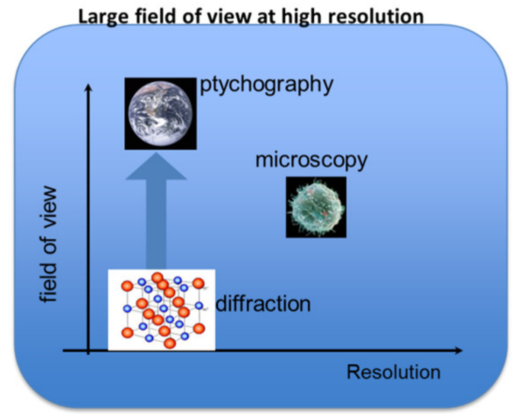
S. Marchesini, D. Shapiro, H. H. Krishnan (LBL) F. Maia (LBL/Uppsala), H-T. Wu (Stanford))

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

CAMERA

BERKELEY LAB

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?

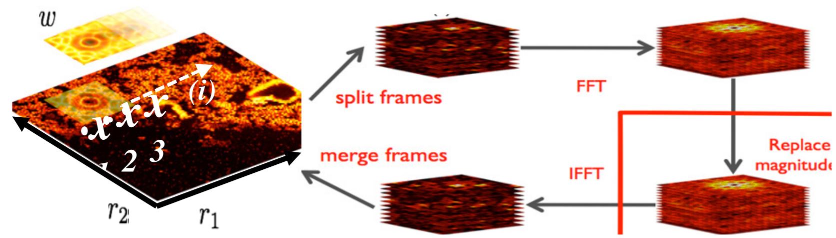
(given a detector, with a limited rate, dynamic range and response function, what is the best scheme to encode and extract more information per detector channel?)

Phase retrieval in high dimensional space

Challenges with basic alternating projection algorithm:

Poor scaling:

long range interactions among frames decay exponentially with distance.

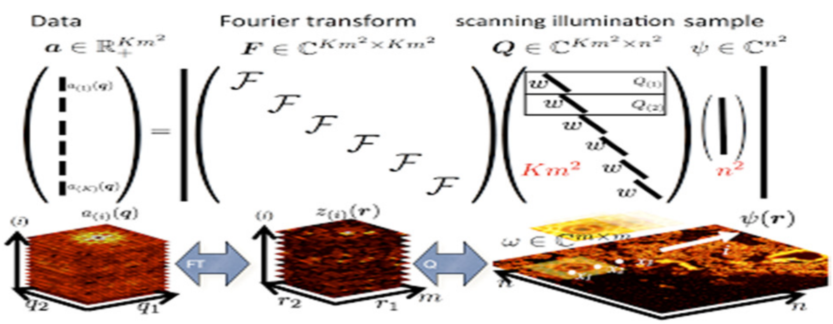


Poor initial guess:

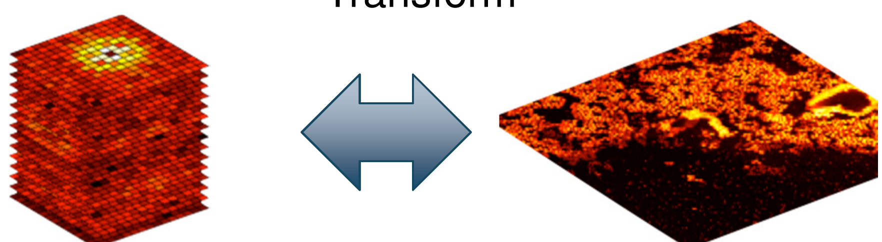
can significantly delay convergence.

Ultimately, an overdetermined problem

in high dimensional space.



Short-time Fourier Transform



Large dimensional data

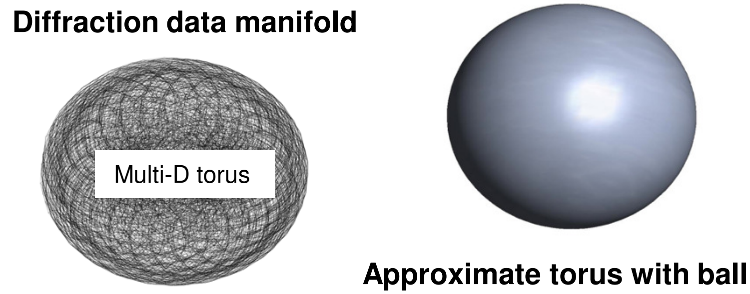
Low dimensional space

How can we speed this up?

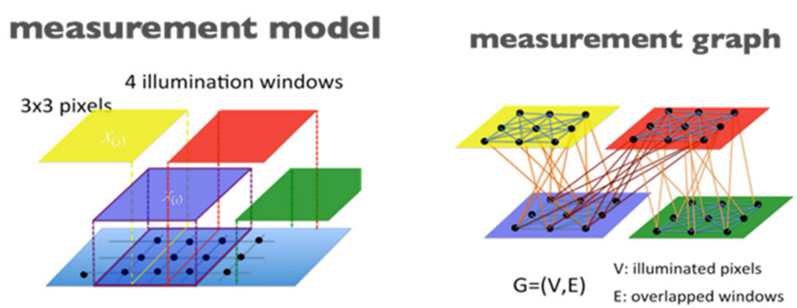
Graph Theory/Graph Laplacians

Building a better starting guess:

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



(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.

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Fast Multiscale Algorithms

BERKELEY LAB
Lawrence Berkeley National Laboratory

Fast multiscale approach:

(1) Above approach can be augmented by alternating long range/short range (framewise/pointwise) relaxations of the connection graph Laplacian. Additionally, use implicit Hessian for fast line search.

(2) This achieves accelerated convergence for large scale phase retrieval problems spanning multiple length-scales.

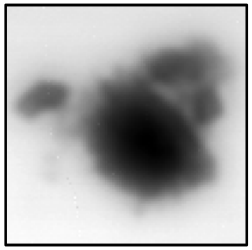
(3) This approach also recovers experimental fluctuations over a large range of time-scales.

(4) Brand-new: Framewise rank-1 **accelerated illumination** recovery by transparency estimation.

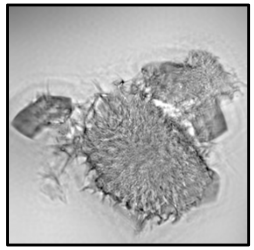
Released Code: SHARP: Scalable Ptychography Solver

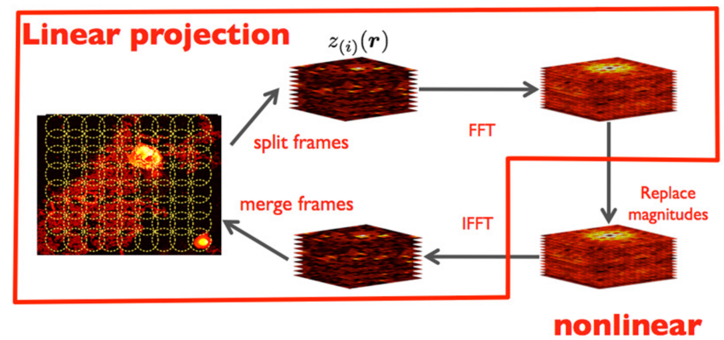
CAMERA

Microscopy



Pytchography



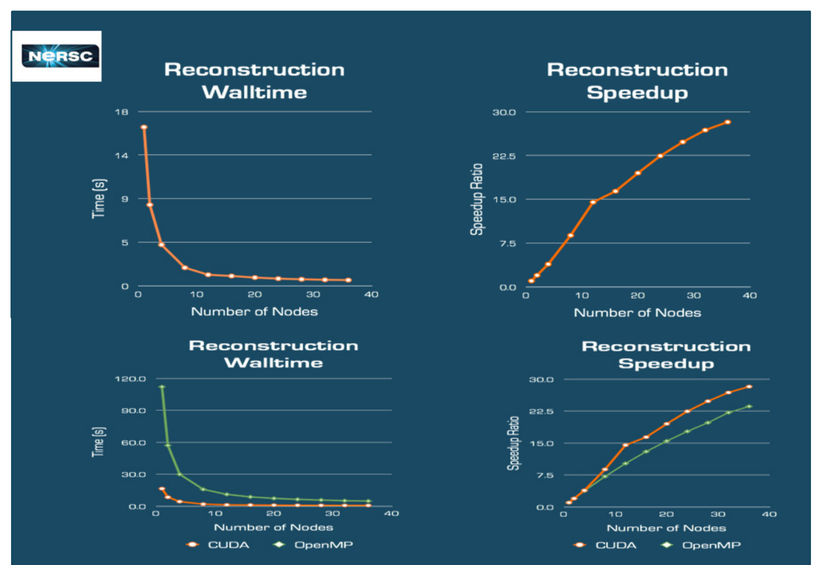


Fast Implementation of Split and Overlap Kernels on GPU

Higher level parallelization for real -time performance (MPI)

Strong Scaling tests for large dimensions

Lens reconstruction, vibrations, background, coherence, multiplexing



U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

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 (invited), CAMERA/ALS/BNL AUG 2014 CAMERA/ALS/APS Sep 10/14, COHERENCE, XRM , SIAMIM, FIO/LS, RACIR summer school, ALS Users workshop

U.S. DEPARTMENT OF ENERGY | Office of Science

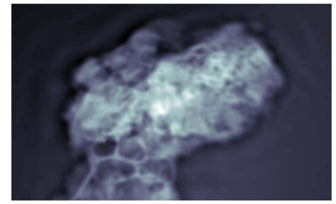
BERKELEY LAB

Toward real-time feedback

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

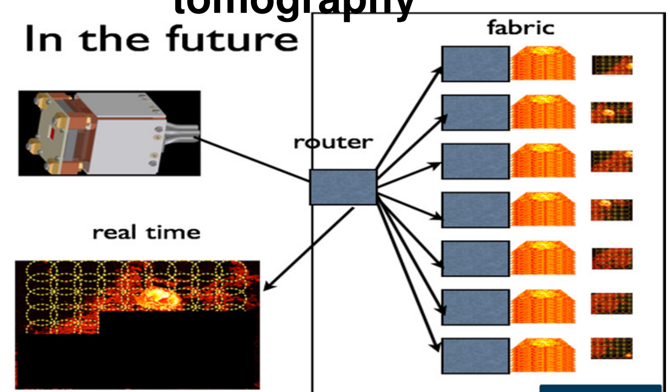
In the future
low Latency (<5 ms) feedback by streaming detector frames on distributed direct memory access fabric.

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



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

High bandwidth 3D torus p2p
For hyperspectral ptycho tomography



QuantCT

Automatic image analysis tools for micro-CT

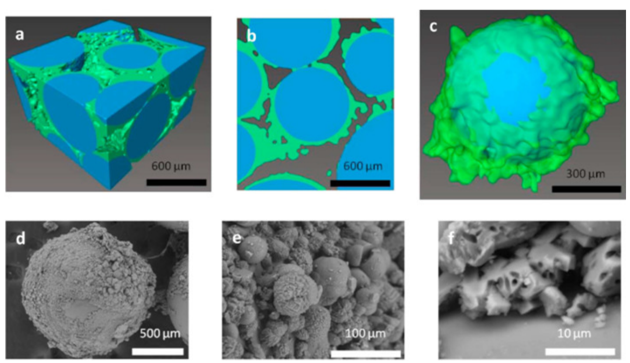
D. Ushizima, D. Morozov, H. Krishnan, T. Perciano (LBNL Computing Sciences)
D. Parkinson (Advanced Light Source)

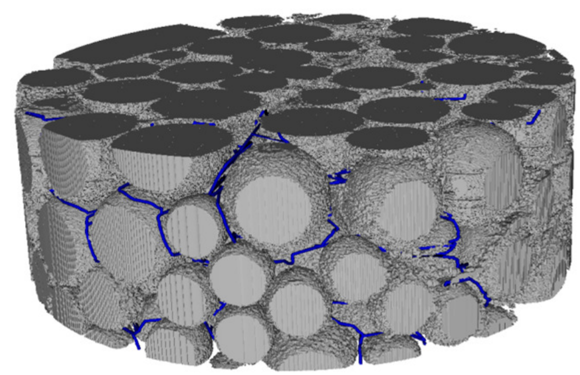
CAMERA: Quantitative Image Analysis of Micro-CT Samples

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

CAMERA

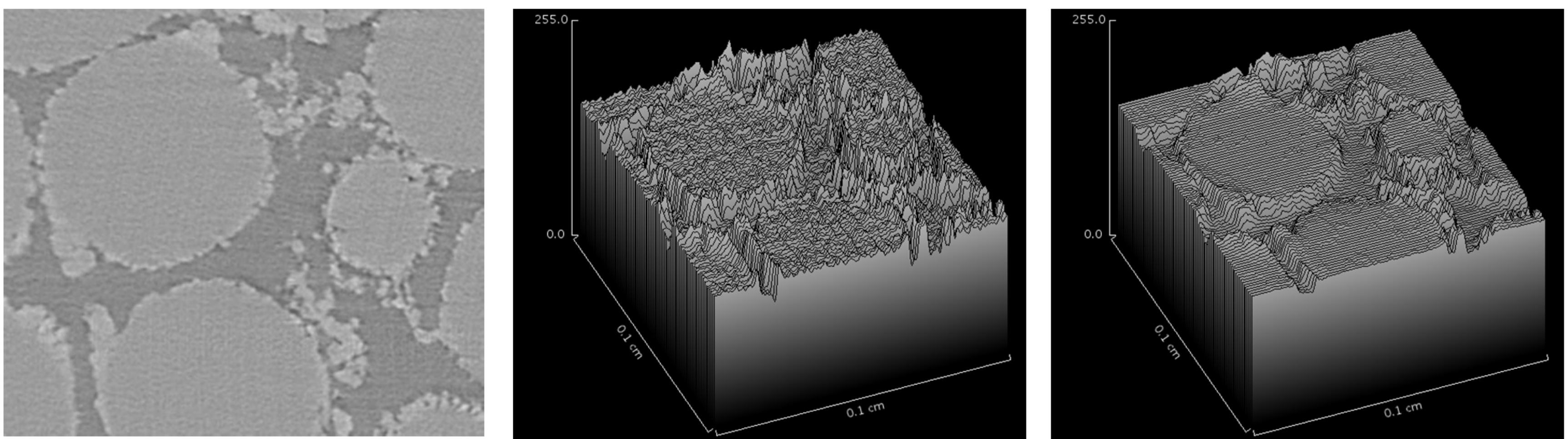
QuantCT: A. Pre-processing/bilateral filters

BERKELEY LAB

- **Smooth and preserve edges: weighted average of local neighborhood – weights based on spatial and intensity (range) distances;**

$$h(x) = k^{-1}(x) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi) \overbrace{c(\xi, x)}^{\text{geometric}} \overbrace{s(f(\xi), f(x))}^{\text{photometric}} \, d\xi$$

$$k(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} c(\xi, x) \, s(f(\xi), f(x)) \, d\xi$$



QuantCT

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

QuantCT: B. PDE-based Automatic Segmentation and Extraction

CAMERA

BERKELEY LAB
Lawrence Berkeley National Laboratory

(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:**

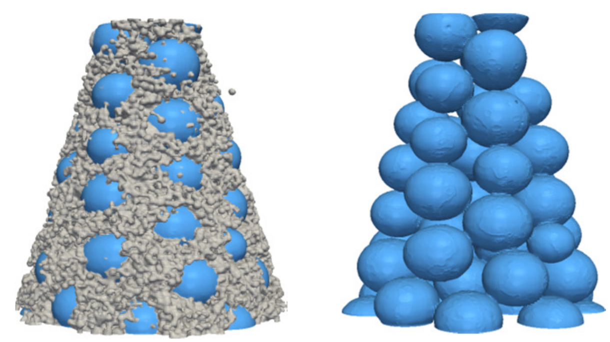
$$\phi_t + F|\nabla\phi| = 0, \text{ where } F = [((I - I_1)^2 + ((I - I_2)^2) - \mu\nabla\cdot(g\nabla\phi/|\nabla\phi|)]$$

(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”

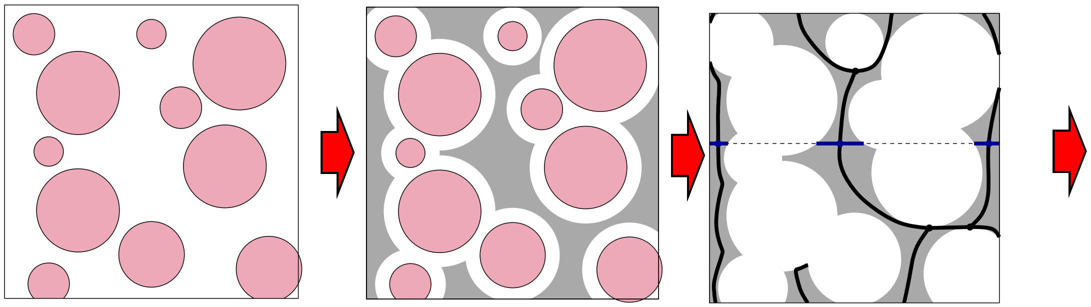
BERKELEY LAB
Lawrence Berkeley National Laboratory

QuantCT: C. Determination of Connectivity and Channel Pathways

CAMERA

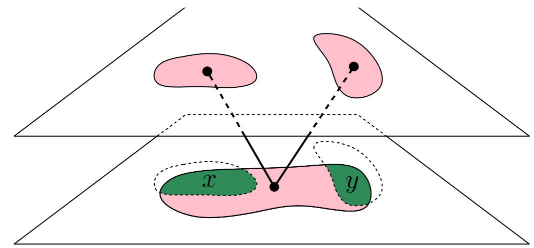
BERKELEY LAB

Augmented Topological Descriptors: Max Flow Graphs and Persistence Diagrams

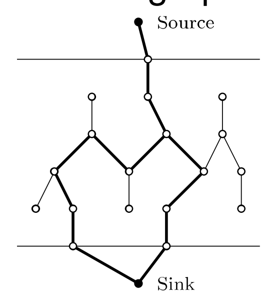


Void Space   Wide pathways   Reeb graph

Edge capacities



Flow graph

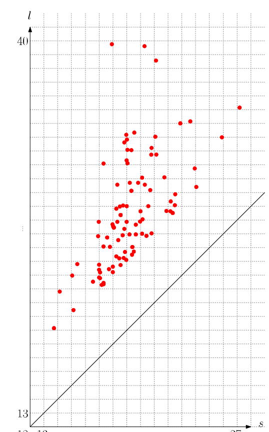


Max-Flow:

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

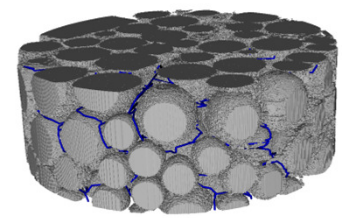
Persistence Diagram:

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



Pocket distribution from persistence diagram

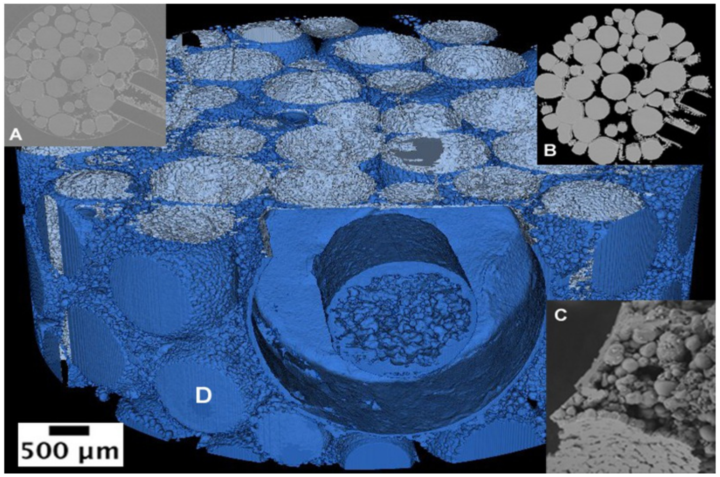
Ford Fulkerson



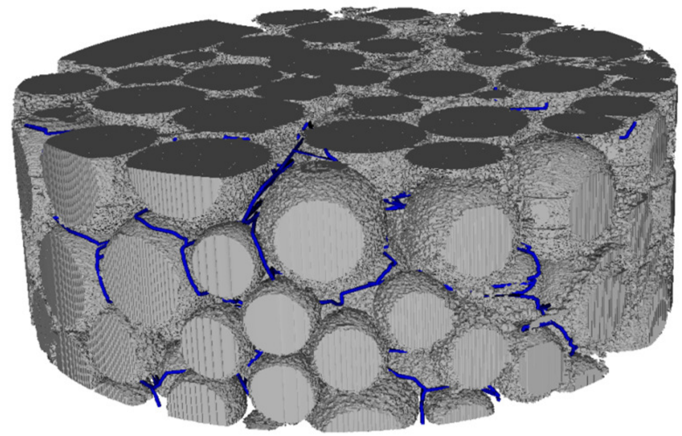
U.S. DEPARTMENT OF ENERGY | Office of Science

CAMERA

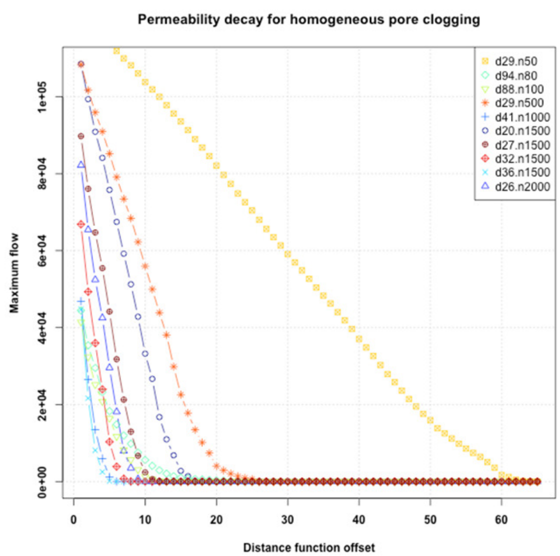
QuantCT: Results

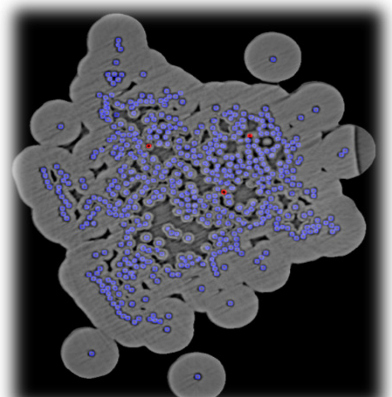


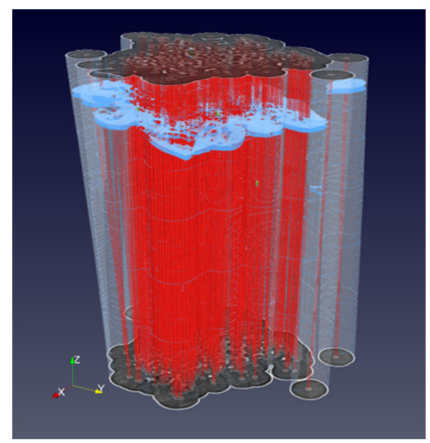
QuantCT software for microCT analysis (0.33 images/s)



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

BERKELEY LAB

QuantCT: Workflow

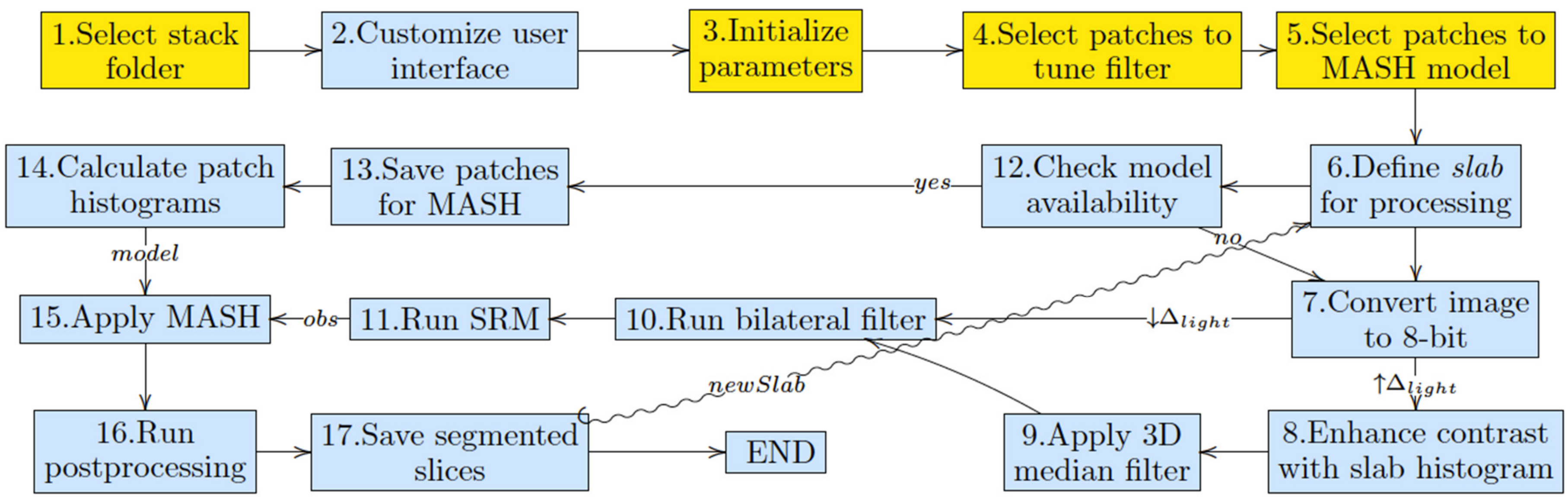


Figure 1: Flow diagram of Quant-CT segmentation workflow: yellow indicates user-interaction event and blue indicates a program action.

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:

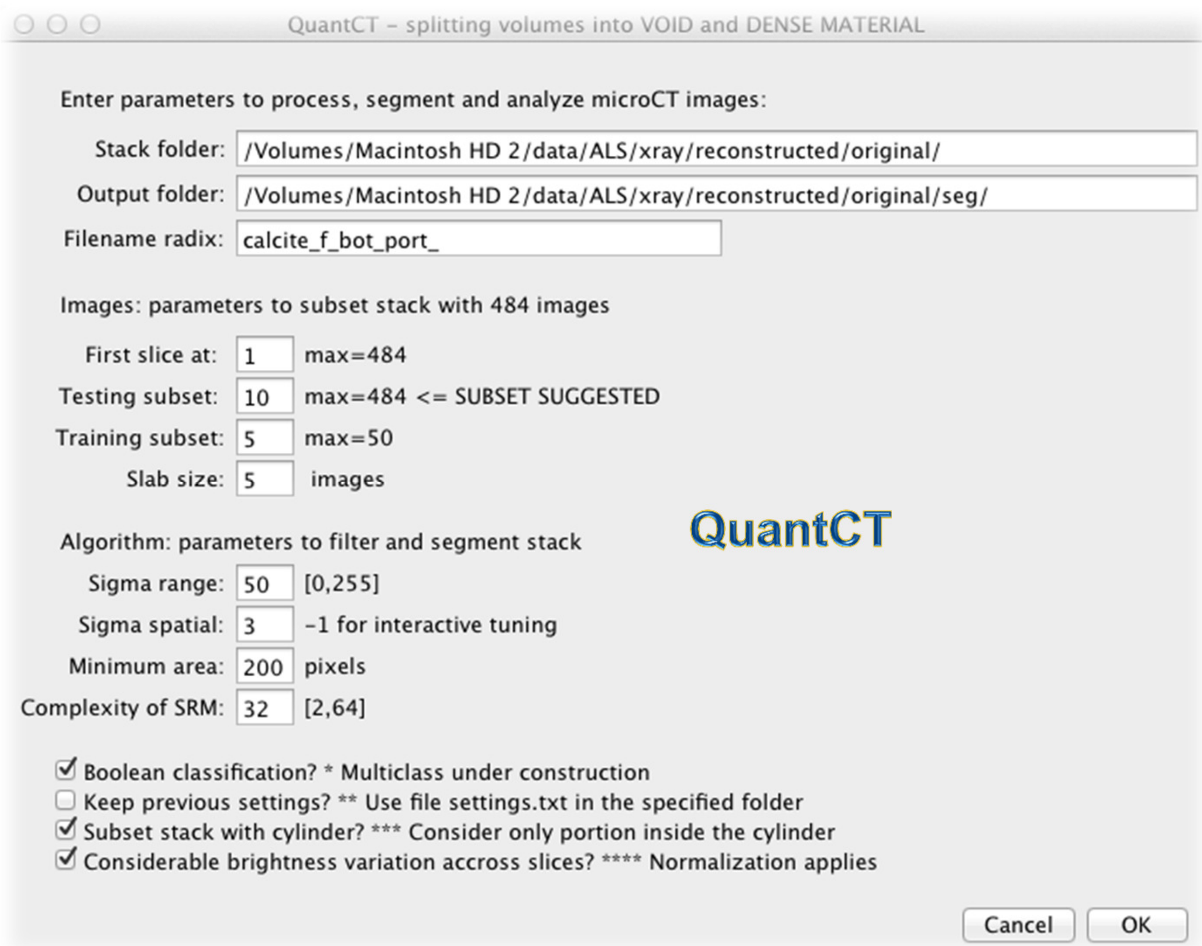
Current:
(1) Browser/computer at ALS

(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).*

PEXSI:

Accelerating electronic structure calculations for large scale materials systems

L. Lin (UC Berkeley Math),
C. Yang (LBNL Computing Sciences)
J. Neaton (Molecular Foundry)

Accelerating electronic structure calculations for large scale materials systems

Starting point: Density Functional Theory:

Reformulates Schrödinger's eqn. as non-interacting electrons moving in an effective potential, which must be determined.

Results in a non-linear eigenvalue problem:

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

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

- Solve for eigenvalues ε & eigenfunctions ψ, which depend on electron density
- Electron density from summing eigenfunctions. Eigenfunctions orthogonal.
- Exchange correlation function V_{xc} (ρ) depends on electron density.
- Big Challenge: Find a self-consistent way to solve this.

One Approach: Iterate:

•First, make initial guess for the electron density $\rho(x)$
•Then, solve eigenvalue/eigenvector problem.
•Then, recompute $\rho(x)$, and repeat.

Problem: Slow, expensive, limited to small systems:

•Scales like cN^3, N = # electrons, c=grid/orbital resolution

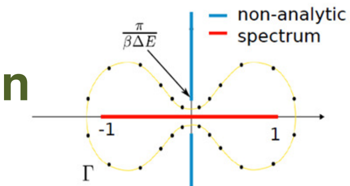
PEXSI:
Pole Expansion and Selected Inversion

CAMERA

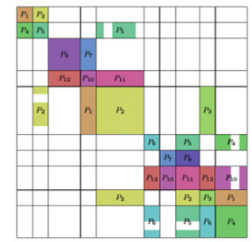
BERKELEY LAB
Lawrence Berkeley National Laboratory

A different approach: (L. Lin and C. Yang, LBNL & UC Berkeley)

(1) PEXSI: Reduce KSDFT cost calculation to at most N^2 scaling without sacrificing accuracy.



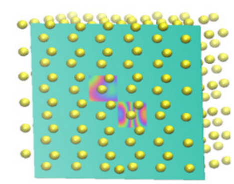
Idea 1a: Represent Fermi operator by pole expansion



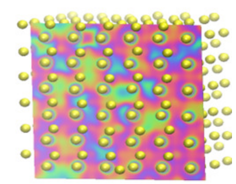
Idea 1b: Develop selected memory inversion. Idea 1c: massively parallel distributed memory implementation

Further ideas:

(2) Develop discontinuous Galerkin basis functions to represent continuous physical quantities with low cost and high accuracy.

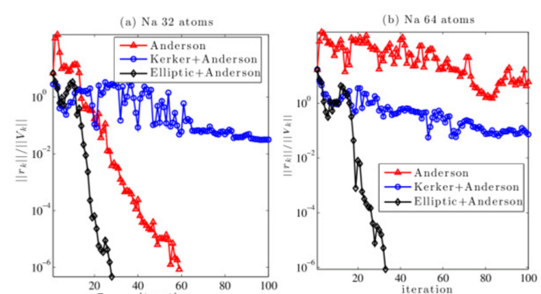


Discontinuous basis function for 3D Na



Nearly continuous electron density recovered

(3) New elliptic preconditioner: reduce number of required self-consistent field iterations (SCF) for large inhomogeneous metallic systems:



Example: Standard systems: 1 million degrees of freedom (DOF).
PEXSI solves 4 billion DOF in 25 min on 4096 processors

PEXSI: An alternative representation of Kohn-Sham

CAMERA

BERKELEY LAB

(1) Density Functional Problem: typically viewed as a non-linear eigenvalue problem.

(2) Instead, solve a non-linear fixed-point problem involving the Fermi-Dirac operator, defined in terms of a matrix function of the Hamiltonian.

(Trading a non-linear eigenvalue problem for a non-linear fixed point has its own challenges—has to be evaluated at every step.)

A Breakthrough: Represent the Fermi operator using a pole expansion.

(a) Electron density typically evaluated by taking diagonals of "Fermi operator"

$$\rho = \mathrm{diag}\frac{2}{1+e^{\beta(H[\rho]-\mu I)}}$$

Want to find ρ through fixed Point iteration

(b) Pole expansion: Represent the Fermi operator by rational functions (single poles), and evaluate the rational function directly without using diagonalization.

$$\rho = \mathrm{diag}\frac{2}{1+e^{\beta(H[\rho]-\mu I)}} = \mathrm{diag}\frac{2}{1+e^{\beta\Delta E\frac{H[\rho]-\mu I}{\Delta E}}} \approx \mathrm{diag}\left\{\sum_{l=1}^{P} c_l\left(\frac{H[\rho]-\mu I}{\Delta E}\right)^l + \sum_{l=1}^{Q}\frac{\omega_l}{\left(z_l I - \frac{H[\rho]-\mu I}{\Delta E}\right)^{q_l}}\right\}$$

$$\rho \approx \mathrm{diag}\sum_{i=1}^{Q}\frac{\omega_i}{H - z_i I}$$

$z_i, \omega_i \in \mathbb{C}$ are complex shifts and complex weights

(c) Evaluate using contour integration techniques.

($\mu =$ chemical potential)

PEXSI: Now, extract key elements of Hamiltonian

CAMERA

BERKELEY LAB

$$\rho \approx \mathrm{diag} \sum_{i=1}^{Q} \frac{\omega_i}{H - z_i I}$$

Next idea: apply selected inversion

- H is a sparse matrix, but $(H - z_i I)^{-1}$ is a full matrix. Fast algorithm for computing the diagonal of an inverse matrix?
- Selected inversion: A=LDL^T: A^{-1} restricted to the non-zero pattern of L is "self-contained". Therefore the selected inversion algorithm only computes all the elements A^{-1}_ij such that L_ij is nonzero, which leads to significant saving of the computational cost.

[Idea of selected inversion dates back to [Erisman and Tinney, 1975], [Takakashi et al 1973]; For electronic structure [LL-Lu-Ying-Car-E, 2009]; For quantum transport [Li, Darve et al, 2008, 2012]]

Other Speedups: Adaptivity and faster convergence

CAMERA

BERKELEY LAB
Lawrence Berkeley National Laboratory

Next idea: Build discontinuous basis functions which are eigenfunctions of the Kohn-Sham Hamiltonian on local domains

A more economical representation, providing systematically higher accuracy.

(a) Discontinuous Galerkin (DG) based adaptive local basis set:
(b) Constructed by solving Kohn-Sham problems locally in the real space.
(c) Automatically and systematically builds the rapid oscillations of the Kohn-Sham orbitals around the nuclei into the basis functions.
(d) Each basis function is discontinuous in the global domain.
(e) The continuous Kohn-Sham orbitals and the electron density are evaluated from the discontinuous basis functions using discontinuous Galerkin (DG) framework.

Next idea: Developed elliptic preconditioners for accelerating the nonlinear SCF iteration for large scale inhomogeneous metallic systems.

(i) This elliptic preconditioner solves an elliptic equation derived from the polarizability matrix with O(N) cost.
(ii) Much smaller computational cost than existing preconditioners, some of which scale as badly as O(N^4).

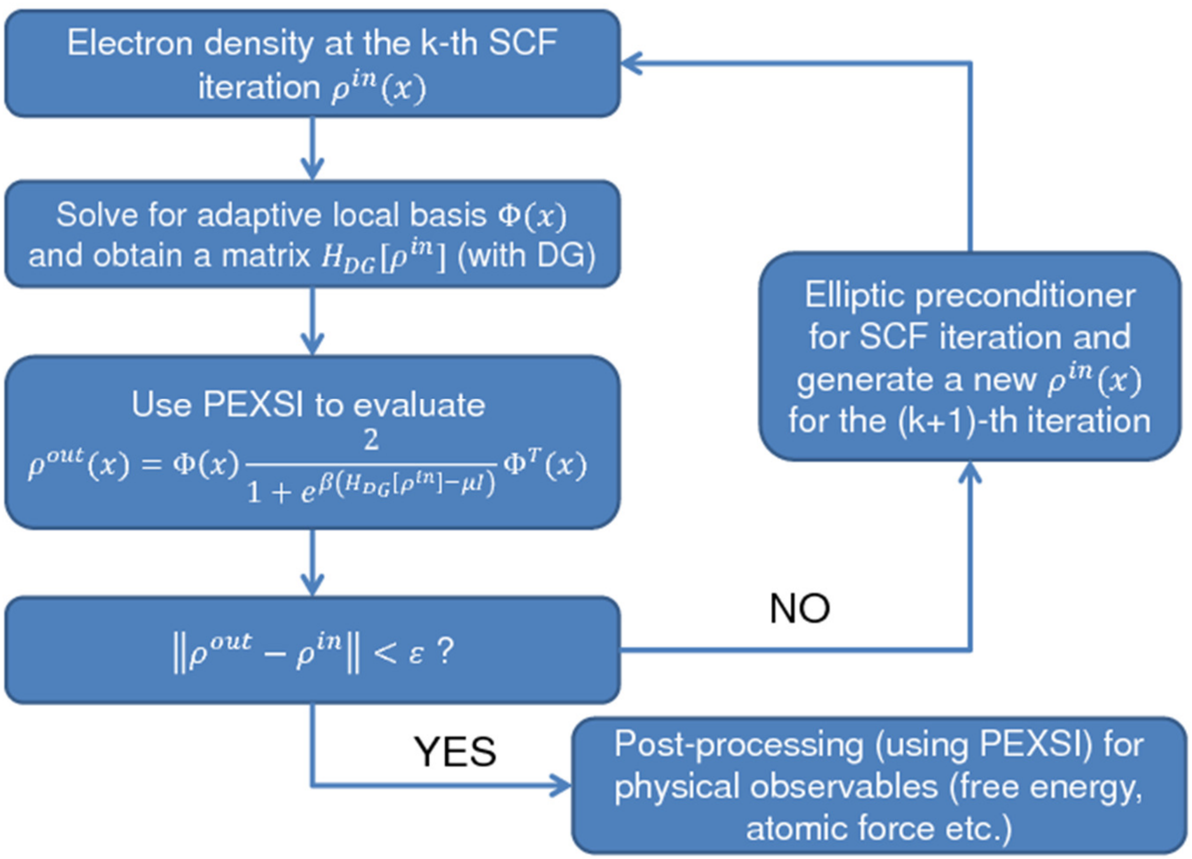
CAMERA

CAMERA: PEXSI: Overall framework for accelerating electronic structure calculation for large scale materials systems

BERKELEY LAB

Lawrence Berkeley National Laboratory

Summary of the algorithmic cycle



CAMERA: PEXSI. Delivery system:

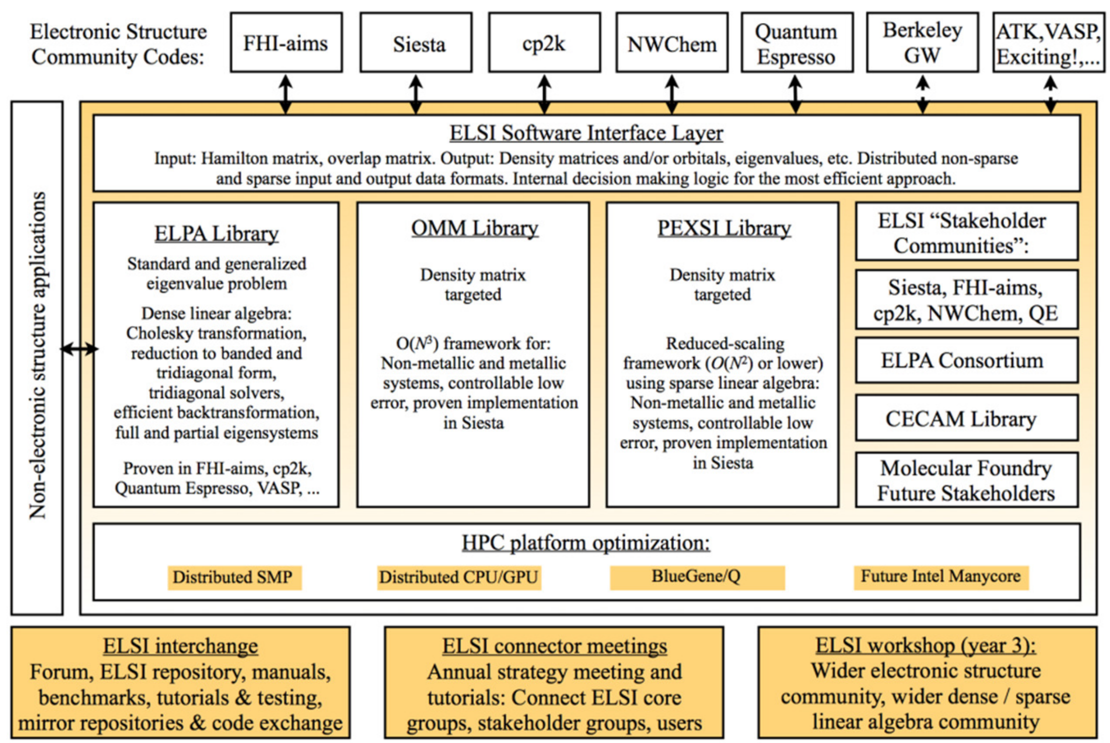
(1) PEXSI is integrated and available within SIESTA
(Spanish Initiative for Electronic Simulations with Thousands of Atoms)

(2) PEXSI currently being integrated into CP2K (Joost VandeVondele et al).

(3) In communication with BigDFT (Basel) and FHI-aims (Fritz Haber Institute)

(4) "Electronic Structure Infrastructure" (ELSI) Proposal (Pending)

Integrate PEXSI, together with companion software, into a large set of Electronic Structure Codes



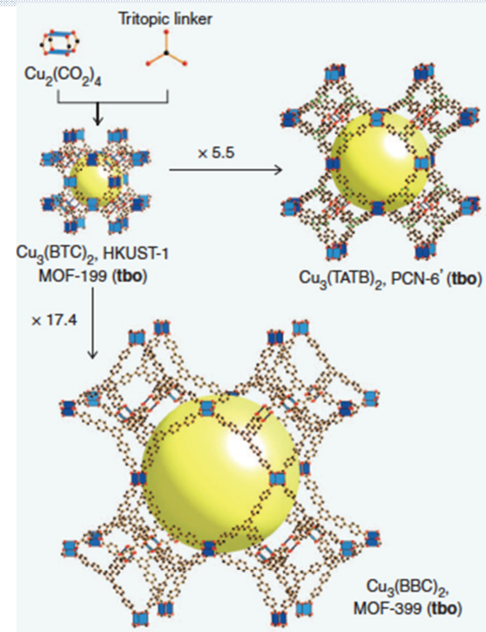
Zeo++
Mathematics and Algorithms for Designing New Porous Materials

M. Haranczyk (LBNL Computing Sciences, Molecular Foundry)
J.B. Neaton (Molecular Foundry)
C. Rycroft, J.A. Sethian (Mathematics: UC Berkeley and LBNL Mathematics)

CAMERA: Mathematics and Algorithms for Designing New Advanced Porous Materials

Scientific Opportunities:

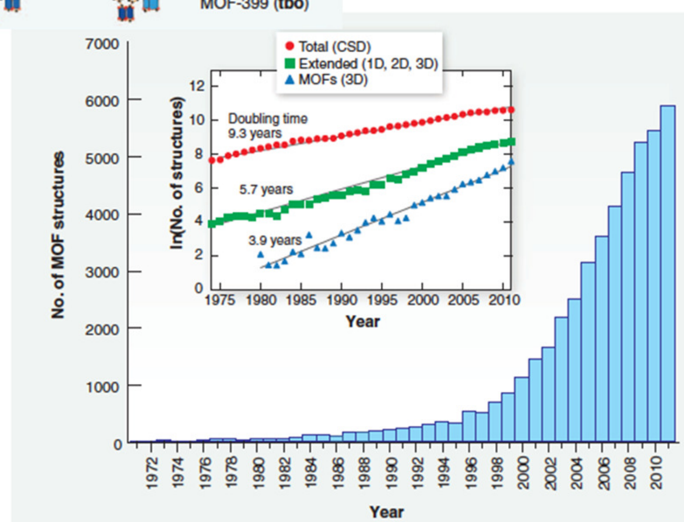
- Numerous families of advanced porous materials synthetized: MOFs, COFs, PAFs
- Demonstrated in multiple applications: separations, gas storage, catalysis, drug delivery etc.
- Structures offer unmatched “tunebility” under *reticular chemistry paradigm: (replacing building blocks within same topology of materials)* – by exchange of building blocks one executes tinker toy chemistry



Assembly of larger structures from smaller “basis” molecular building blocks.

Mathematical/Scientific Challenges:

- Infinite search space pose challenge to identify optimal materials for particular applications as well as to explore possibilities
- Need tools to analyze, characterize and classify porosity in very large number of structures



Growth in built materials

Figures from Nature 2003, 423, 705, Science 2013, 341, 1230444

CAMERA: Mathematics and Algorithms for Designing New Advanced Porous Materials

Mathematical issues/Transforming into Math:

- **Abstracting real molecules/materials into mathematical models**
  - Statistical approaches to describe abstract molecules
  - Replace atoms with geometric blocks,
    (don't know the mass of building blocks, use model to link surface area and mass.)

- **Mapping porosity ("dual" of an real material)**
  - Computational geometry and Voronoi diagrams to quickly characterize void space.
  - Fast PDE-Eikonal solvers to navigate, characterize, and refine void space.

- **Optimization and optimal selection of building blocks to achieve desired properties**
  - Gradient descent optimzation
  - Genetic algorithms.
  - Are using these to optimize structure to maximize surface area
  - (internal surface area important for MOFs).

Specific Linkage and Utility:

For the Molecular Foundry: On-going research on MOFs

For the EFRC: Discovering new advanced porous materials for gas separation)

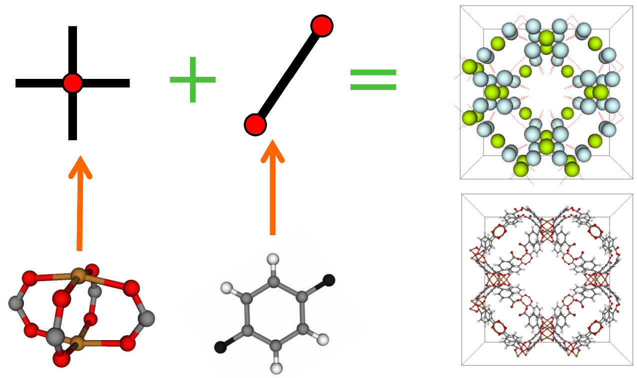
For the Material Genome: Materials discovery, high-throughput materials analysis, and data mining tools

CAMERA: Mathematics and Algorithms for Designing New Materials

Example Algorithms and Math Involved:

(1) Assembling Potential Materials

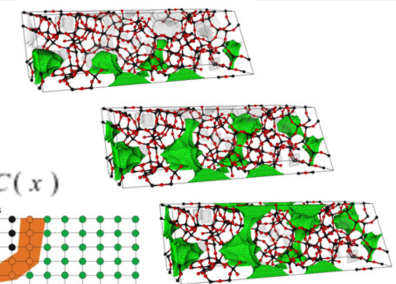
Tools for 3D assembly of porous polymer models from enumerated periodic graphs



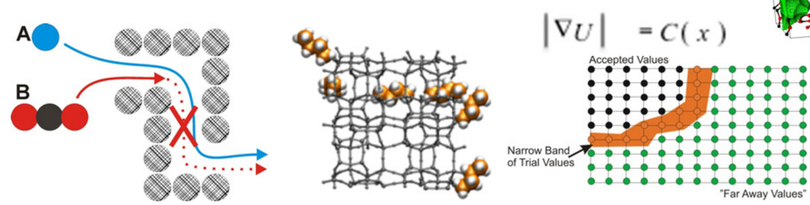
Map building blocks to vertices and edges of topology graph

Final 3-D model exhibits desired topology

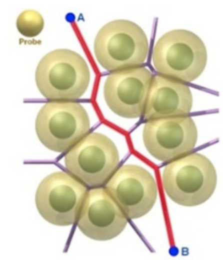
(2) Analyzing Proposed Materials



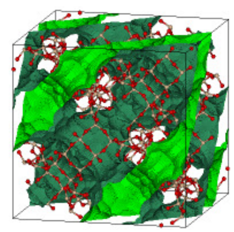
Fast PDE-based algorithms for porosity analysis



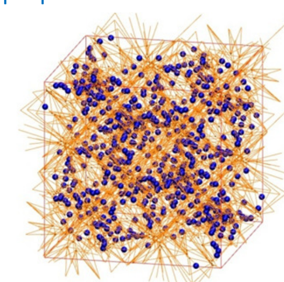
Voronoi decomposition-based tool for characterization of porosity



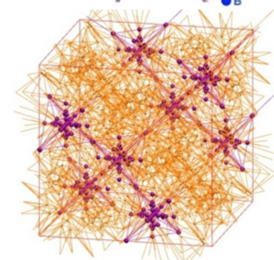
Fast discrete algorithms for calculation of guest-diffusion paths, pore size distributions and other properties with sub 0.1 A accuracy



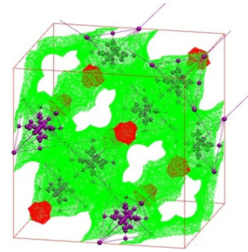
Material structure



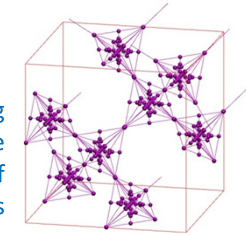
3D Voronoi network



Probe-accessible Voronoi network in pink



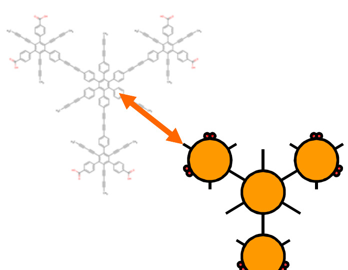
Monte Carlo sampled surface area



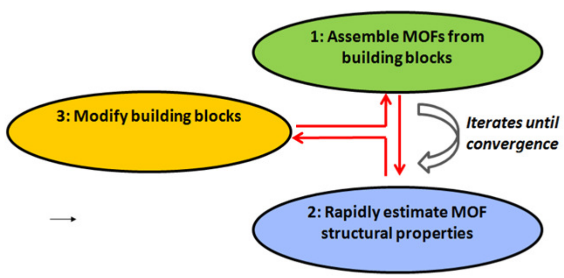
Simplified network representing void space is used for structure similarity and calculation of structural descriptors

(3) Steering the Design

Efficient material design with optimization algorithms



Abstract representation captures shape of real building block of a material



Optimization in abstract space representing a material reveals high-performing material designs

Haranczyk, and Sethian, PNAS, 2009; Willems, Rycroft, Kazi, Meza, Haranczyk, *Microporous Mesoporous Mater.* **2012**; Martin, Haranczyk, *Chem. Sci.* **2013;** *J. Chem. Theory Comput.* **2013**

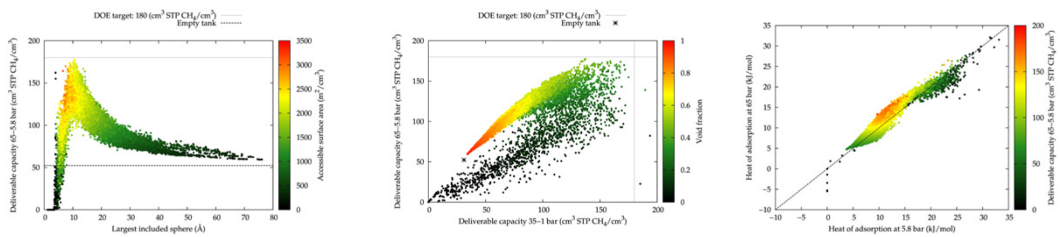
CAMERA: Mathematics and Algorithms for Designing New Advanced Porous Materials

CAMERA

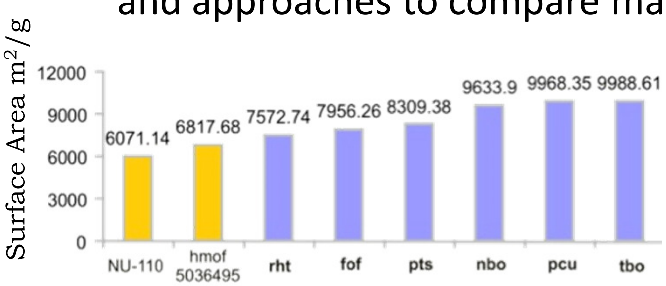
BERKELEY LAB

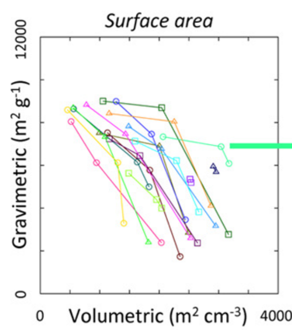
Applications and Early Successes:

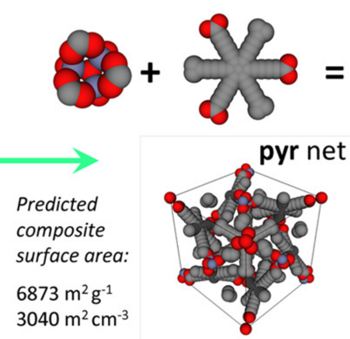
- High-throughput material characterization tools used to analyze porosity in multiple material databases (zeolites, MOFs, PPNs, etc.)
- Material and porosity descriptors used, together with results of molecular simulations, to perform data-mining for structure-property relationships



- Various approaches to material sampling, diversity selection and similarity searching developed and demonstrated
- Demonstrated (Multi-)Objective optimization of porous material structures, and approaches to compare materials topologies







Nature Chemistry (2014)

research highlights

AMORPHOUS SOLIDS

Cage calculations

J. Am. Chem. Soc. http://doi.org/p7q (2013)

Resulting publications:

Cryst. Growth Des. (2014), Phys. Chem. Chem. Phys. (PCCP) (2014), J. Chem. Phys. C (JPCC) (2014), J. Am. Chem. Soc. (JACS) 136 (2014) 5006, *JACS* 136 (2014) 2228, *PCCP* 16 (2014) 5499-5513, *JACS* 135 (2013) 17818, *PCCP* 15 (2013) 20937, *CrystEngComm* 15 (2013) 7531, *JPCC* 117 (2013) 20037, *Crystal Growth Des.* 13 (2013) 4208, *Microporous and Mesoporous Mat. (MMM)* 181 (2013) 208, *J. Mol. Graph. Model* 44 (2013) 208, *JPC C* 117 (2013) 12159, *J. Chem. Theory Comput.* 9 (2013) 2816, *Chem. Sci. 4 (2013) 1781, JACS* 134 (2012) 18940, *ChemPhysChem* 13 (2012) 13, 3595, *Langmuir 28 (2012),11914, Nature Materials* 11 (2012) 633, *J. Chem. Inf. Model.* 52 (2012) 308, *MMM* 149 (2012) 134, *Mol. Sim.* 37 (2011) 986

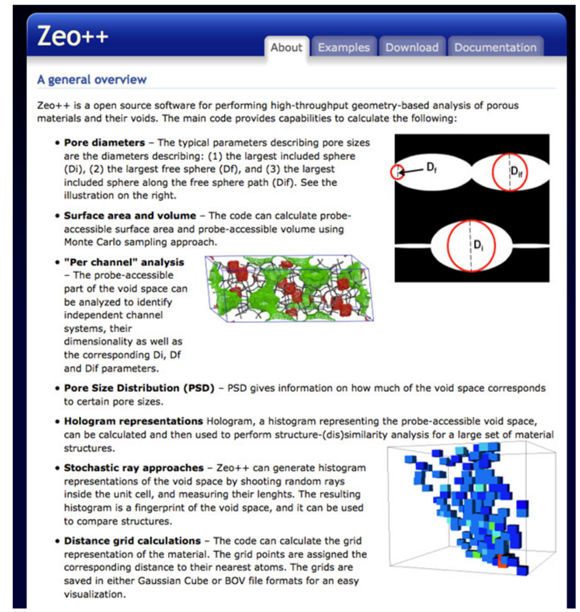
CAMERA: Mathematics and Algorithms for Designing New Advanced Porous Materials

CAMERA

BERKELEY LAB
Lawrence Berkeley National Laboratory

Deliverables:

- Algorithms implemented in an open-source package - Zeo++ - www.zeoplusplus.org
- Zeo++ has been adopted and become a default tool for two BES Materials Genome Centers (Nanoporous Materials Genome Center (Minnesota) and Center for Functional Electronic Materials (LBNL)) and the EFRC for Gas Separations (LBNL)
- Ca. 200 registered users world-wide in both academia and industry (e.g. Bosch, Samsung)
- Initial work on web-interfaces to allow easy access to structure enumeration and analysis capabilities (target users: experimentalists and material designers alike)



http://urchin.lbl.gov:8080/draw

Prototype Web Interface:

Welcome to MOF builder Wizard - Powered by Zeo++

Input SMILES string defining the linker (you can use the drawing tool below to draw your linker and then press "Get SMILES" to paste the string here.

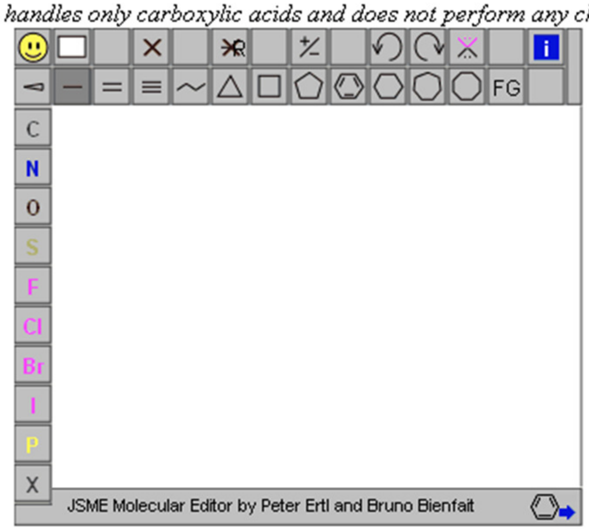
Enter organic linker (SMILES format):

Select metal cluster: [ ] and topology [Please select the metal cluster first]

[Submit Query]

[Get SMILES] [Clear Editor] [Clear Fields]

Note: The current version of the builder handles only carboxylic acids and does not perform any checks (make sure your molecule is valid)!



GISAXS and HipGISAXS

(Grazing-incidence small-angle X-ray scattering)

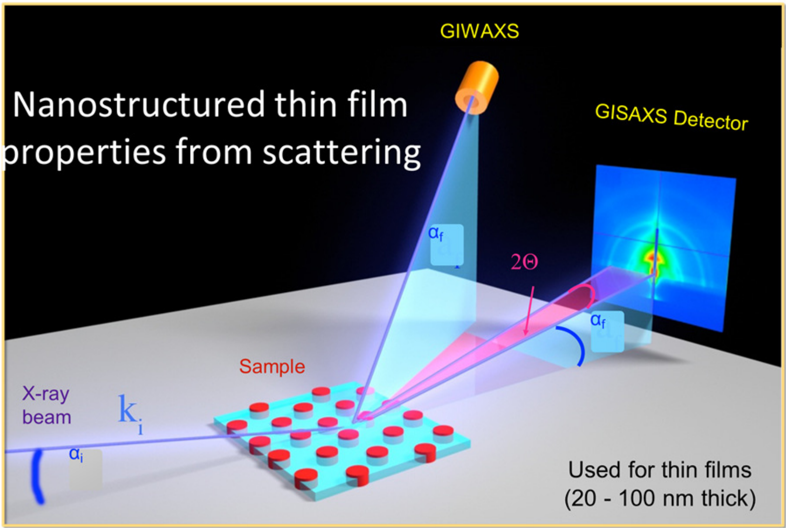
Faster Analysis for X-ray Scattering Data

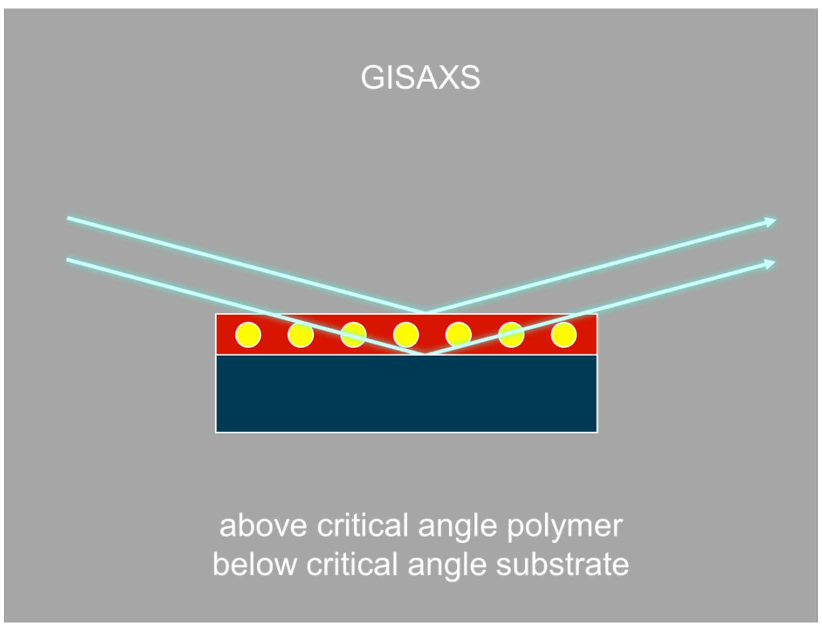
A. Hexemer (Advanced Light Source)
X. Li and C. Yang (Computing Sciences Division)
J. Donatelli, J. A. Sethian (UC Berkeley Math, LBNL Computing Sciences)

GISAXS: Overview

CAMERA

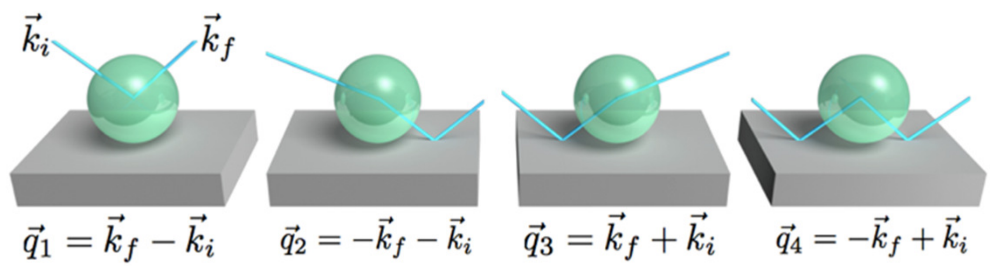
BERKELEY LAB





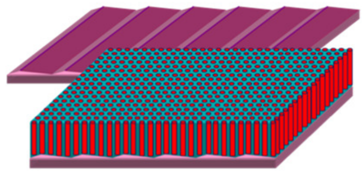
GISAXS nanorod calculation

- Received data not just Fourier transform
- Use distorted wave Born approximation(DWBA)
  Treat scattering by nanorods as perturbations of incident, reflected, and refracted scattered waves

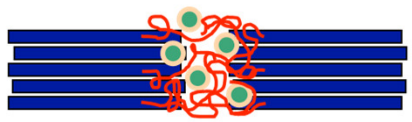


GISAXS: Sample Uses

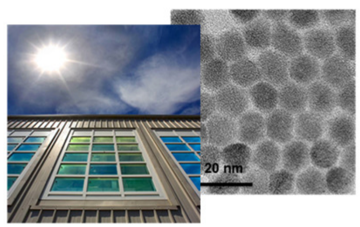
Long-range ordering of block copolymers for dense storage media
(Russell, UMass Amherst, Xu, UCB/MSD, A. Hexemer LBNL)



Nanoparticle/polymer composites for solar cells
(Segalman, UCB/MSD & Urban, TMF)

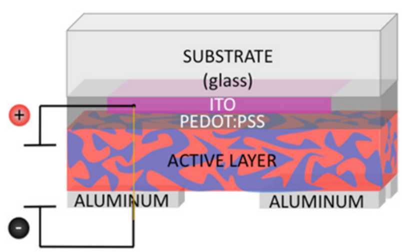


Electrochromic windows
(Milliron, TMF)

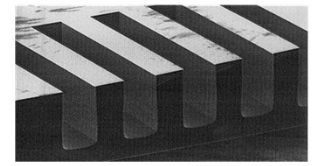


Battery electrolytes
(Balsara, UCB/MSD/EETD)

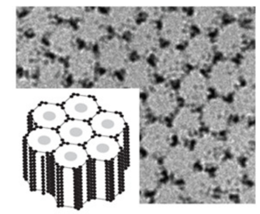
OPV BHJ materials
(McGehee, Stanford; Toney, SSRL/SLAC; Gomez, PSU; Kline, NIST; Liu, TMF; Ade, NCSU; Kramer, UCSB; Russell, UMass Amherst; Amassian, KAUST, A. Hexemer LBNL)



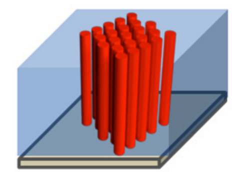
Lithographic patterning
(Soles, NIST; Ocko, BNL)



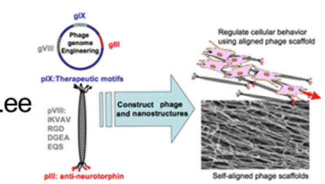
Self-assembly of nanoparticles in block copolymer thin films (Xu, UCB/MSD)



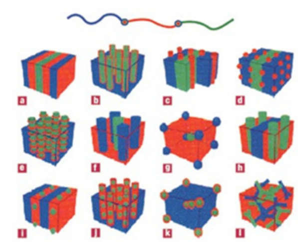
Composite membranes for artificial photosynthesis
(Segalman, UCB/MSD)



Virus nanofiber tissue engineering materials (Lee UCB/PBD)



Block copolymer self-assembly
(Kramer, UCSB; Russell, UMass Amherst; Xu, UCB/MSD)



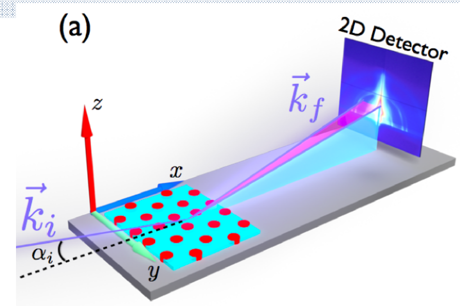
GISAXS: Mathematical/Computational Issues

CAMERA

Forward Simulation:

Design input structure

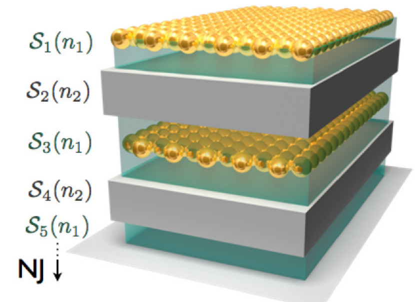
→ compute scattering pattern



Next Generation Computing for X-ray Science

- **HipGISAXS simulation code based on DWBA**
- **High Performance Parallel Code**
- **Orders of magnitude faster than before**
- **Resulted from designing the mathematics to parallelize the algorithms.**

Sample Structure

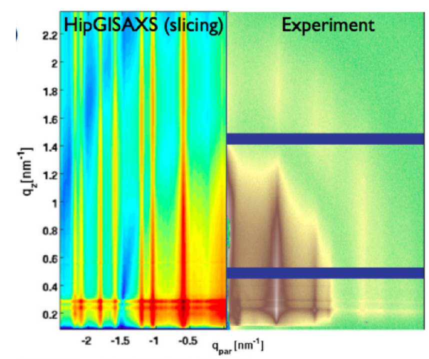


GISAXS (forward simulation):

simulation: 20 sec → **0.05 sec** /frame

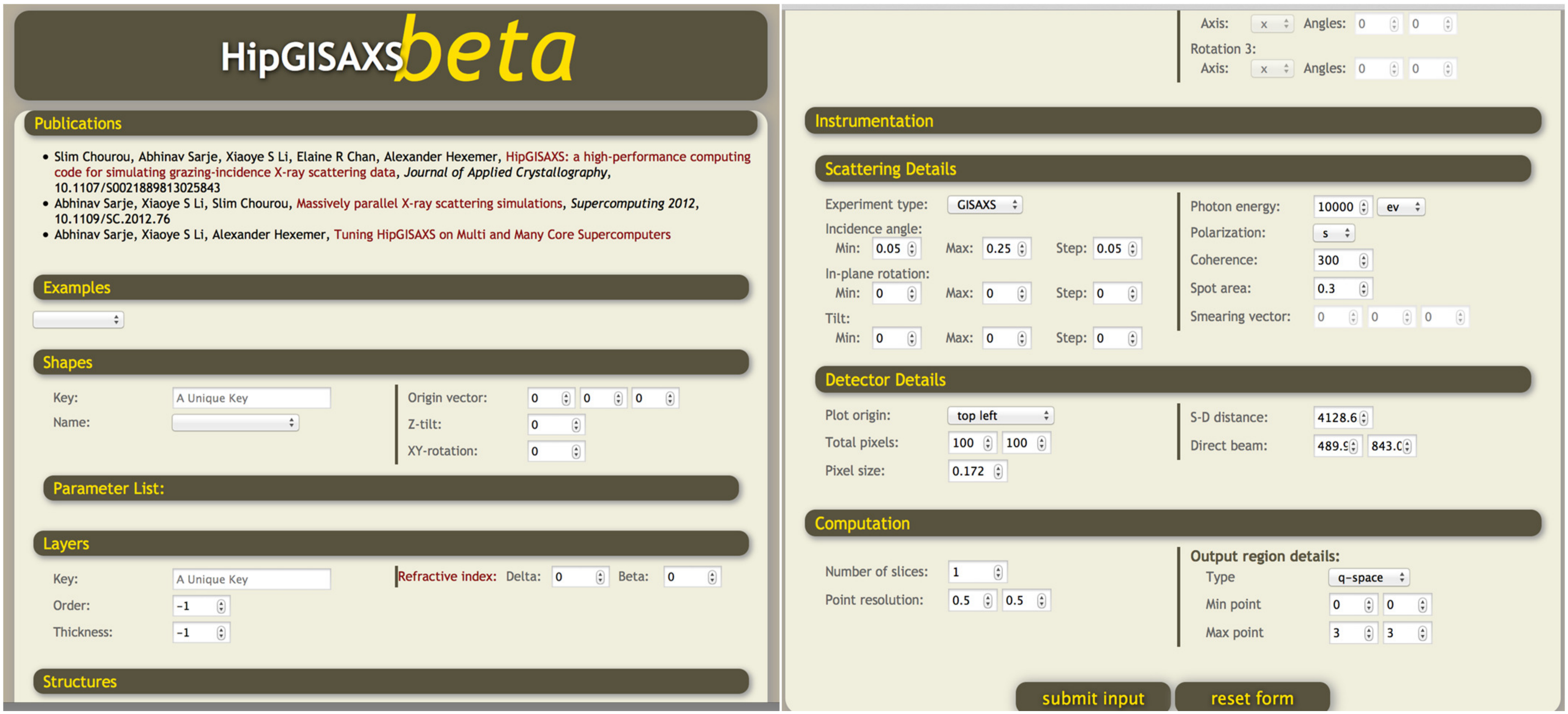
before after

GISAXS Image



(In progress: allow "CAD" input: building high order integration of form factors)

GISAXS: Deliverables

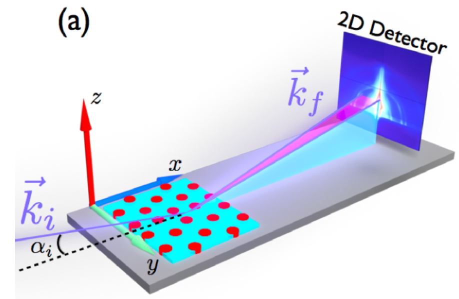


GISAXS: Mathematical/Computational Issues

Backward Simulation:

Given scattering pattern

→ what was the structure?



Current "beta" approach: Reverse Monte Carlo (recover structure):

1 frame in 240 min → **100 frames in 15 min**

before after

Mathematical Questions:

- Not unique: is there theory to be developed within class constraints?
- Can you guide toward plausible solutions with constrained optimization?
- Can you use machine learning, coupled to image/pattern analysis to help steer?
  (identify peaks from image analysis and help identify crystal structure?)
- Better optimization methods: (non-linear methods, genetic algorithms)

Software: Can we expand the software to work better with other techniques, such as TEM, SEM maybe Tomography ?

Reconstruction Algorithms for X-Ray Nanocrystallography

J. Donatelli (LBNL Math)
J. A. Sethian (UC Berkeley Math, LBNL Computing Sciences)

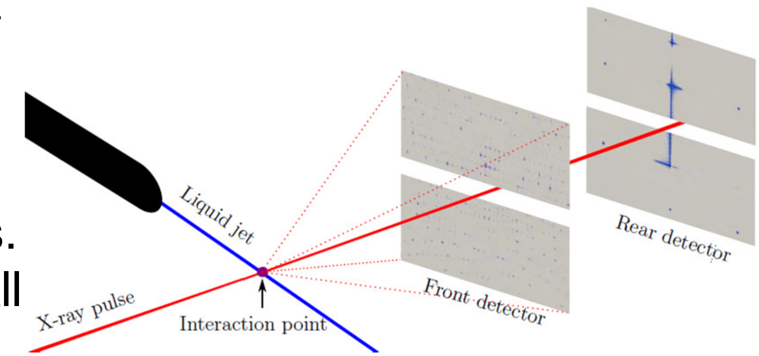
Reconstruction Algorithms for X-ray Nanocrystallography

CAMERA

BERKELEY LAB

X-ray nanocrystallography allows structure of a macromolecule to be determined from a large ensemble of nanocrystals.

Each nanocrystal is destroyed during the process. The goal is to assemble diffraction images from all the crystals to determine structure



X-ray Nanocrystallography Experimental Setup

Central issues:

Several parameters, including crystal sizes, orientations, and incident photon flux densities, are initially unknown.

Additionally, images are highly corrupted with noise.

Tremendous amount of advanced mathematical algorithms for these problems:

John Spence's group, Kay Diederich's group.

Large amount of working software in use every day.

Reconstruction Algorithms for X-ray Nanocrystallography

CAMERA

BERKELEY LAB

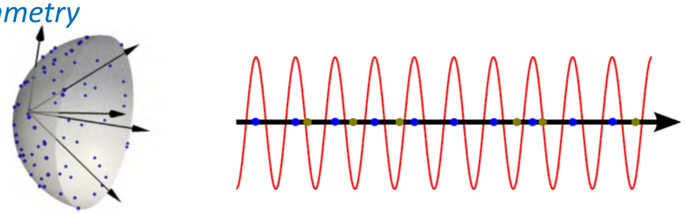
We're trying to bring some new math to these challenges:

Computational harmonic analysis, graph clique analysis, dimensional reduction, compressed sensing, multi-modal expectation, PDE segmentation…

Hope: Use fewer images, deal with more noise, harder cases

(1) Autoindexing

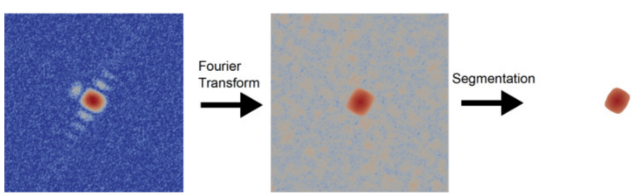
Techniques for orienting images up to crystal lattice symmetry



Periodicity analysis of reflections yields partial orientation information

(2) Peak Shape Analysis

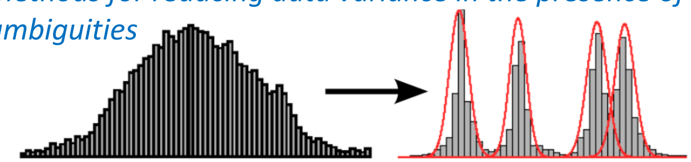
Tools for determining crystal shape and size



Fourier analysis of finely sampled low angle data coupled with image segmentation reveals crystal features

(3) Multi-Modal Modeling

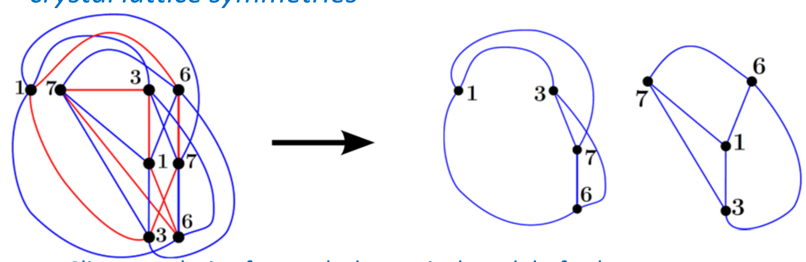
Methods for reducing data variance in the presence of indexing ambiguities



Multi-stage expectation maximization/scaling locates histogram modes from ambiguously oriented data

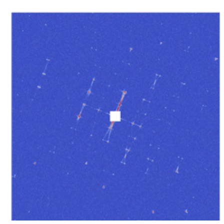
(4) Resolving Indexing Ambiguities

Algorithms for removing orientation ambiguities resulting from crystal lattice symmetries

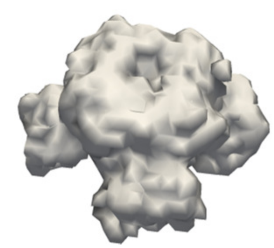


Clique analysis of a graph theoretical model of value concurrency resolves indexing ambiguities

Structure determination of puuE allantoinase from simulated data



Simulated diffraction pattern



Reconstruction via iterative phase retrieval

Donatelli and Sethian, An algorithmic framework for x-ray nanocrystallographic reconstruction in the presence of the indexing ambiguity, PNAS, 2013 (to appear)

CAMERA: What is Happening Now

BUILDING, TESTING, EXPORTING:

•Building the new mathematics required to partner with experimental facilities

•Together with experimental partners, testing algorithms on data and "on the shop floor"

•Exporting codes, software to other DOE Facilities, Labs, and to advanced computing environments (HIPGISAXS, MicroCT, PytchoPS, PEXSI, ZEO++,…)

camera.lbl.gov

CAMERA: Future?

CAMERA

Current Status:

LBNL LDRD: Expires in one month (35 out of 36 months completed)
ASCR-BES Pilot Project: Expires in 13 months (10 out of 23 months completed)

Immediate Tasks:

Supporting on-going software development/delivery to wider DOE community.
Bring current projects into deliverables

Tremendous Opportunities and Impact:

Small DOE investment in mathematics/algorithms. .
Gains in productivity could be massive, orders of magnitude more than now.

Expand the range and influence:

More beams, facilities: (inversions, reconstructions, segmentations, extractions)
More nanofacilities (high throughput, computational speedups)
More experimental partnerships (lab/faculty/postdocs/grad students)
Aim the next generation of ASCR mathematicians at these problems.

CAMERA

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, most users need not worry about becoming mathematicians, and can instead just *use* tools that transforms their data into the information they really want.

U.S. DEPARTMENT OF ENERGY | Office of Science

camera.lbl.gov

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory