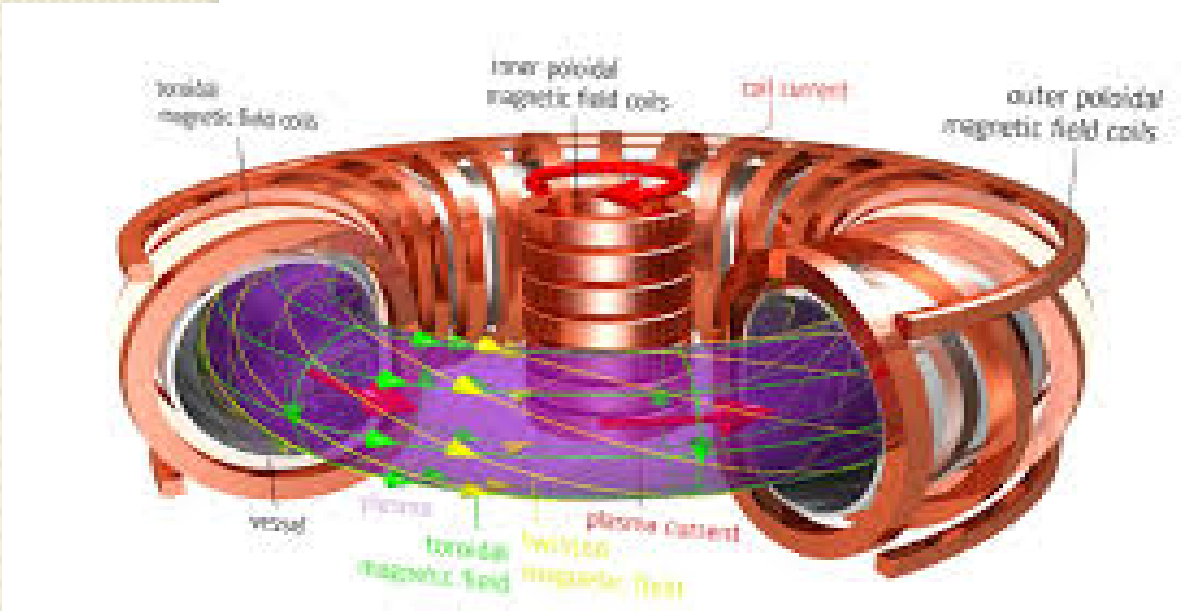
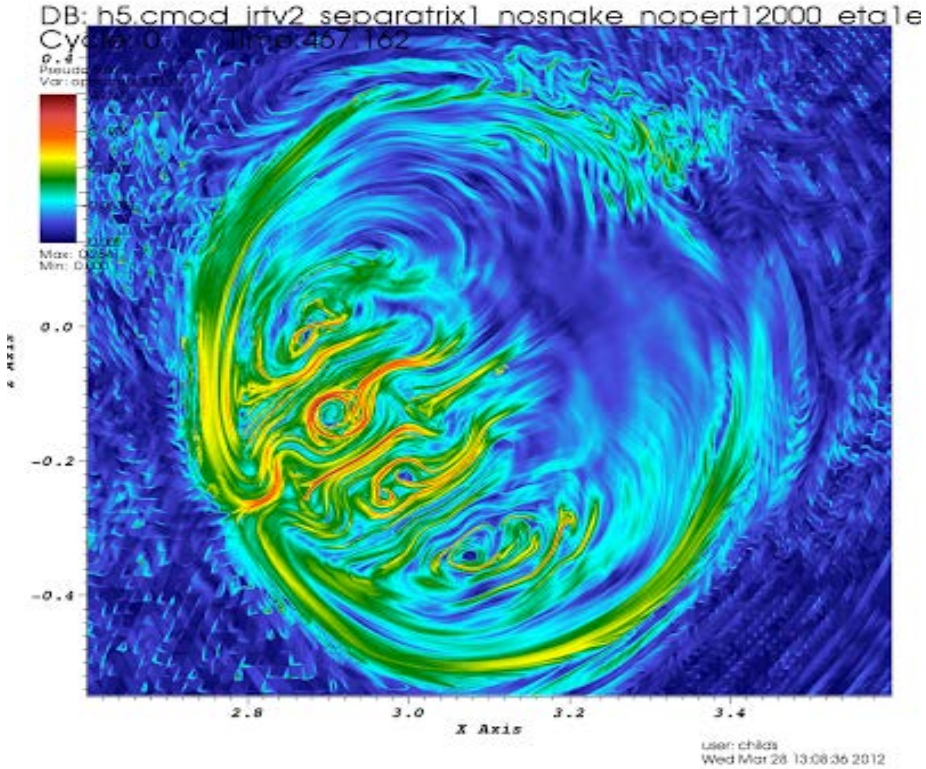
Understanding Scalable Realtime Collaborative Workflows

Hari Krishnan, Lawrence Berkeley National Laboratory, Computational Research Division.

Research at the Lab

Fusion - Relationships between magnetic and velocity fields in a tokomak.

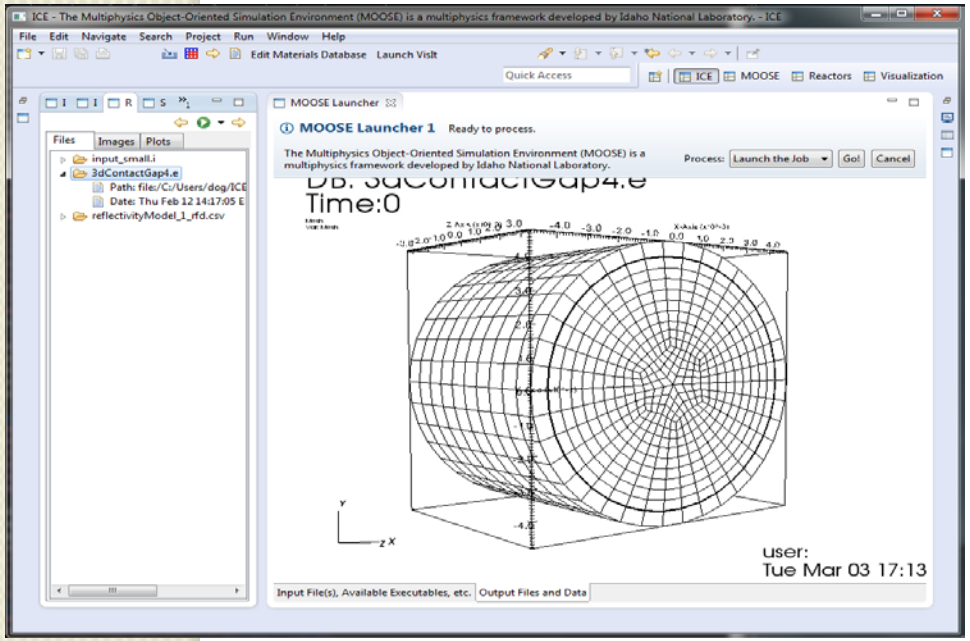


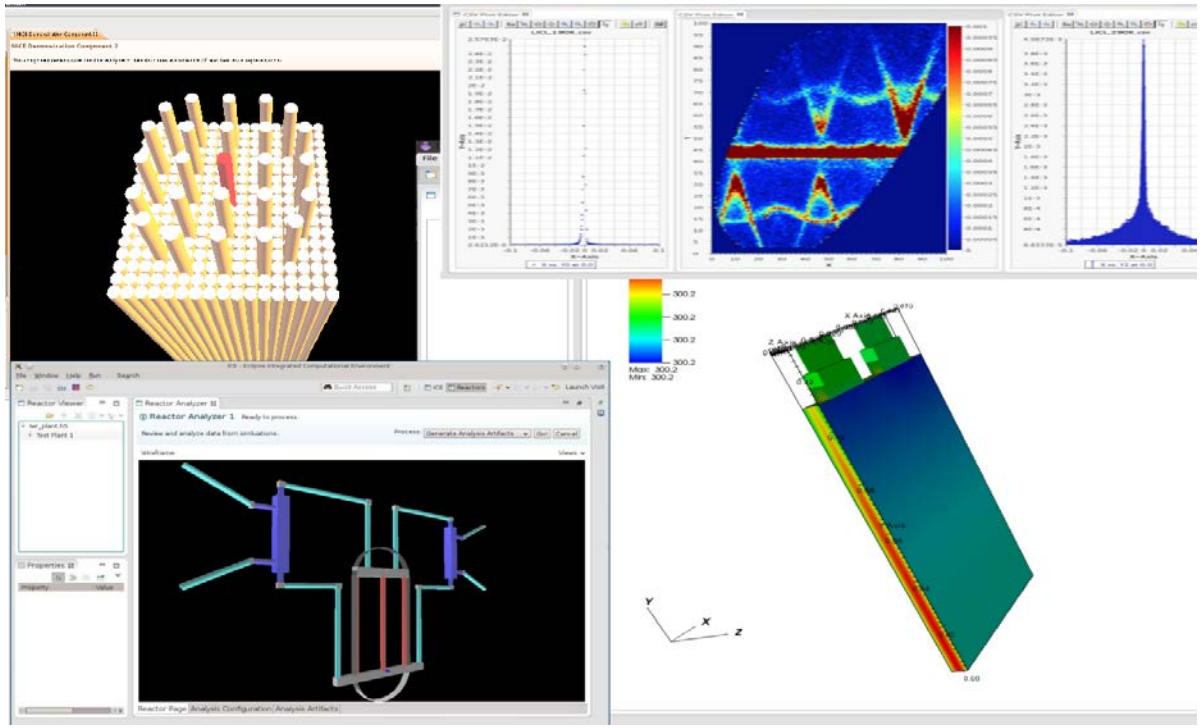


http://adsabs.harvard.edu/abs/2012APS..DPPYP8009S

Research at the Lab

Nuclear Energy - Modeling a Nuclear Power Plant from pellet to plant

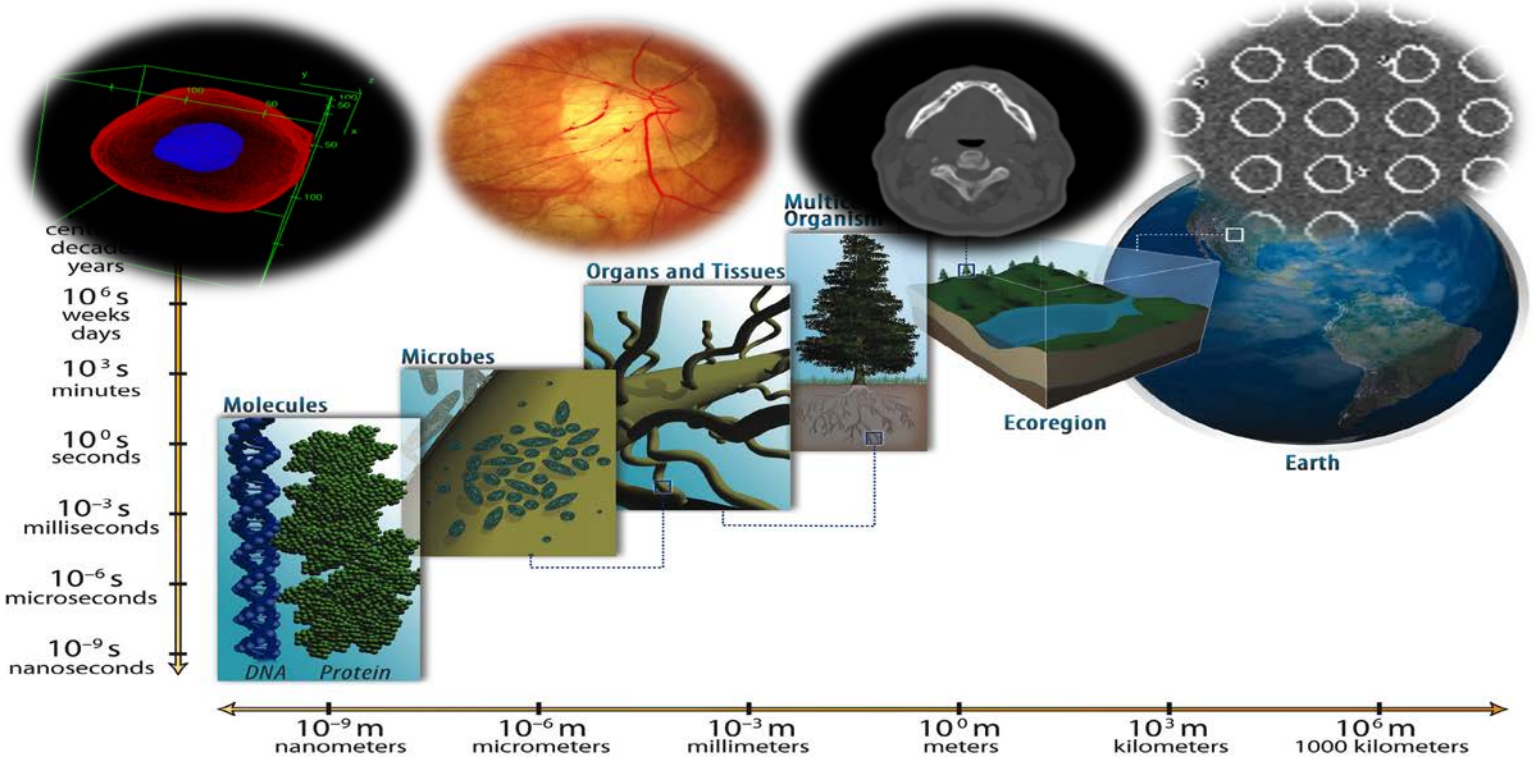




https://www.eclipse.org/community/eclipse_newsletter/2015/january/article1.php

Research at the Lab

Understanding Biological, Chemical, and Material Properties

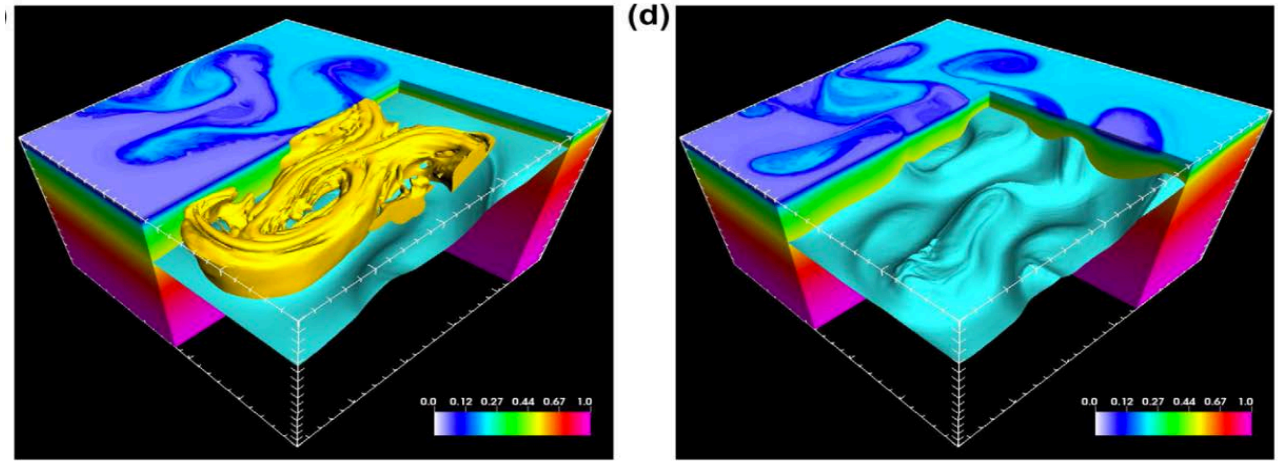


https://arxiv.org/abs/1602.01448
http://ieeexplore.ieee.org/document/7004292/

Research at the Lab

Ocean Modeling (Visualizing Oil Dispersion) - Deep Water Horizon Oil Spill in Gulf of Mexico

Study of currents in Gulf of Mexico



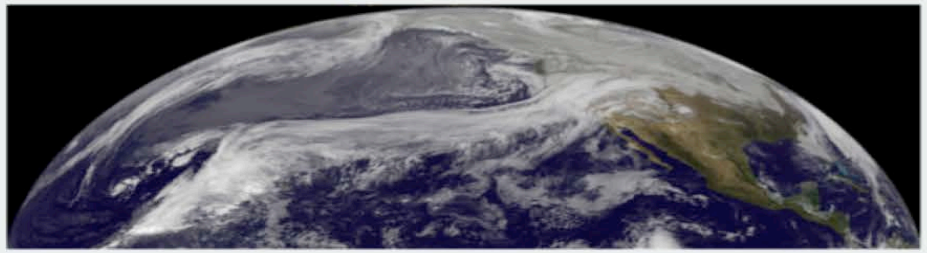
Distributed Finite-Time Lyuponov Exponent Computation

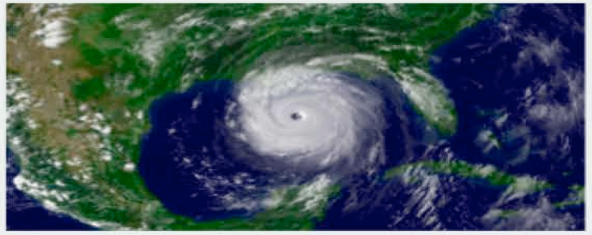
http://cs.lbl.gov/news-media/news/2012/visualizing-oil-dispersion/
http://www.rsmas.miami.edu/users/tamay/ftp-pub/omod12b.pdf

Research at the Lab

Extreme Climate Event Detection - Hurricane, Tropical Cyclone, Atmospheric Rivers Detection, etc…

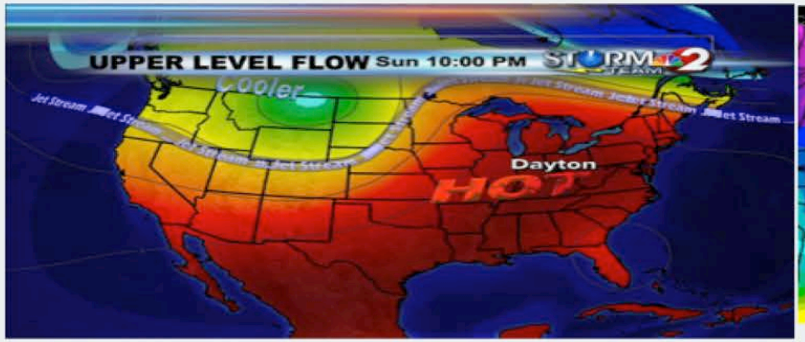
Atmospheric Rivers - NASA





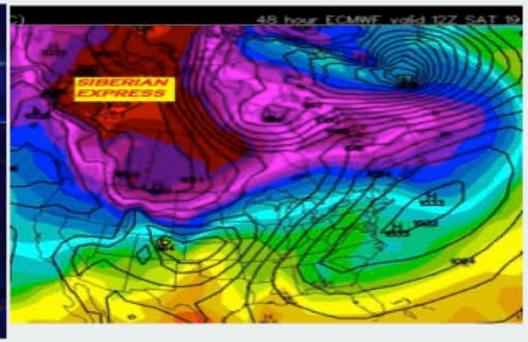
Hurricanes (Tropical Cyclones) – NASA/NOAA

WDTN Dayton, Ohio

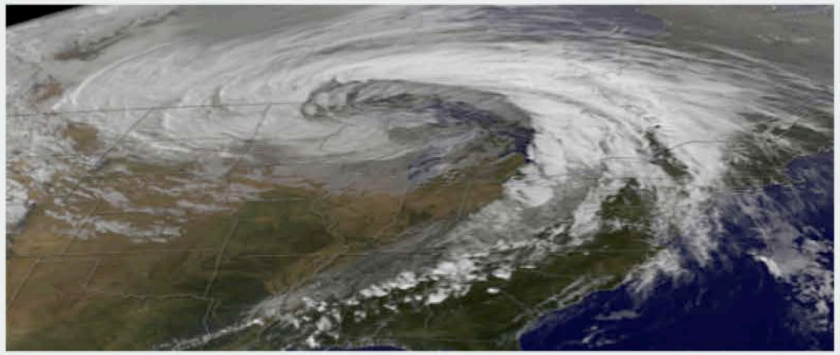


Heat waves and droughts – WDTN (Dayton, OHIO)

ECMWF - Weather underground



Cold Snaps – ECMWF Weather Underground



Extra Tropical Cyclones – NASA/GOES

http://www.sciencedirect.com/science/article/pii/S1877050912002141

Big Data Challenges

- Cataloging the universe & determining the fundamental constants of cosmology
- Characterizing extreme weather events in a changing climate
- Extracting knowledge from scientific literature
- Investigating cortical mechanisms for speech production
- Google Maps for Bio-Imaging
- Perform extreme scale genome assembly
- Precision toxicology
- Seeking designer materials
- Determining the fundamental constituents of matter

https://www.oreilly.com/ideas/big-science-problems-big-data-solutions

What do many large DOE Projects have in common.

- Multi institutional (just a few)
  - Labs: LBNL, LLNL, PNNL, LANL
  - Facilities: ALS, BNL, SLAC (SSRL & LCLS), NSLS2
  - Sites: Hanford (Washington), F-Area (Savannah River)
  - Resources: NERSC, ORNL, SDSC, TACC

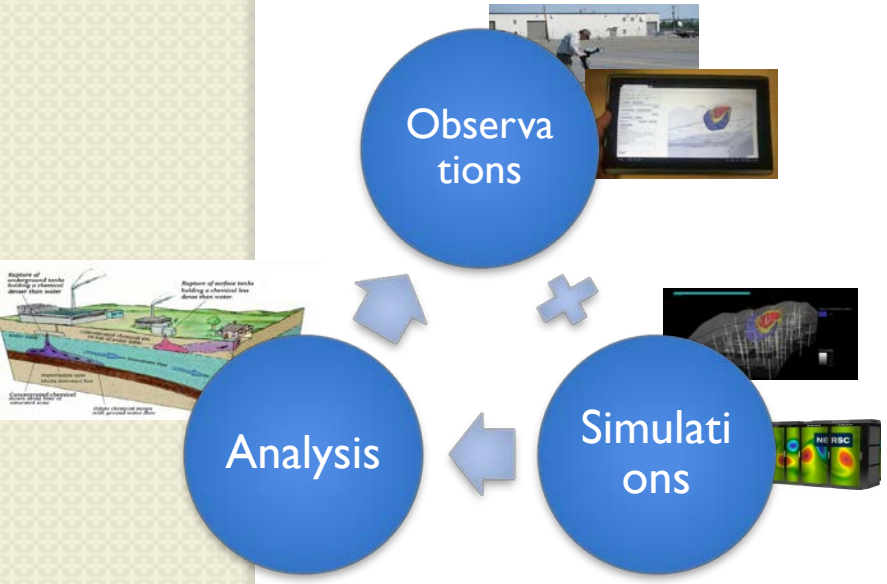
- Expertise from several domains working together.
  - Domain Scientists, Physicists, Mathematicians, Statisticians, Engineers
  - Research Focused (Fair amount of software development)
  - Complex workflow – Highly specialized hardware and custom software.

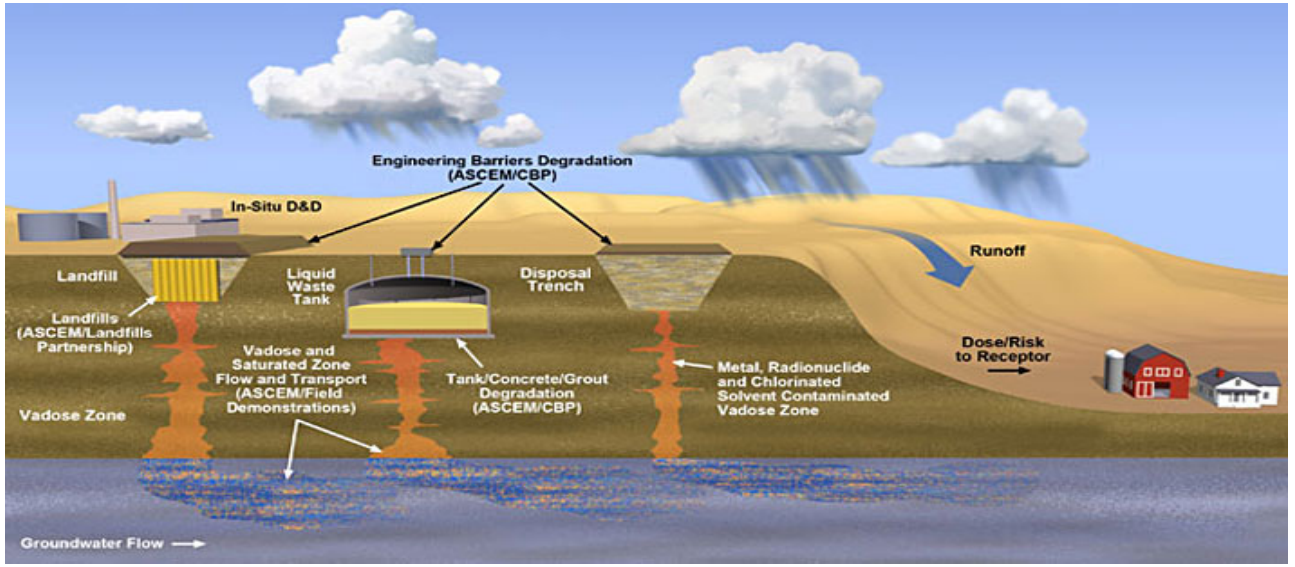
The rest of the Talk will delve into two specific projects

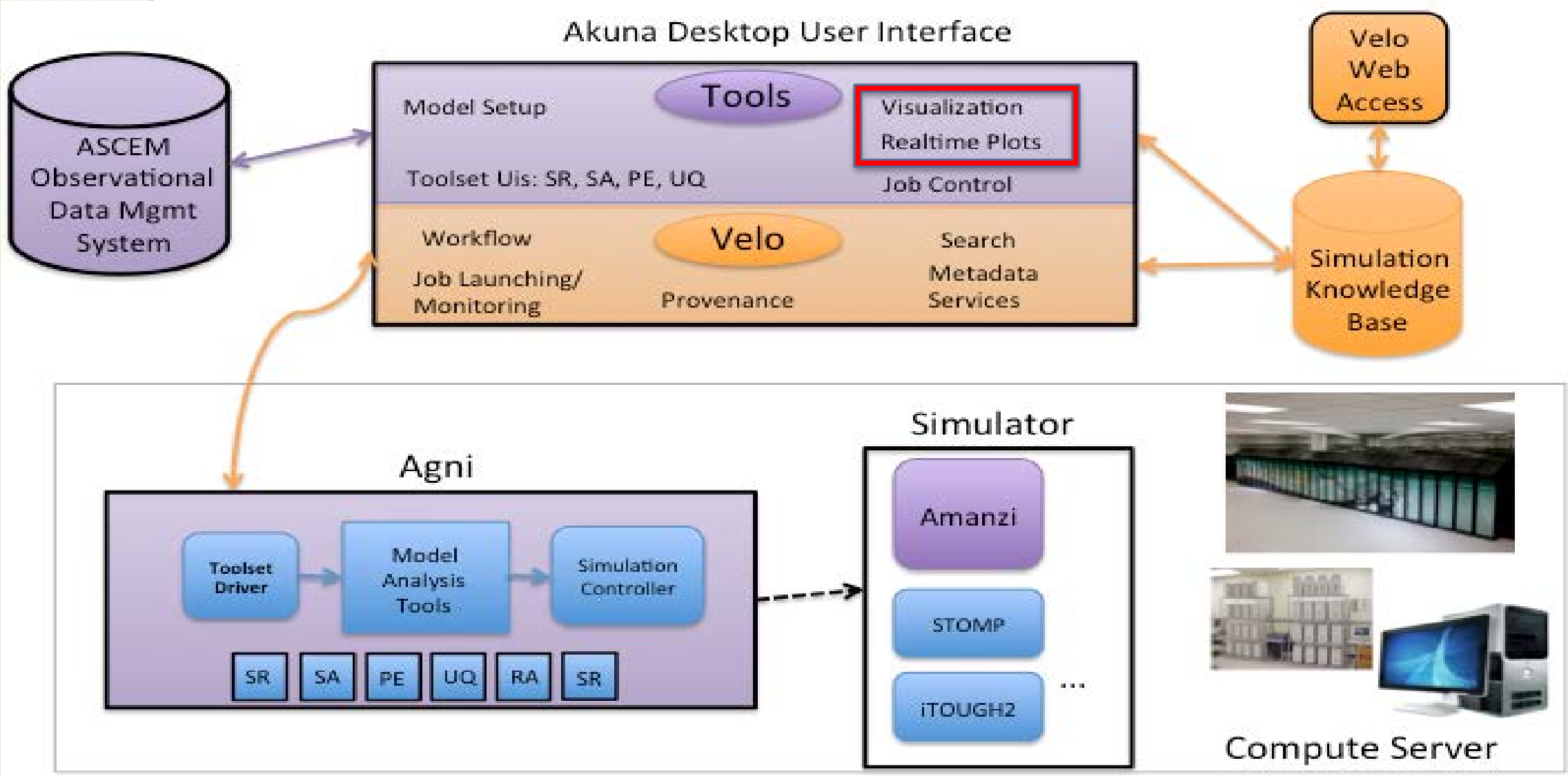
Science Use Case #1: Environmental Management (Macro)

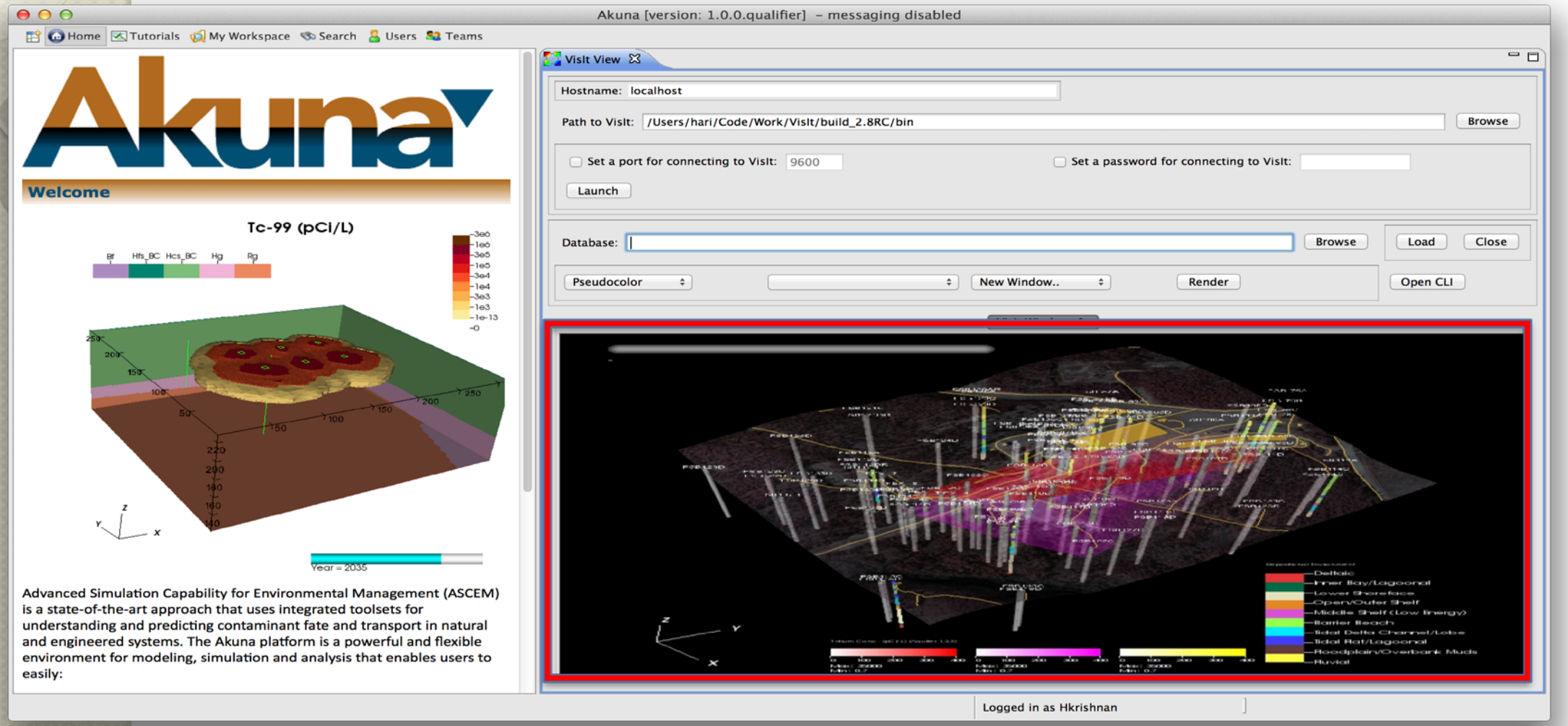
- Understand Cleanup efforts at the Hanford & F-Area Savannah River Sites.
  - Hanford - the first full-scale plutonium production reactor in the world.
  - F-Area (Savannah River) – Site for refinement of nuclear materials

- Create a process combining combining strengths of observed data, modeling, analysis, and simulation to gain insight.





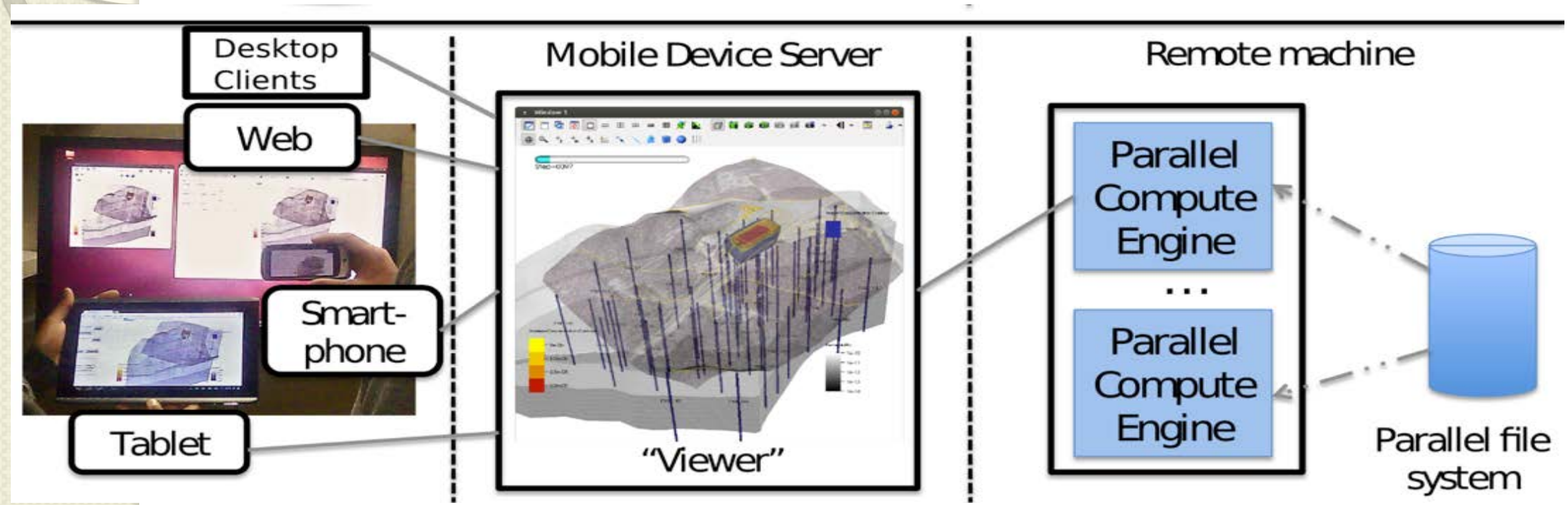

Akuna Desktop User Interface
ASCEM Observational Data Mgmt System
Model Setup
Tools
Visualization
Realtime Plots
Toolset Uis: SR, SA, PE, UQ
Job Control
Workflow
Velo
Search
Job Launching/ Monitoring
Provenance
Metadata Services
Velo Web Access
Simulation Knowledge Base
Agni
Toolset Driver
Model Analysis Tools
Simulation Controller
SR
SA
PE
UQ
RA
SR
Simulator
Amanzi
STOMP
iTOUGH2
...
Compute Server



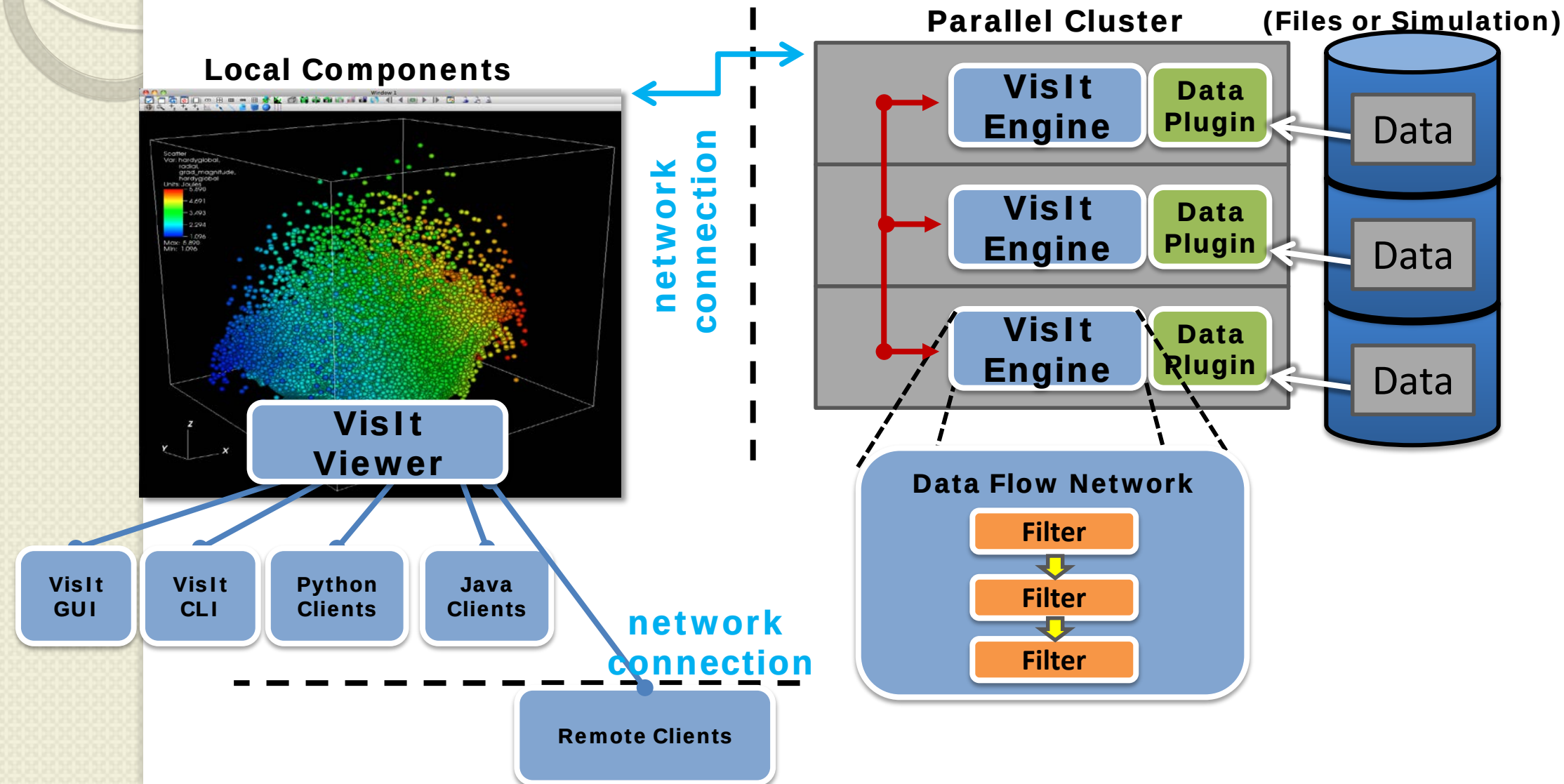
Java Eclipse application al

Provides Model-Setup, Inverse Parameter Estimation, UQ, Remote Job Launching & Monitoring of Simulations, and Visualization.

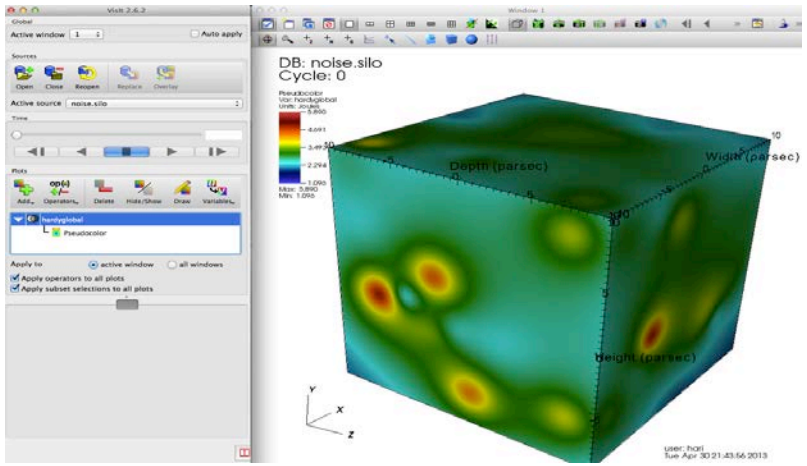
Vislt visualization framework



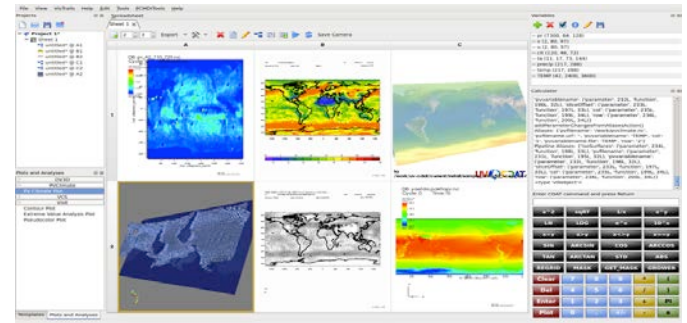
VisIt visualization framework



VisIt: Customizable Interfaces

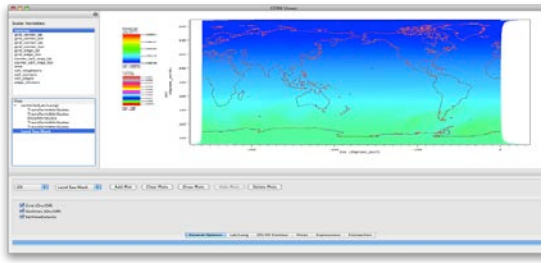


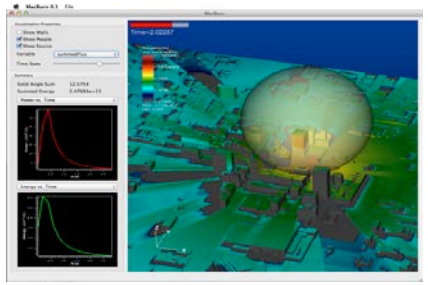
Embedded

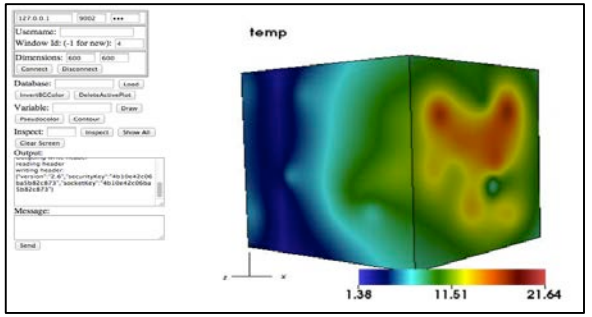


Lightweight, Collaboration

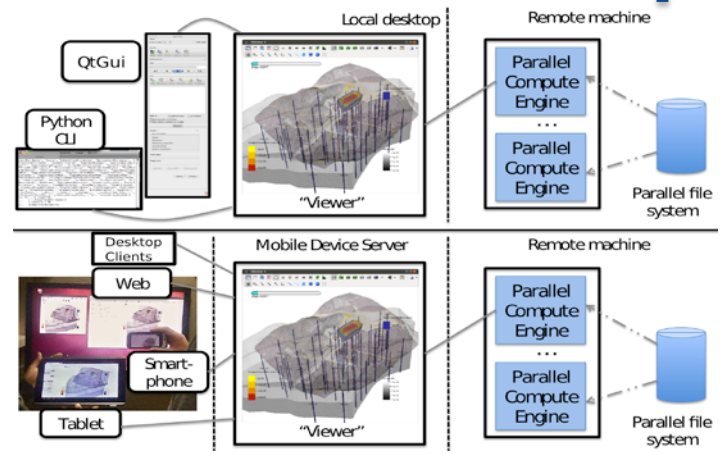
Tailored Vis



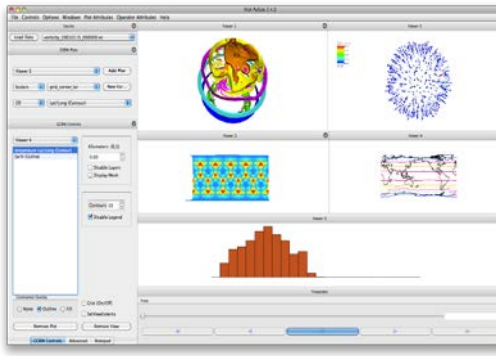


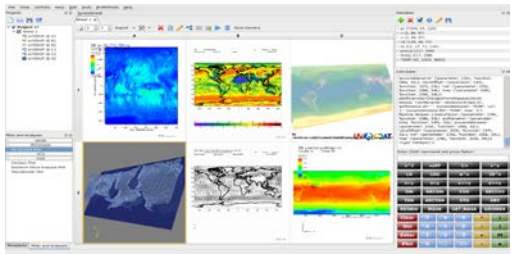


VisIt: Collaborative Capabilities

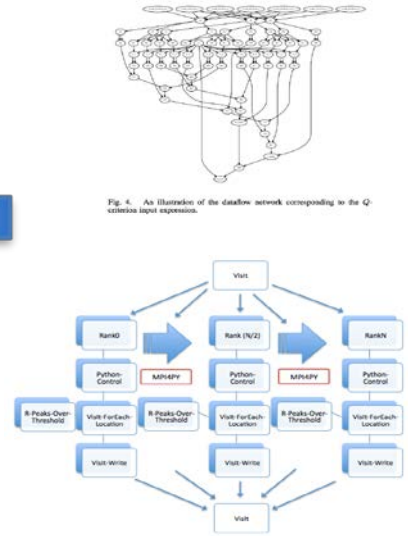


Custom UI

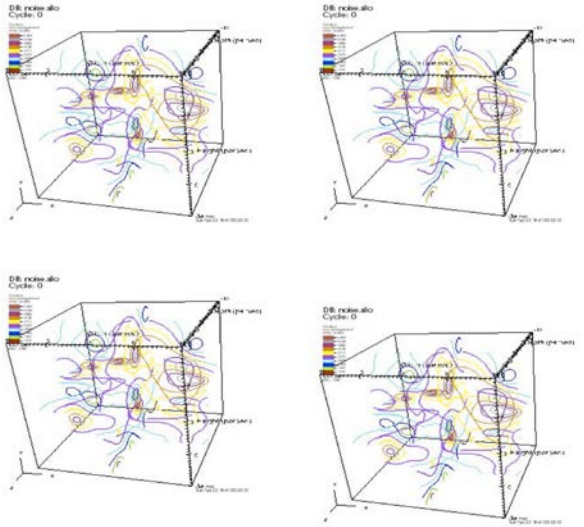




Domain Processing



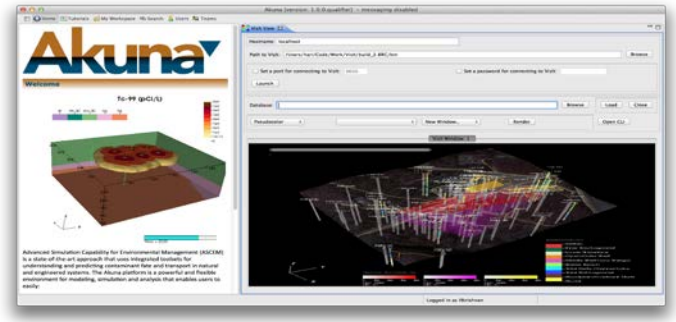
Collaborators



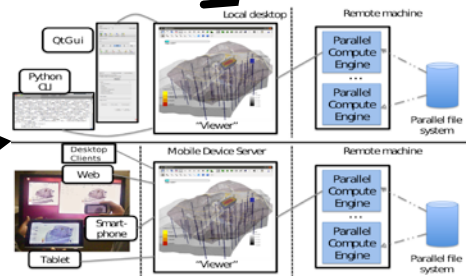
Visualization Services

Provenance
Data Storage

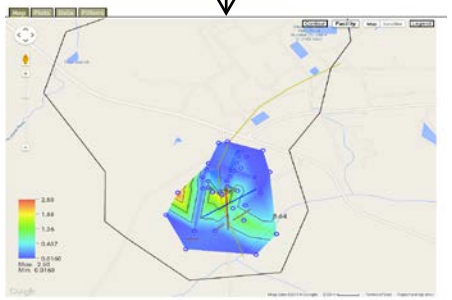
3D visualization



ASCEM Data Browser



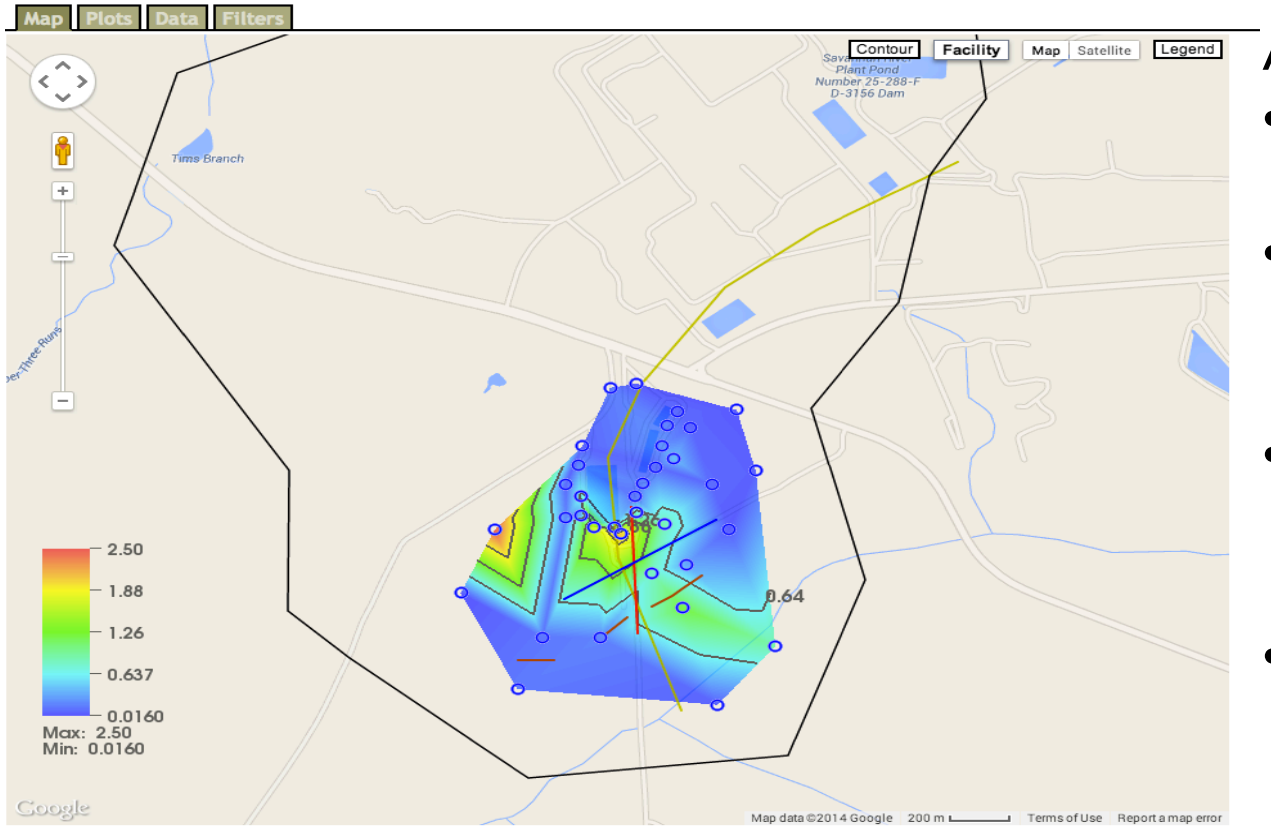
Visualization Service



2D visualization

http://sti.srs.gov/fulltext/SRNL-STI-2015-00027.pdf

2D Visualization (F-Area)



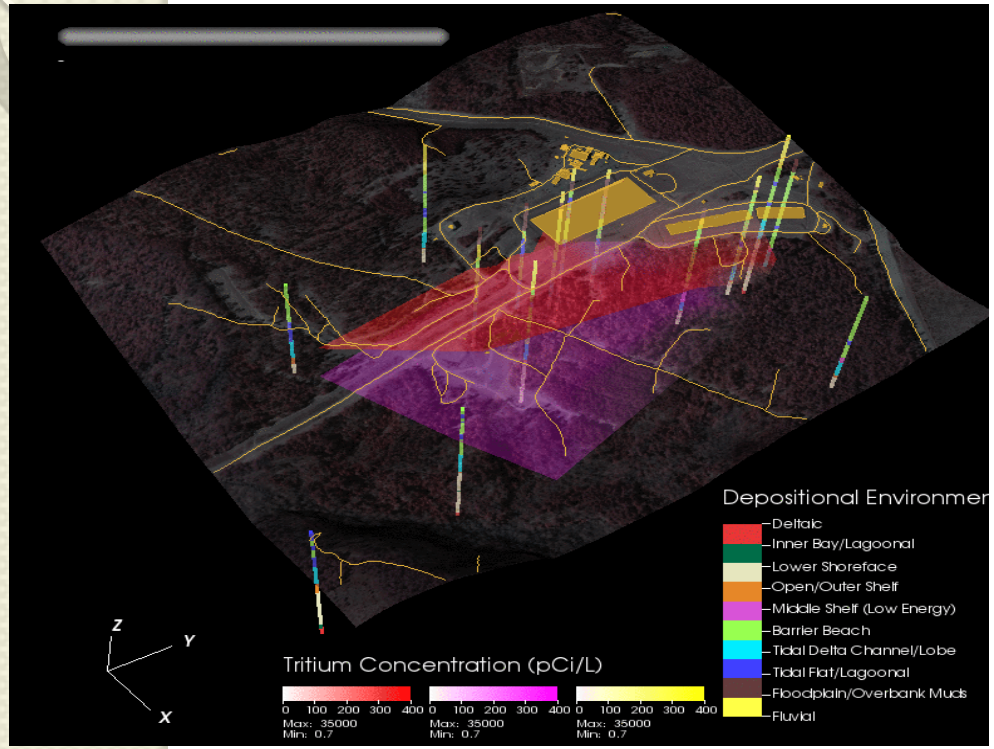
ASCEM Data Browser

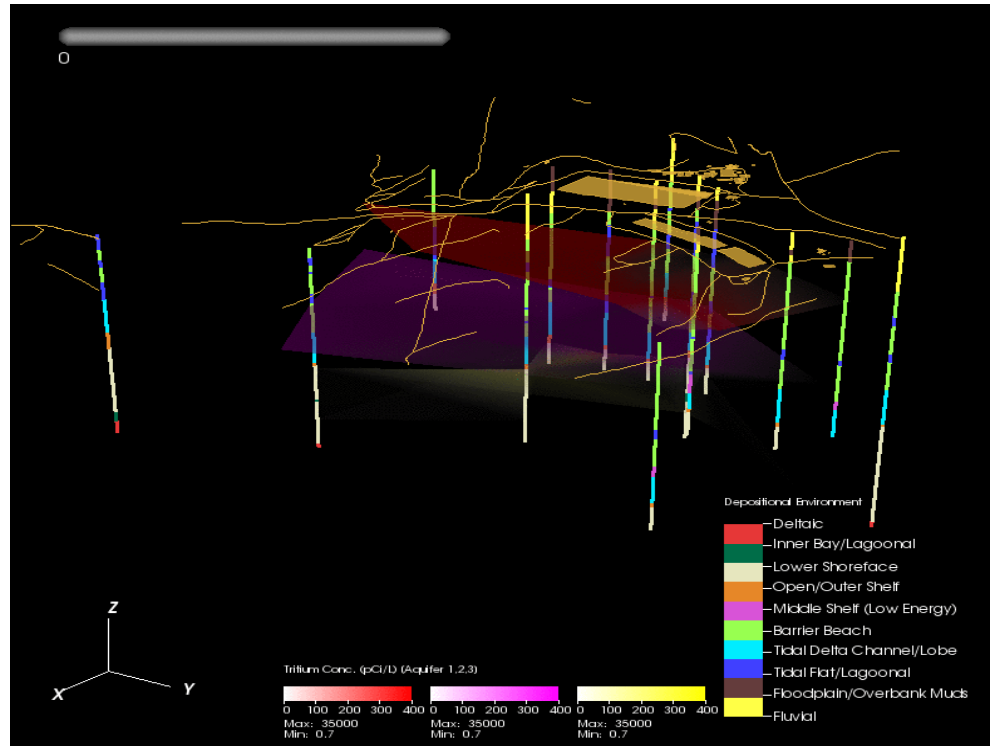
- Google Map Overlay
- Query by: Aquifer Zone, Analyte, and Year
- Contours of concentration levels
- Time-varying data

http://babe.lbl.gov/ascem/maps/SRDataBrowser.php

Tritium Concentration 1996-2011 (F-Area)

3D Visualization (F-Area)





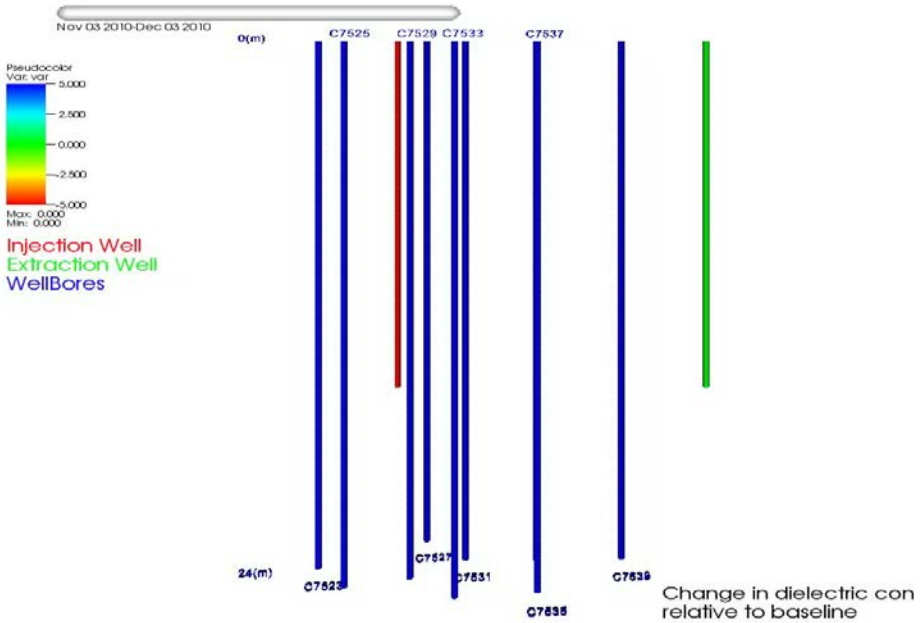
Evolution of Tritium Concentration from 1990-2009

Time Sliders
Depositional Environment
All Aquifer Layers

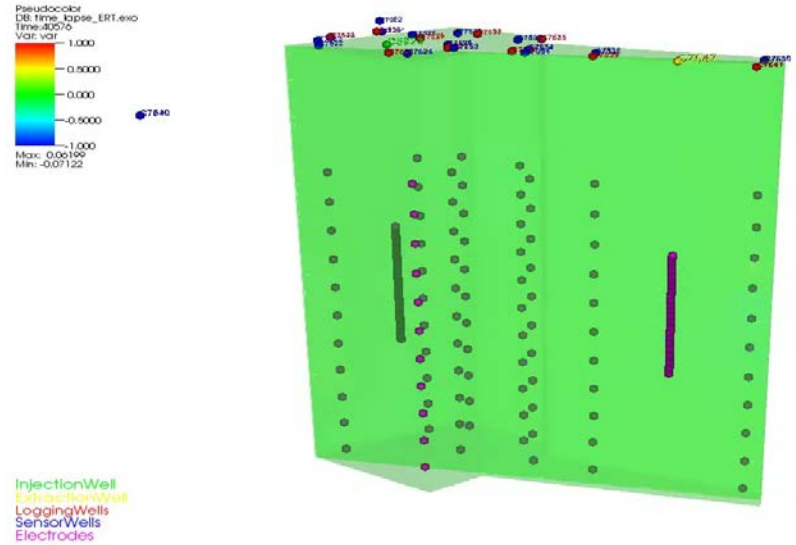
Context: Overlay, Well Sites, Legend, Concentration Levels, Contours/IsoSurfaces

3D Visualization (Hanford Site)

Ground Penetrating Radar

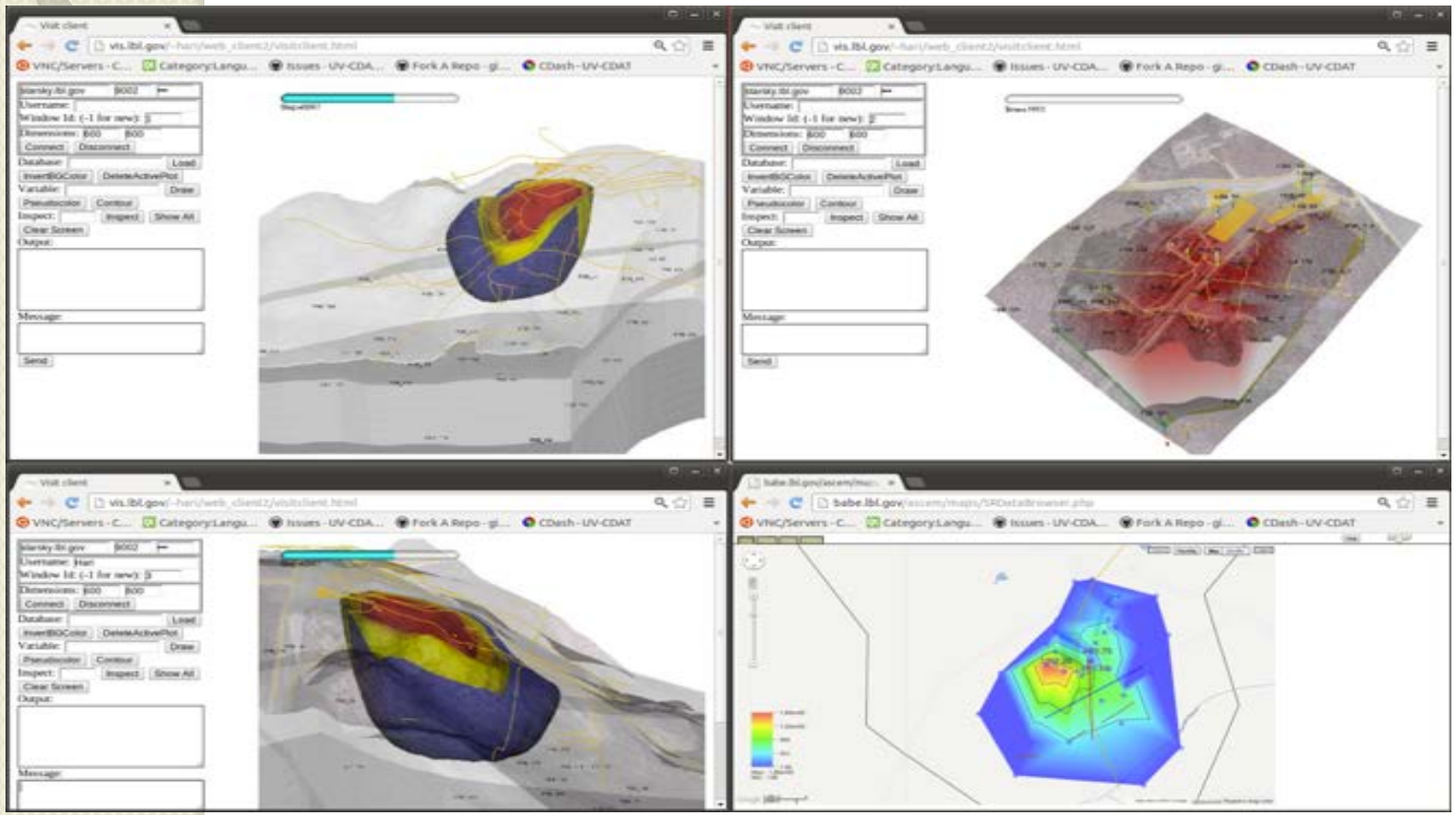


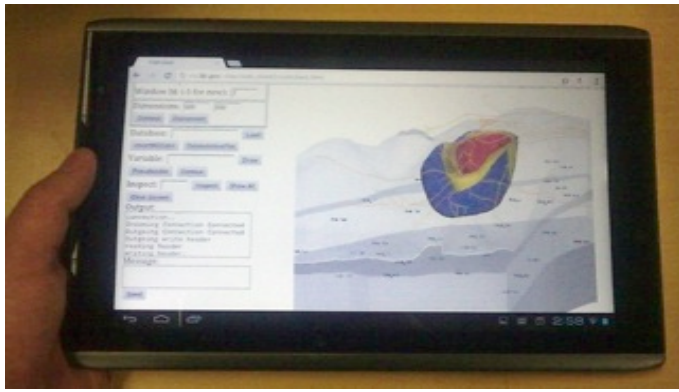
Simulation



Observation vs Simulation

Domain Centric Collaborative Visualization





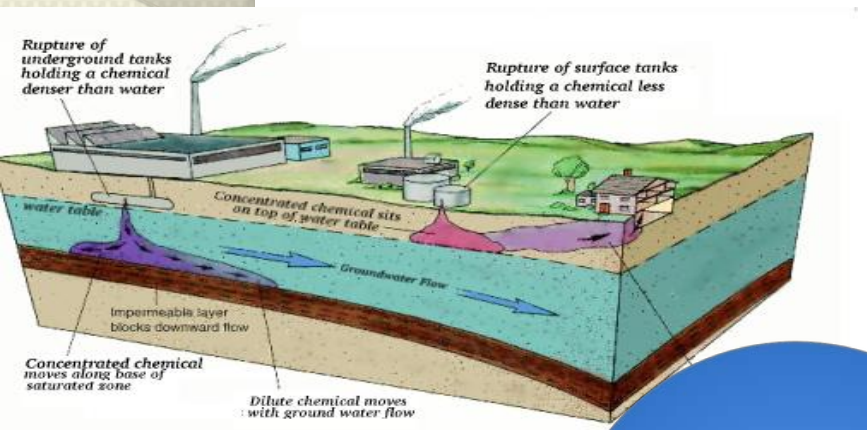
2D Visualization

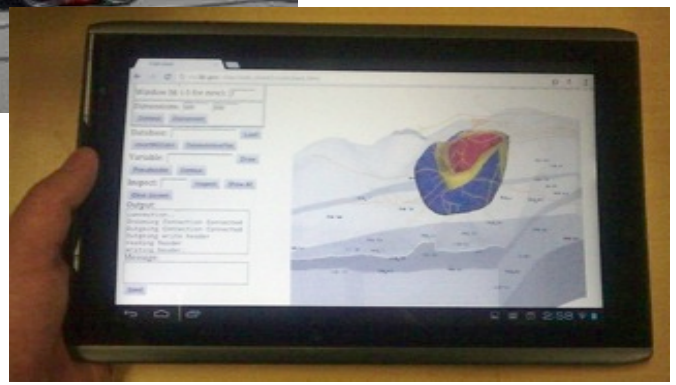
- Google Map API
  - Intuitive, Easy to use, Familiar, Powerful

- Delaunay Triangulation Overlay (VisIt-backend)
  - Shows concentration levels
  - API allows for Custom Color-maps and Concentration levels
  - Temporal view provides powerful and intuitive understanding of concentration levels over time. (Impact of proposed mitigation solutions)

3D Visualization

- Interactive – Supports visualization of multiple layers
- Visually coherence
  - Sensors, Injection + Logging Sites, Well Bores, Image Overlays
- Provides easy to use spatial + temporal visualization
- Visual Comparisons:
  - Same information different sources.
  - Observed and simulated data.

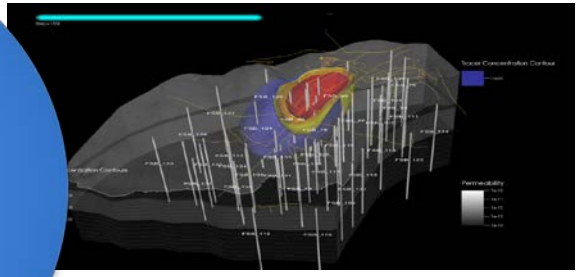


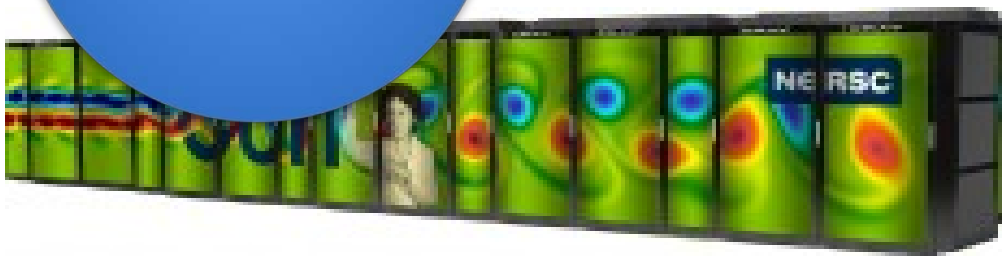


Observations

Analysis

Simulations





Project Summary

- Challenge: Provide a diverse team of scientists together to understand and mitigate a major environmental issue.

- 2D + 3D Visualizations (Provide a complete picture)
  - GIS information, Sensor data, Well Site location, Depositional Environments, Spatial + Temporal information, Comparative visualization

- Domain Centric Collaborative visualization.
  - Allows tools to address needs of complex and diverse team.

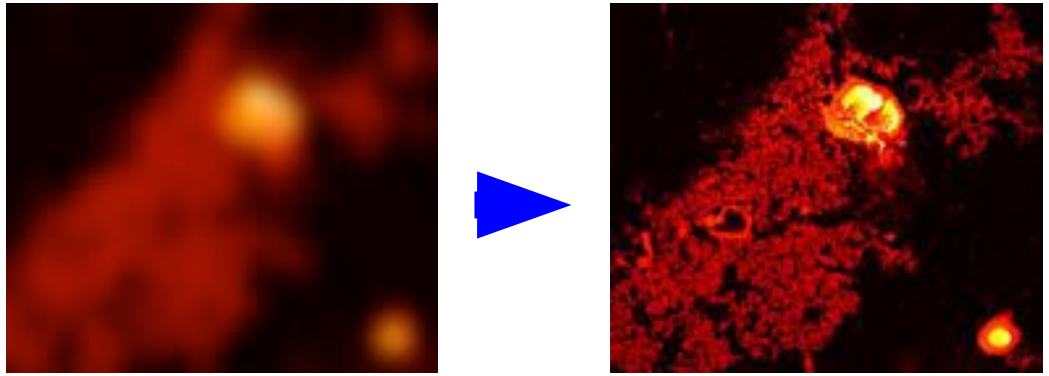
- ASCEM-Akuna Software Toolkit (Open Source)
  - Provides Model-Setup, Inverse Parameter Estimation, UQ, Remote Job Launching & Monitoring of Simulations, and Visualization.

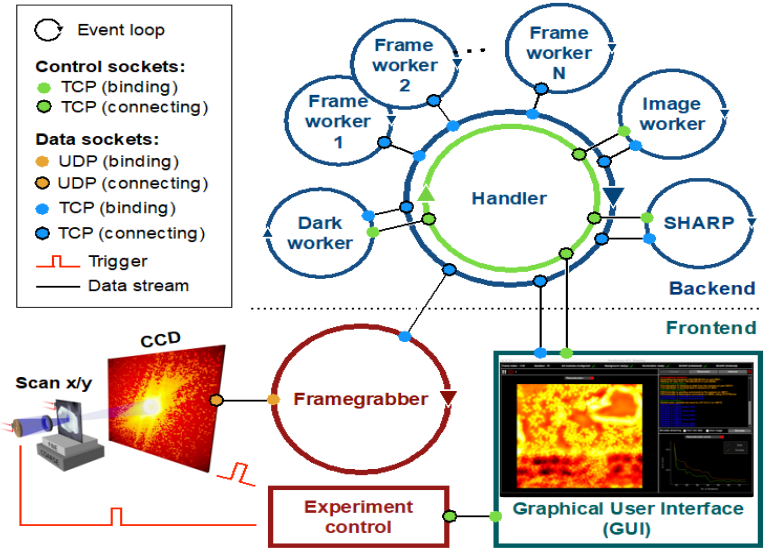
https://akuna.labworks.org/download.html

Science Use Case #2: X-ray Light Sources (Micro/Nano)

- Image reconstruction images from multiple lower resolution diffraction patterns (Ptychography).

- A high throughput realtime data analysis pipeline.





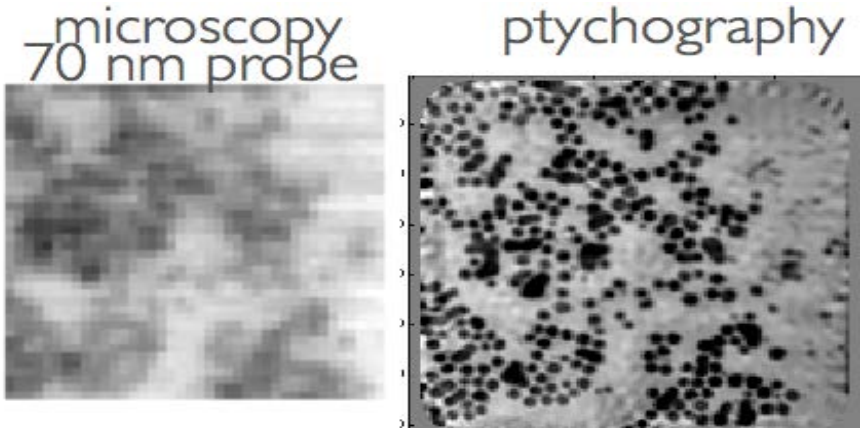
https://arxiv.org/abs/1602.01448 (Multi-node GPU-based Ptychography)
https://arxiv.org/abs/1609.02831 (Streaming Ptychography)

X-ray microscopes, spectrometers, and scattering instruments

- Characterization of structure and properties of materials for example:
  - New drug synthesis
  - Dust particles from space
  - New super conductors
  - Battery research on nanoscale internal structures to understand reactivity
  - Carbon sequestration by porous rock at nanometer scale

- New generation of 3D microscopes
  - brighter x-ray light sources
  - fast parallel detectors

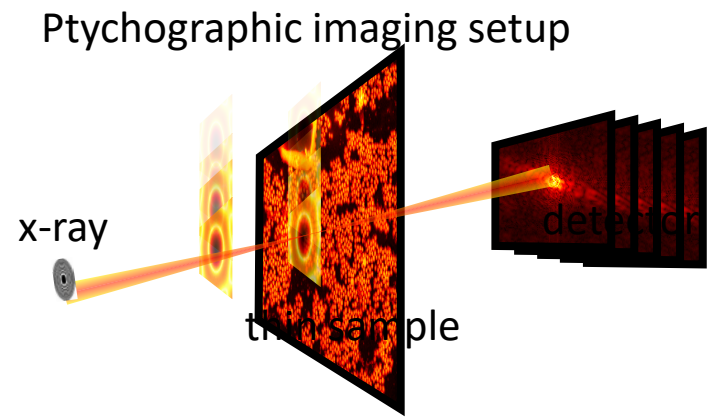
Improvements in image resolution enables this work



Ptychography

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.

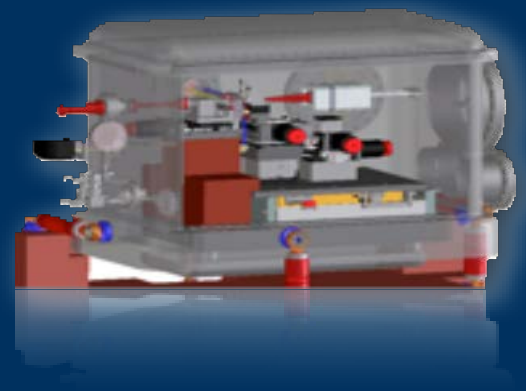
Pytchography:

- 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.

Several ptychographic equipment/codes throughout DOE, universities, world-wide

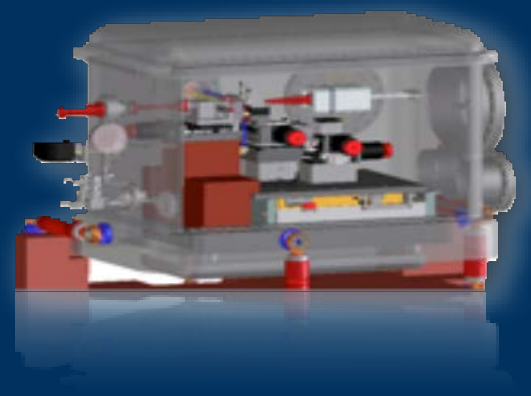
ALS beamline

Nanosurveyor chamber



ALS beamline

Nanosurveyor chamber

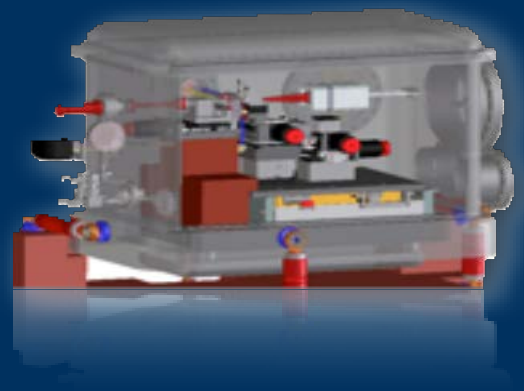


FastCCD detector

200x1024x1024 pixels/s

ALS beamline

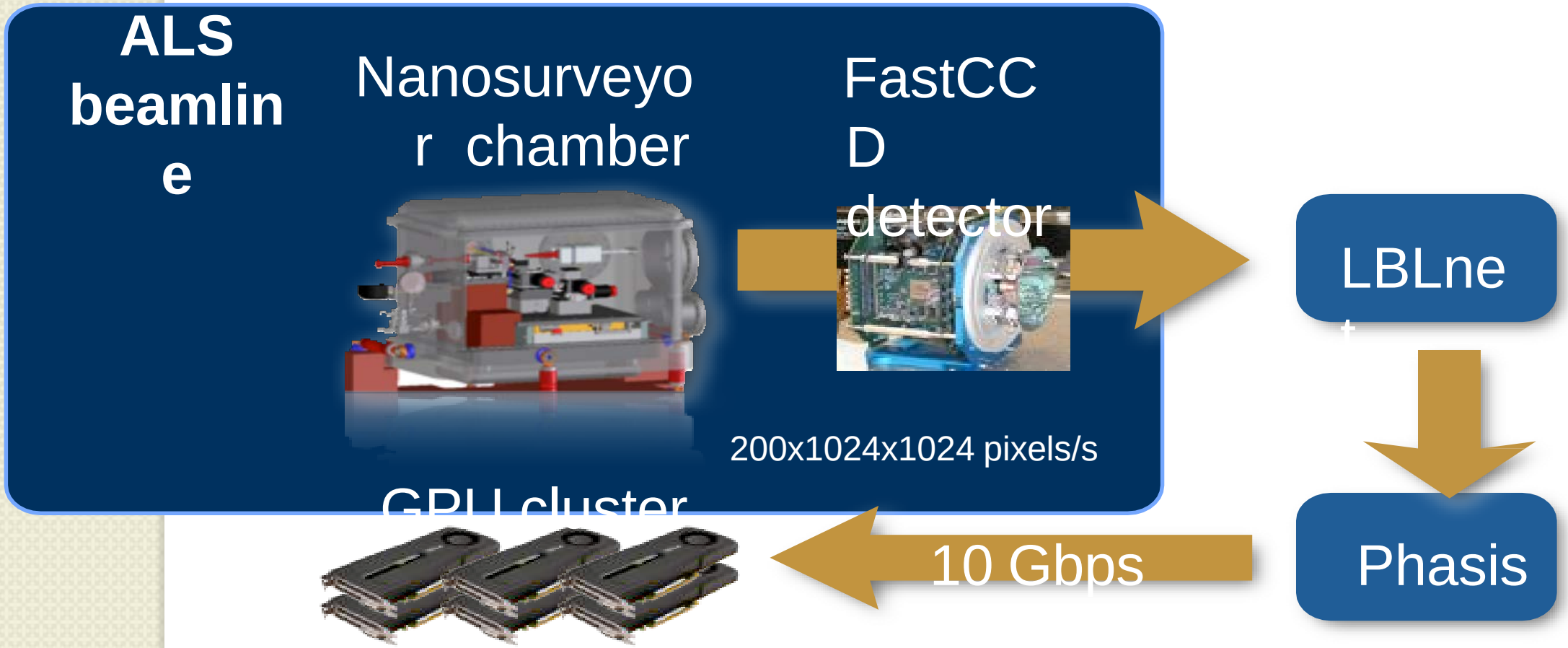
Nanosurveyor chamber

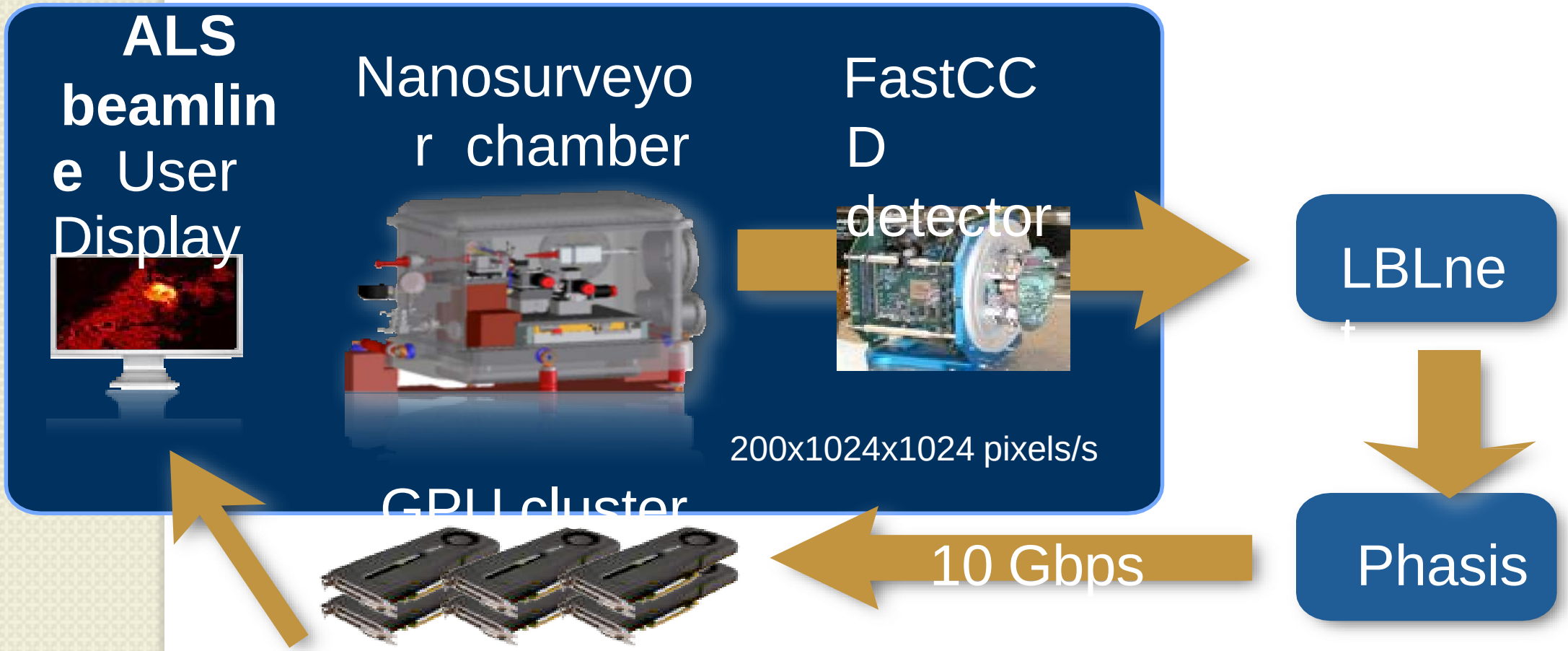


FastCCD detector

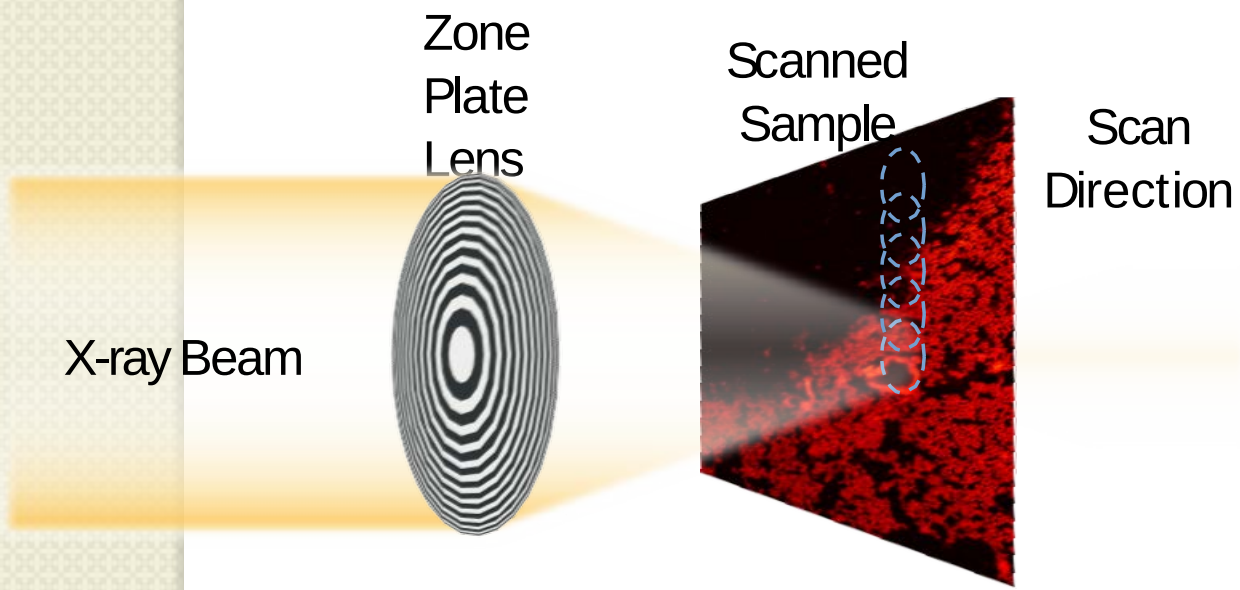
200x1024x1024 pixels/s

LBLnet


ALS beamline
Nanosurveyor chamber
FastCCD detector
LBLnet
200x1024x1024 pixels/s
GPU cluster
10 Gbps
Phasis

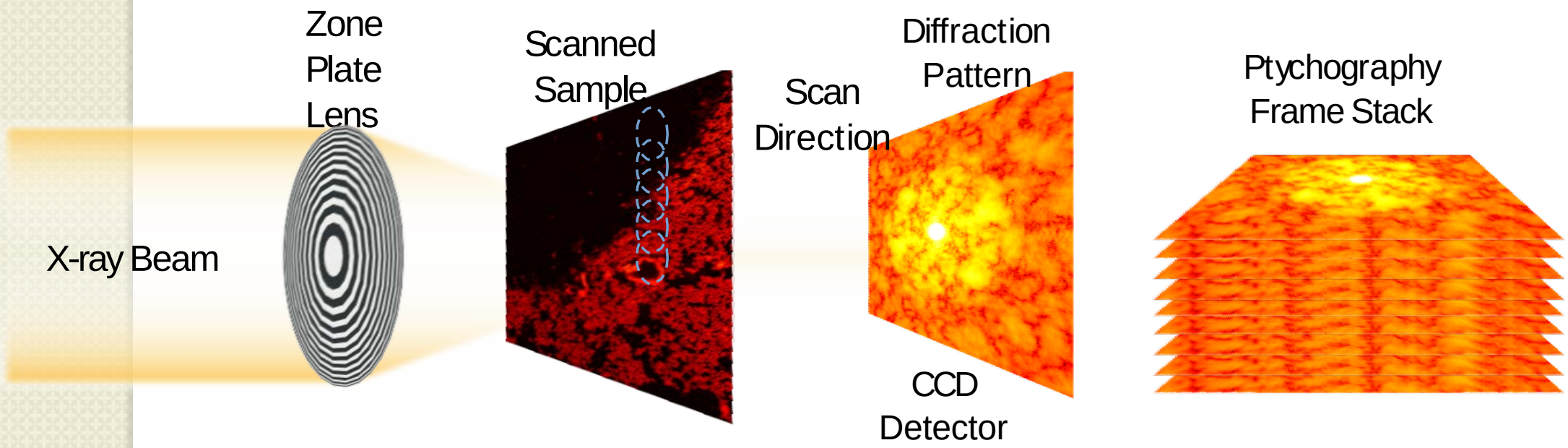

ALS beamline User Display
Nanosurveyor chamber
FastCCD detector
LBLnet
200x1024x1024 pixels/s
GPU cluster
10 Gbps
Phasis

Ptychography is similar to Scanning Microscope but trades greater complexity for higher resolution.



Scanning Microscopes are the most oversubscribed instruments at ALS and other Synchrotrons

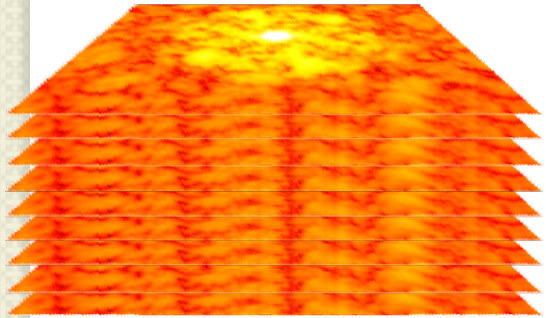
Ptychography is similar to Scanning Microscope but trades greater complexity for higher resolution.



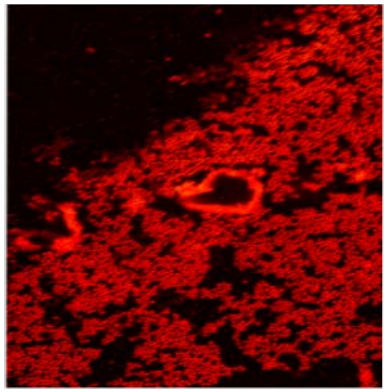
Scanning Microscopes are the most oversubscribed instruments at ALS and other Synchrotrons

Ptychography is similar to Scanning x-ray microscope but trades greater complexity for higher resolution.

2D Diffraction measurements



Phasing



$$I = |F(P_i \cdot O)|^2$$

I = Recorded intensities

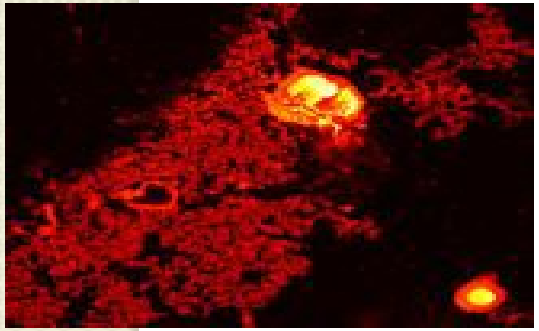
P_i = Illumination probe of frame i

F = Fourier transform

O = Sample Object

Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

Split
kernel

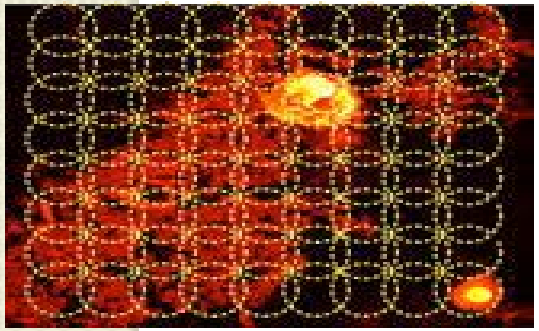


Start with a
random image

Merge kernel

Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

Split
kernel

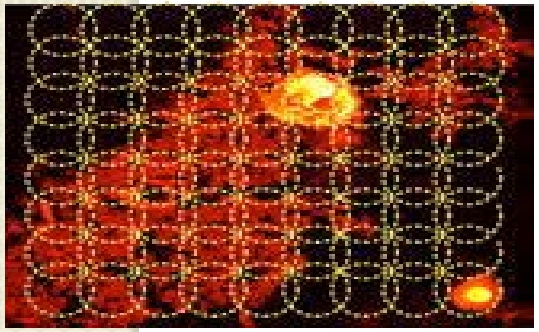


Start with a
random image

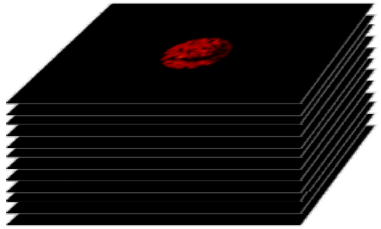
Merge kernel

Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

Split kernel



Multiply Object with Probes



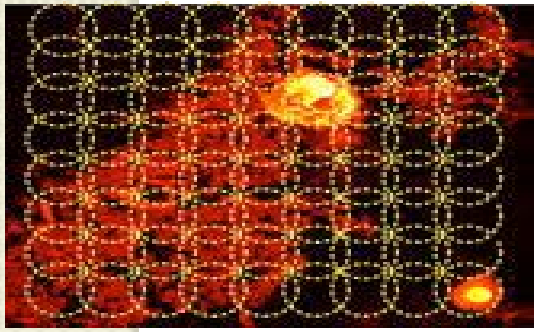
Split kernel

Start with a random image

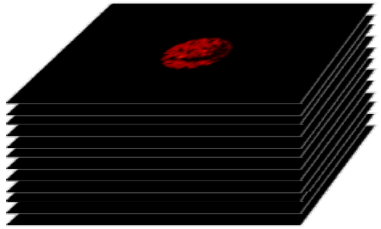
Merge kernel

Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

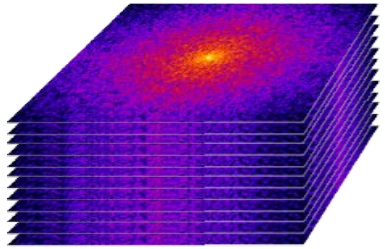
Split kernel



Multiply Object with Probes



FFT frames



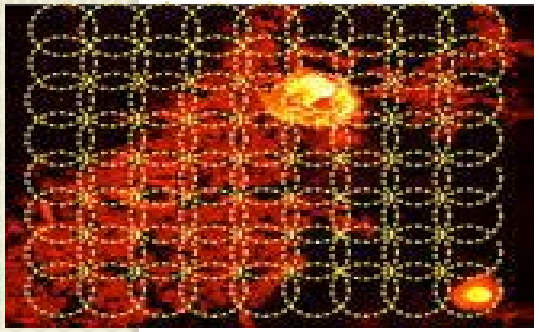
Split kernel

Start with a random image

Merge kernel

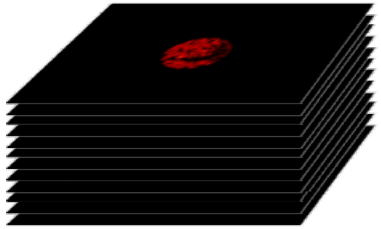
Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

Split kernel



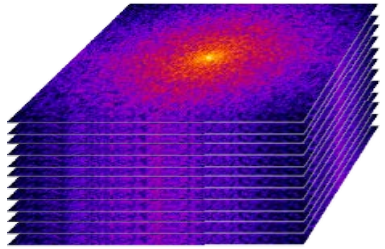
Multiply Object with Probes

Split kernel

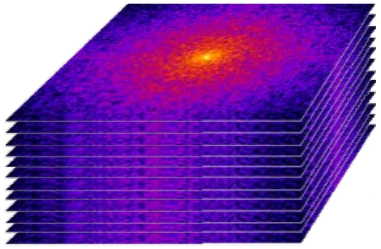


FFT frames

CUFFT



For each pixel replace magnitude with experimental value

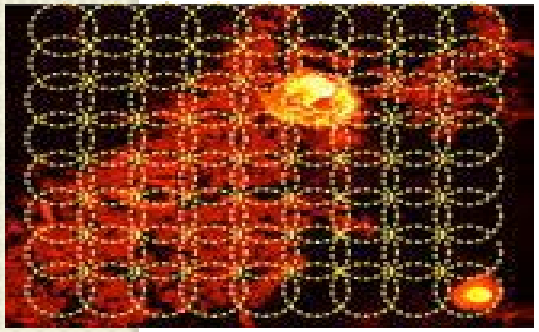


Start with a random image

Merge kernel

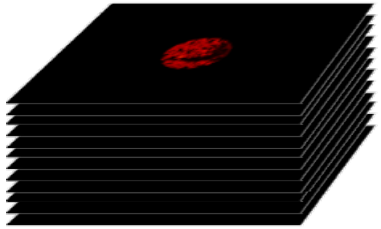
Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

Split kernel



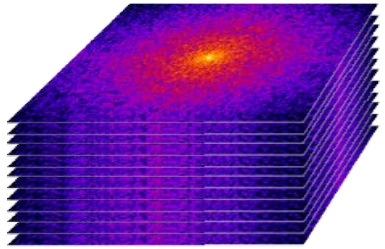
Multiply Object with Probes

Split kernel

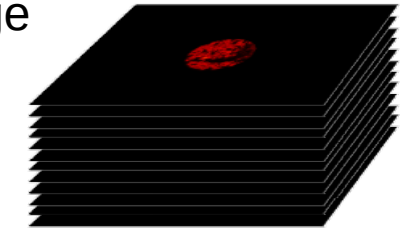


FFT frames

CUFFT

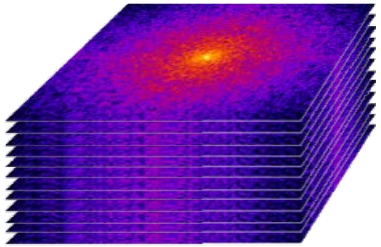


Start with a random image



IFFT frames

CUFFT

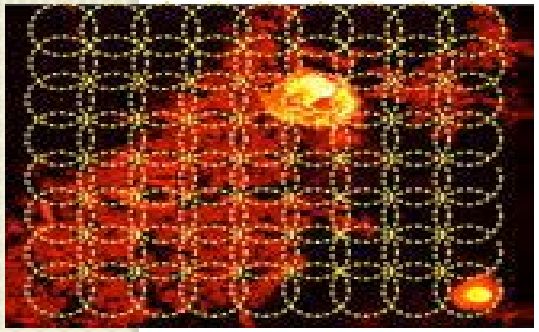


For each pixel replace magnitude with experimental value

Merge kernel

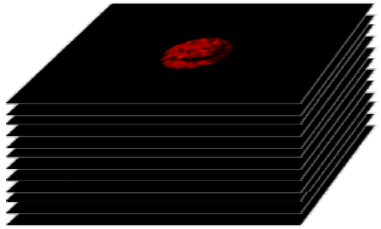
Only a few kernels are necessary to implement basic ptychographic reconstruction on a GPU.

Split kernel



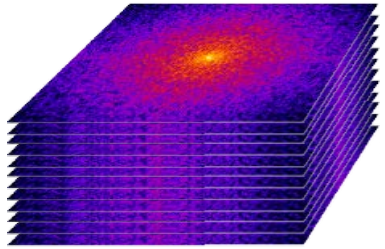
Multiply Object with Probes

Split kernel



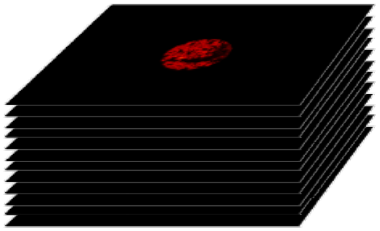
FFT frames

CUFFT



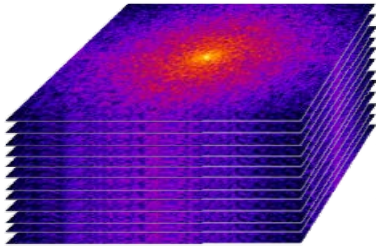
Overlap kernel

Overlap and average frames.



IFFT frames

CUFFT

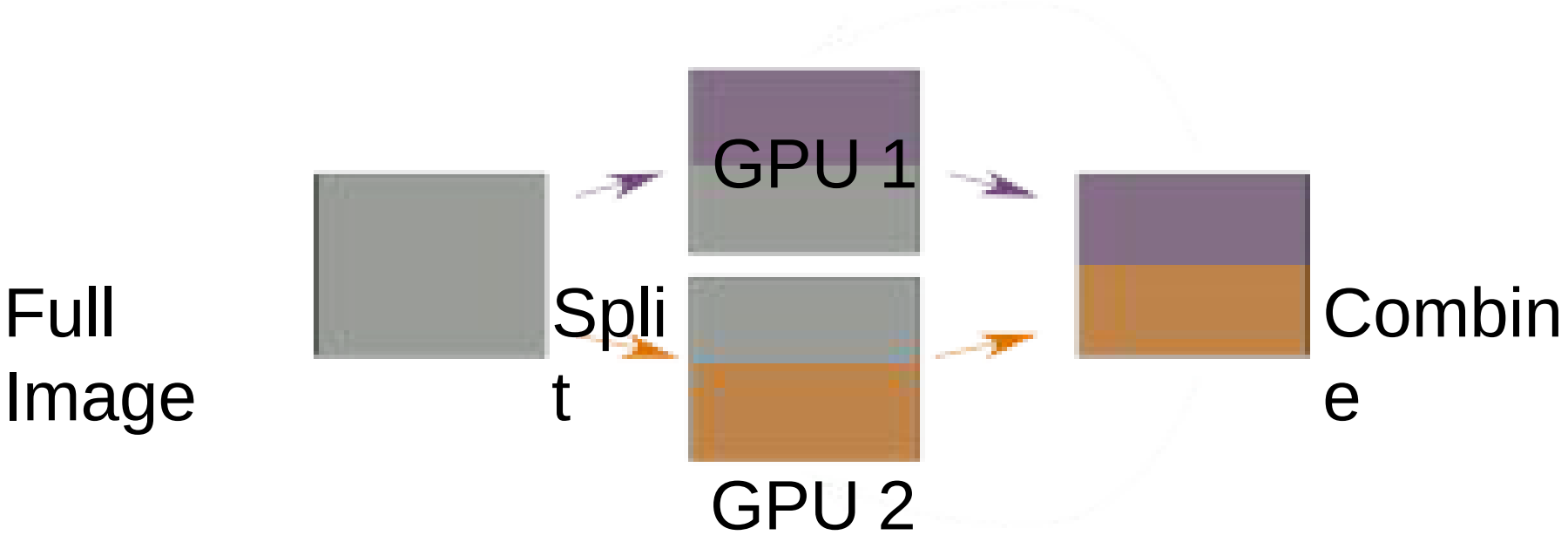


For each pixel replace magnitude with experimental value

Merge kernel

Higher level parallelization

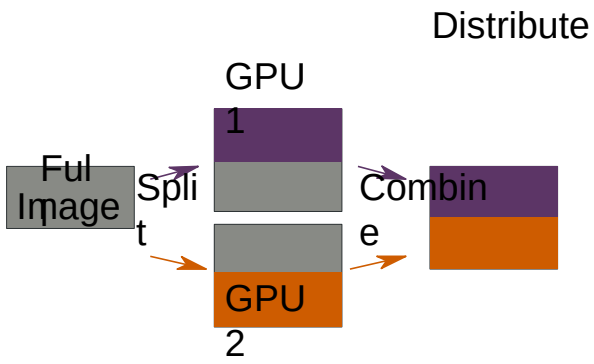
- To be able to process data in real time (200Hz) we need to use multiple GPUs.

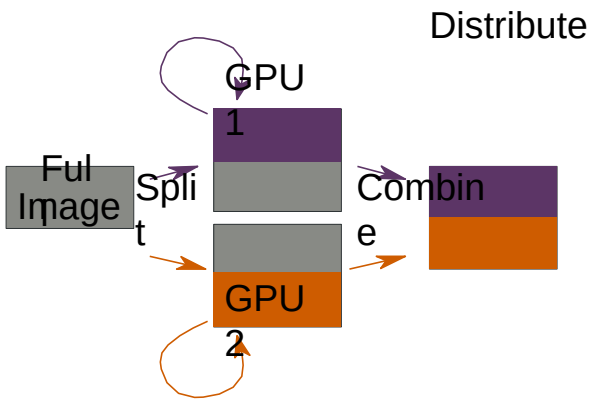


Phase

Higher level parallelization

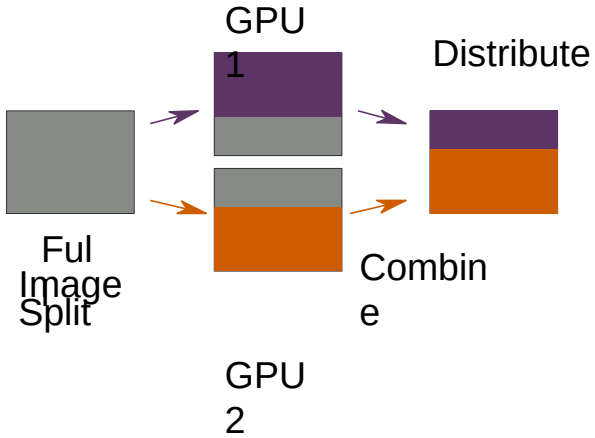
- Split without overlap
- Synchronize every iteration

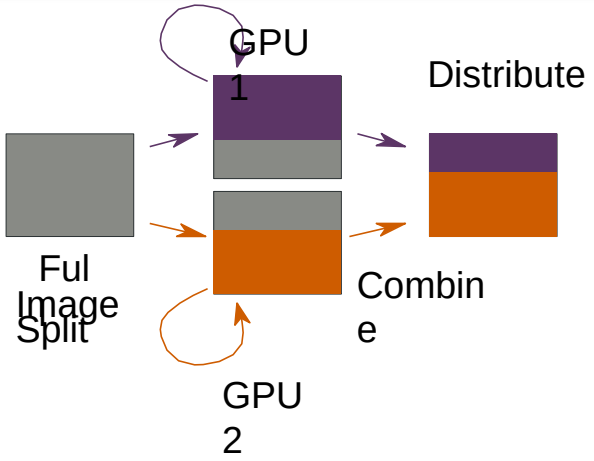




- Split without overlap
- Do not synchronize every iteration

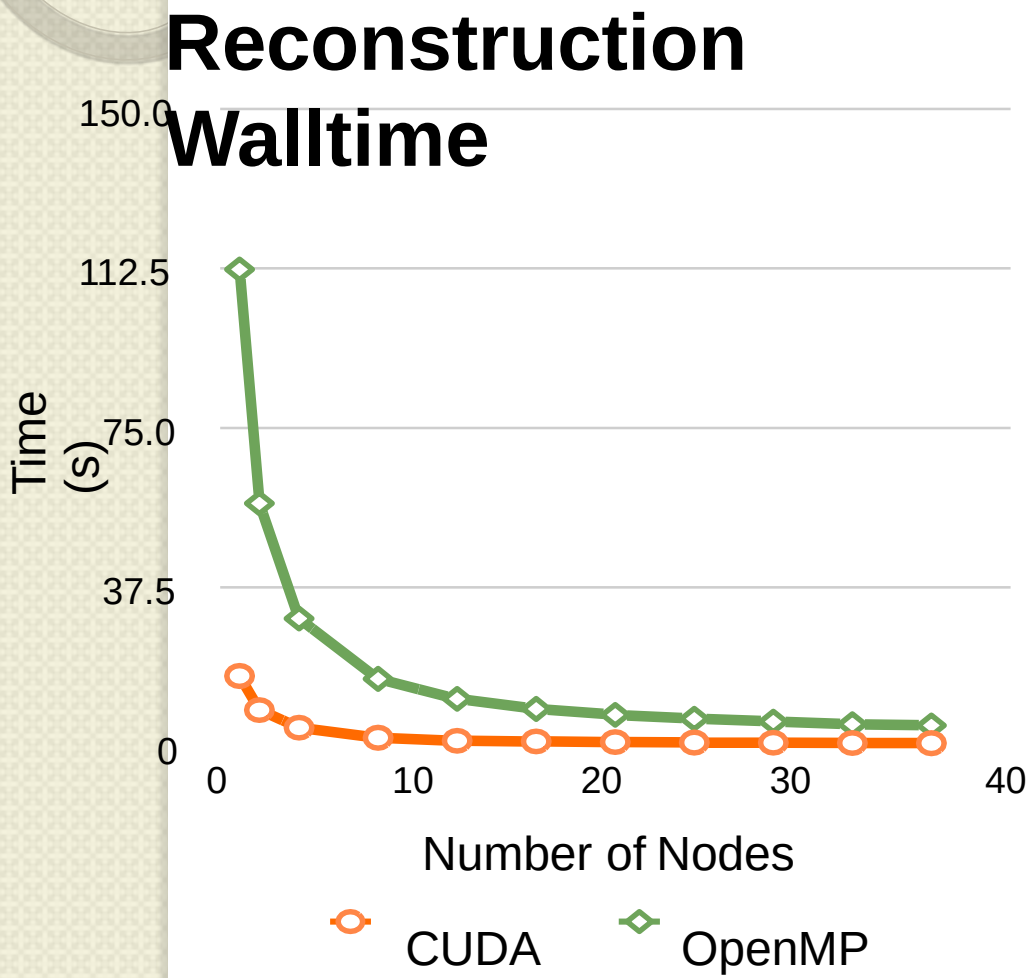
- Split with overlap
- Synchronize every iteration

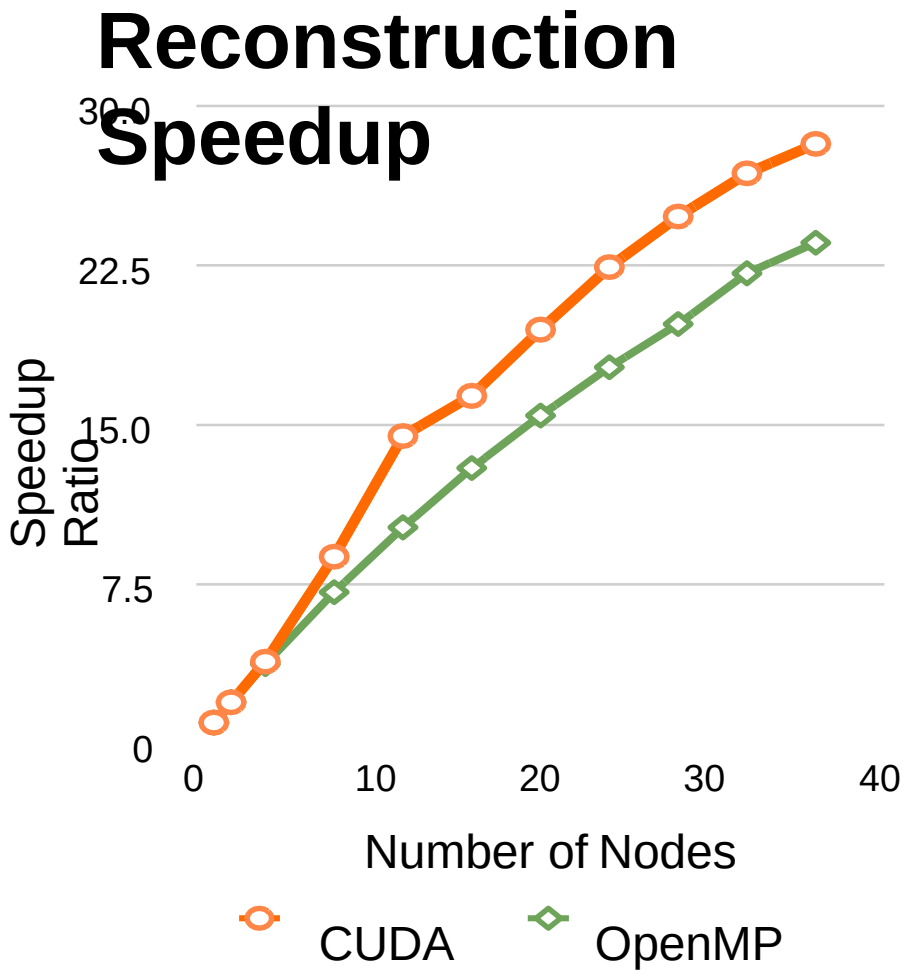




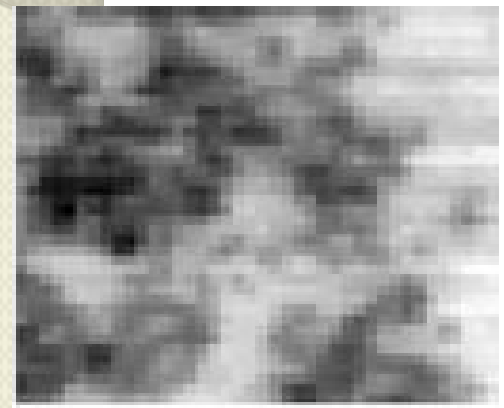
- Split with overlap
- Do not synchronize every iteration

Strong scaling tests on an experimental dataset show the code is scalable.



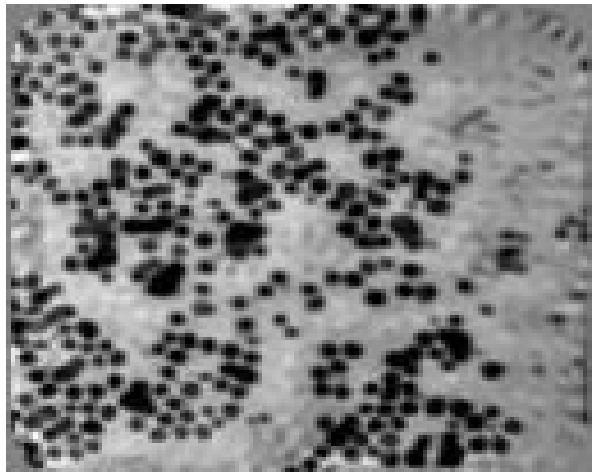


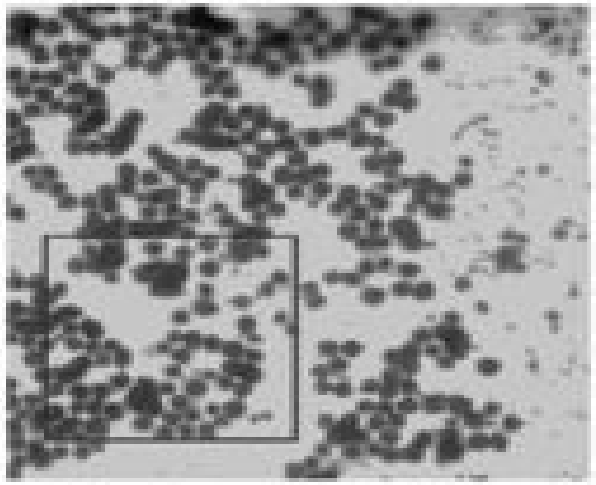
First experimental results show a large improvement in resolution over STXM.



Traditional STXM image.

Ptychography image using the same data.





SEM image

.

Resolution of about 10 nm.

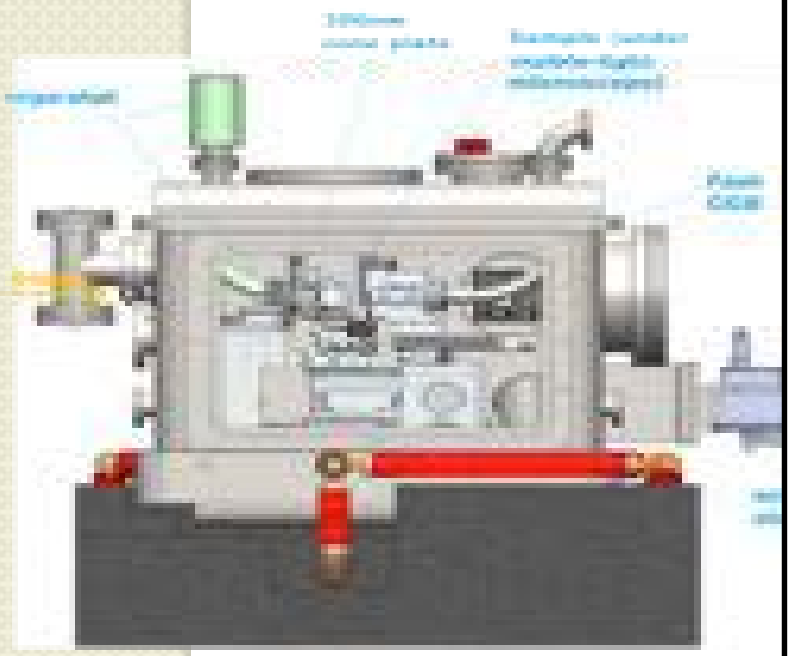
COSMIC-Nanosurveyor

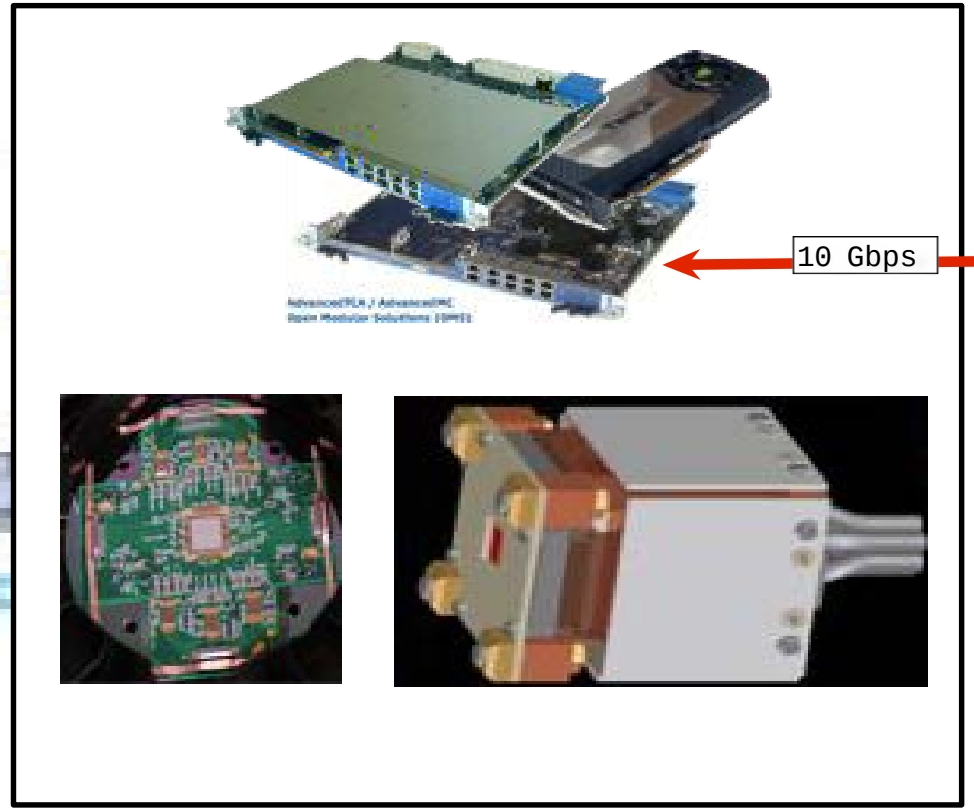
Microscope
-under construction

100 frame / sec CCD
- developed at LBNL

High performance computing
- use of NERSC infrastructure

1 MHz CCD in 3 years





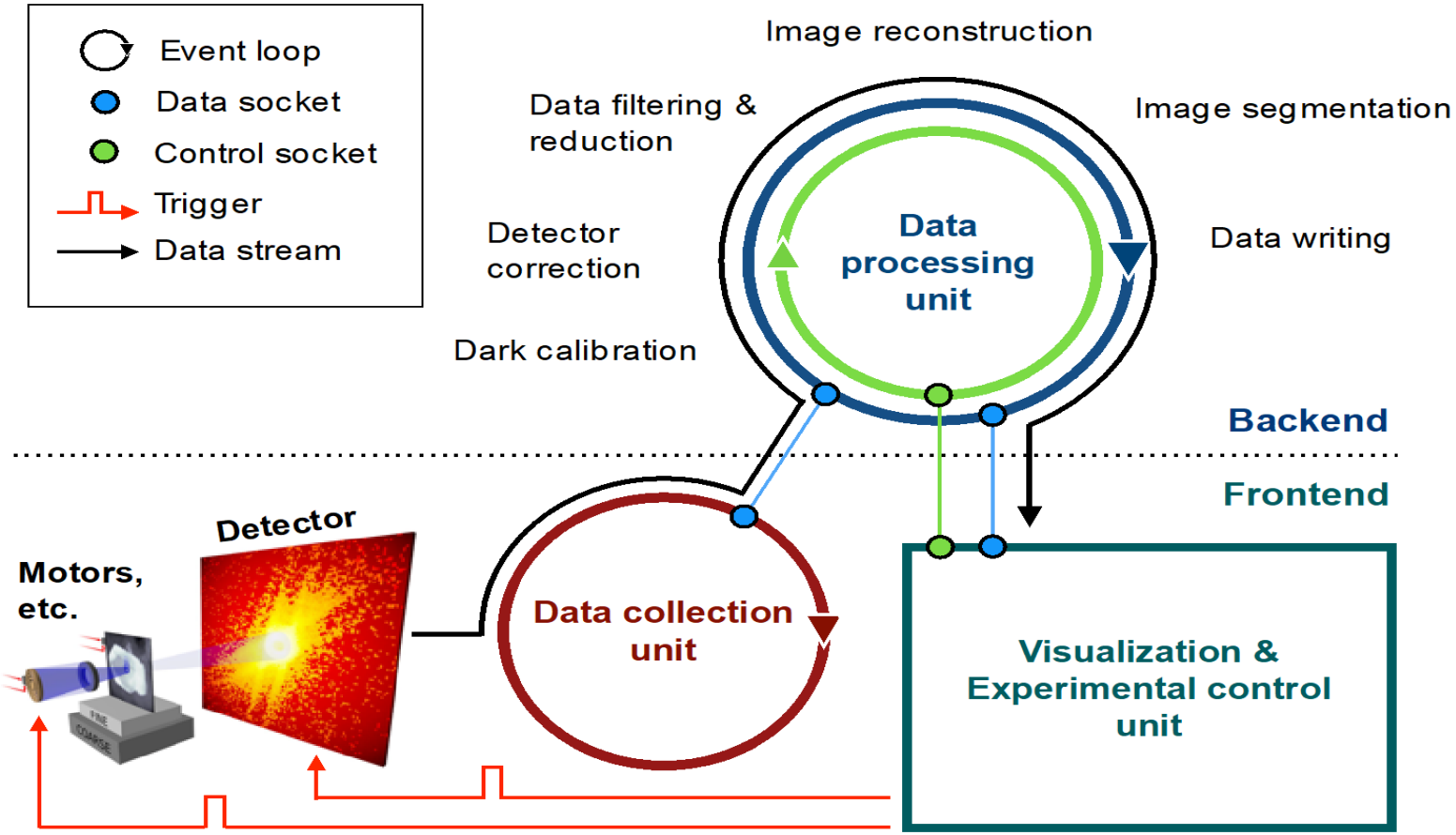
COMPUTATIONAL RESEARCH DIVISION

LBLnet

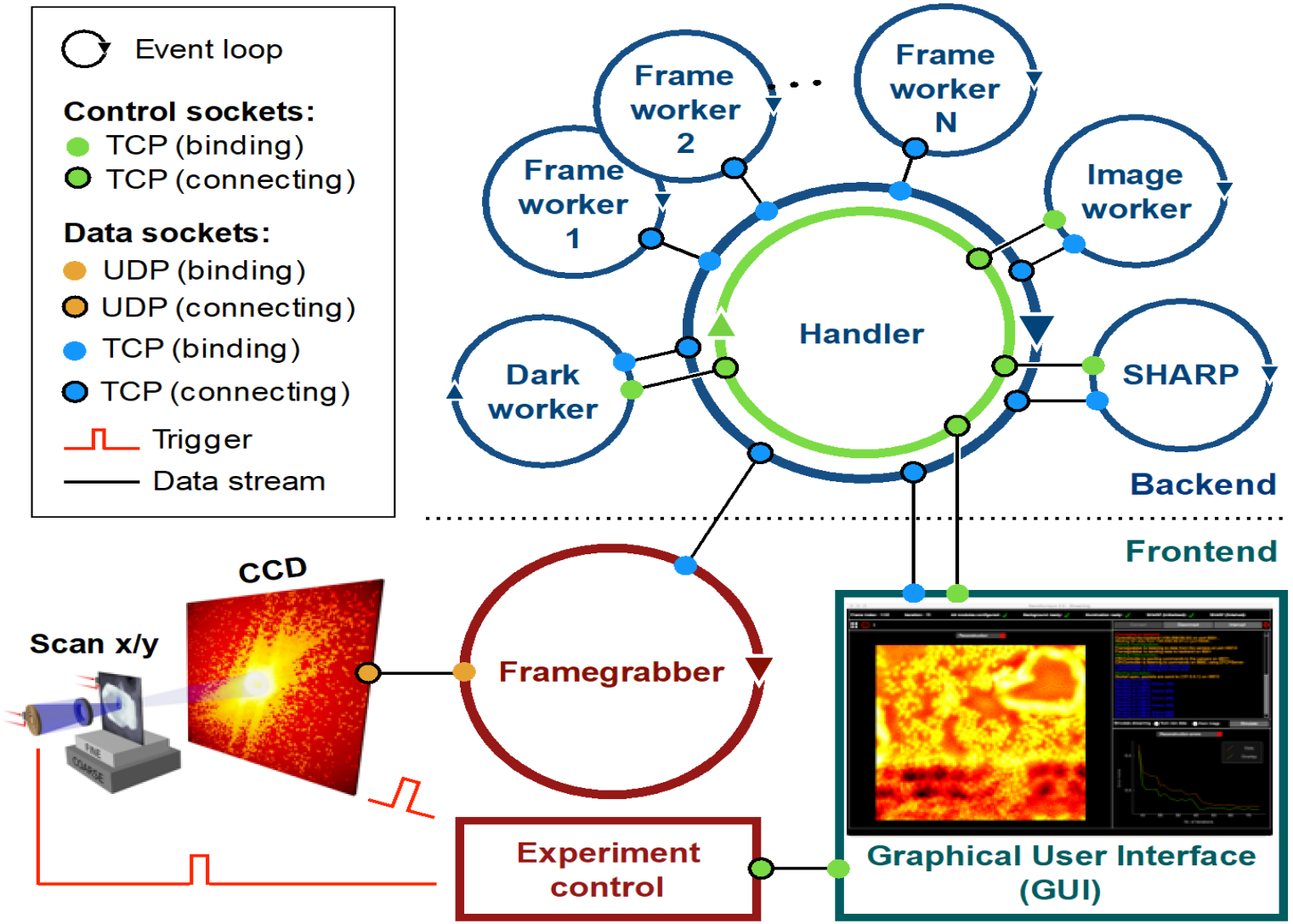
ESnet

GPU Cluster

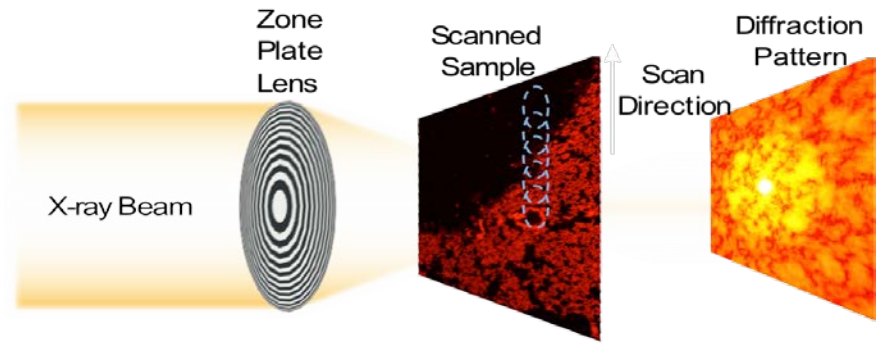
Enabling Streaming Ptychography

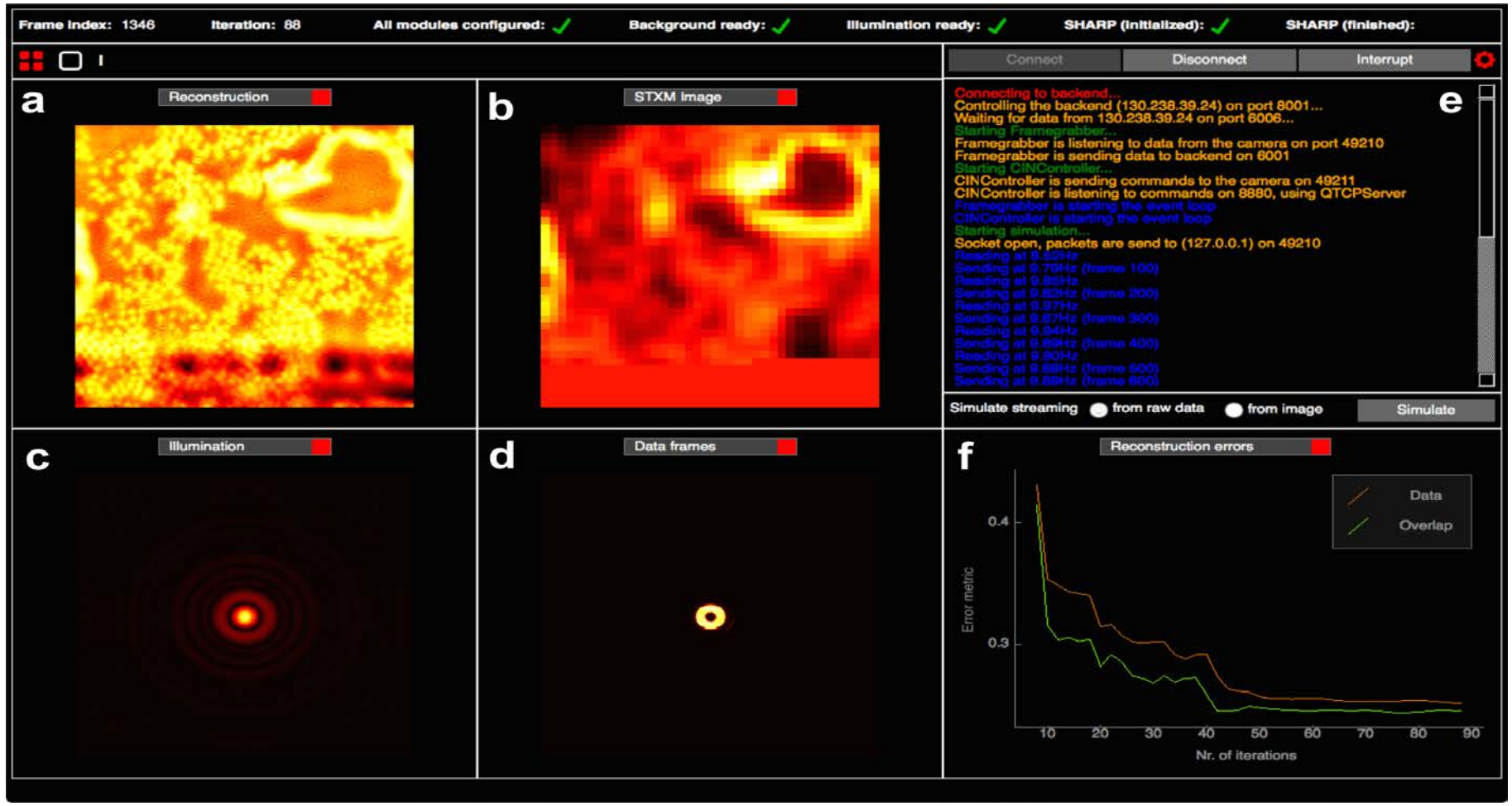


Enabling Streaming Ptychography



Nanosurveyor





Conclusions

- Image reconstruction at nanometer scales enables to new science insight.
- New light sources, parallel detectors, and computational hardware now makes novel algorithms such as real-time Ptychography and tomography possible.
- The rate of data acquisition is also increasing and need for immediate feedback is necessary to ensure optimal use of X-ray beamline.

Final Thoughts

- Thank you!

- Acknowledgements:
  - ASCEM – ASCR/DOE funded Environmental Management project
  - CAMERA/ALS – ASCR/BES funded project Team members – X-ray light sources

Publications

- http://adsabs.harvard.edu/abs/2012APS..DPPYP8009S
- http://scripts.iucr.org/cgi-bin/paper?S1600576716008074 http://www.sciencedirect.com/science/article/pii/S1877050912002141
- http://www.rsmas.miami.edu/users/tamay/ftp-pub/omod12b.pdf
- https://www.eclipse.org/community/eclipse_newsletter/2015/january/article1.php
- http://link.springer.com/article/10.1007/s11837-016-2098-4
- http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.4952921
- https://publications.lbl.gov/islandora/object/ir%3A1005825
- http://onlinelibrary.wiley.com/doi/10.1002/cpe.3697/abstract
- http://onlinelibrary.wiley.com/doi/10.1002/adma.201502276/abstract
- http://sti.srs.gov/fulltext/SRNL-STI-2015-00027.pdf
- http://www.tandfonline.com/doi/abs/10.1080/08940886.2015.1013413

Software

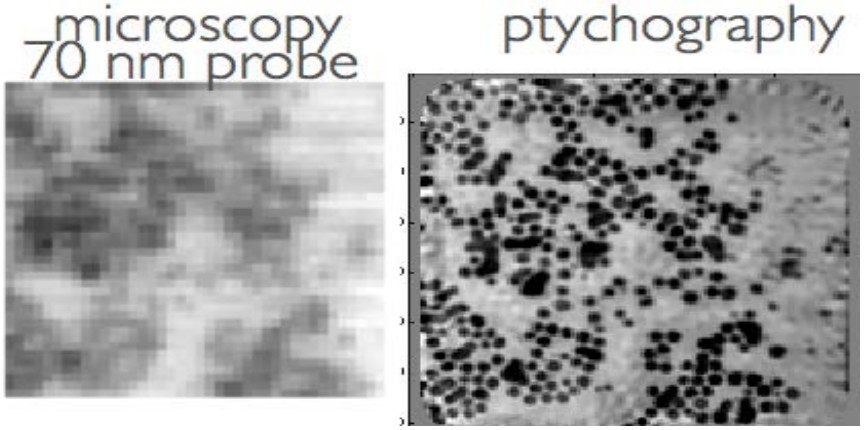
- https://wci.llnl.gov/simulation/computer-codes/visit/
- https://akuna.labworks.org/download.html
- https://github.com/eclipse/ice
- https://github.com/CameraIA/F3D
- https://bitbucket.org/lbl-camera/xi-cam
- https://github.com/UV-CDAT/uvcdat
- https://github.com/LBL-EESA/TECA
- https://github.com/visit-vis/visit_java_client
- http://www.camera.lbl.gov/software

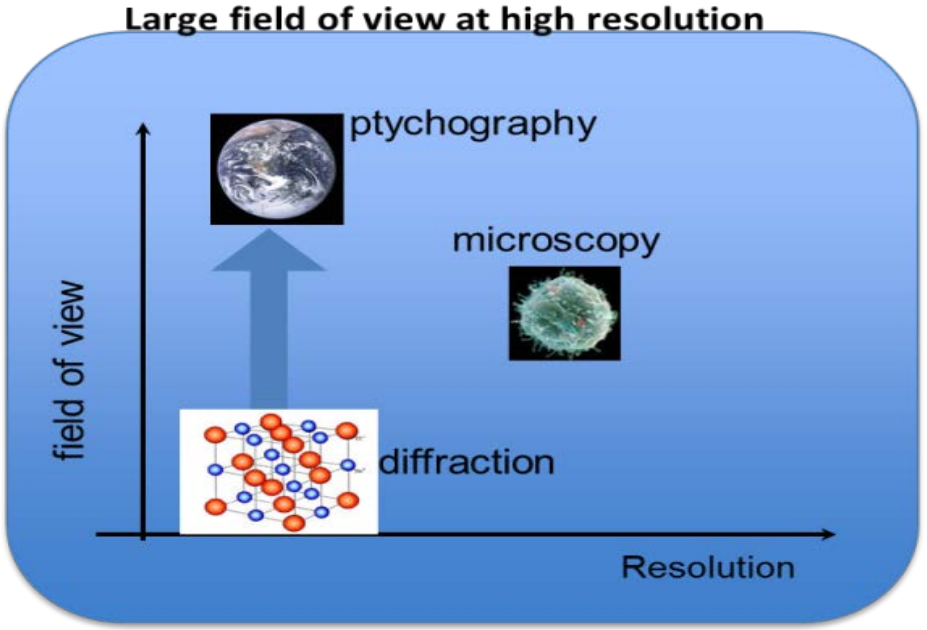
Why is ptychography so interesting?

- Diffraction resolution
- Macroscopic field of view
- Increased contrast through phase
- In-situ optical metrology (blind ptychography)
- Turns more data into better resolution
- extendible to spectro-ptychography, ptycho-tomography, near field, Fourier Ptychography, time resolved dynamics

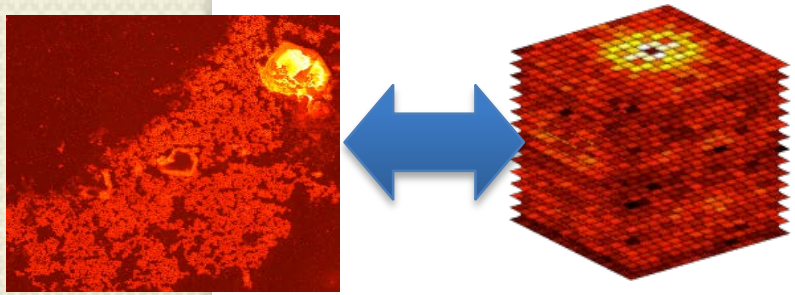
Why not is everyone doing it (cons)

- requires fast detectors
- requires a bright source
- requires mathematics
- requires parallel code
    - Alternating Projections
    - "RAAR",





an overdetermined problem
in high dimensional space.



Low dimensional space	⟷	**Large dimensional data**
fit model	⟷	**fit data**

How to solve it?
Algorithms in this talk

- **Projection algorithms**
  - **Alternating Projections**
  - "**RAAR**",
  - Augmented Lagrangian
  - "Difference Map", "HIO"

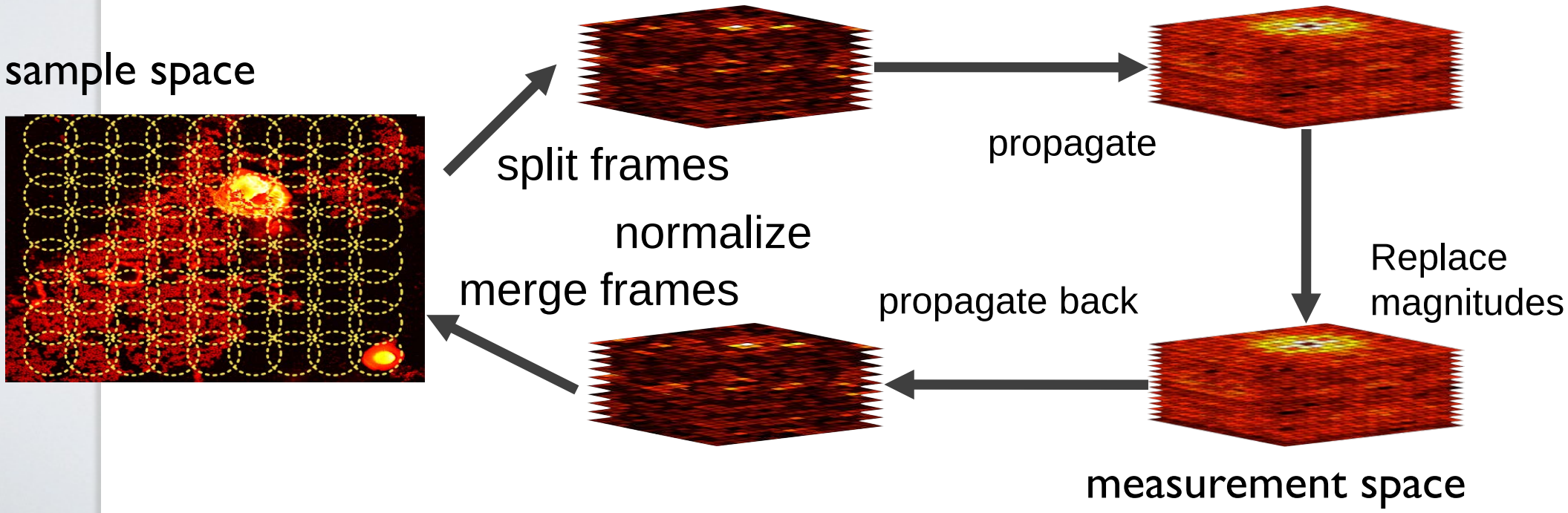
  tutorial
  in use
- **(Weighted) Least Square methods,** maximum likelihood:
  - **Conjugate Gradient**,
  - Newton,
  - CG Newton

  { **acceleration**
  noise model
- **Spectral methods**
  - **synchronization**
  - Graph Laplacian

  large scale
  robust

fit model ⟷ **fit data**

Alternating projections



How to simulate it?

Ptychographic imaging setup

Fourier transform data

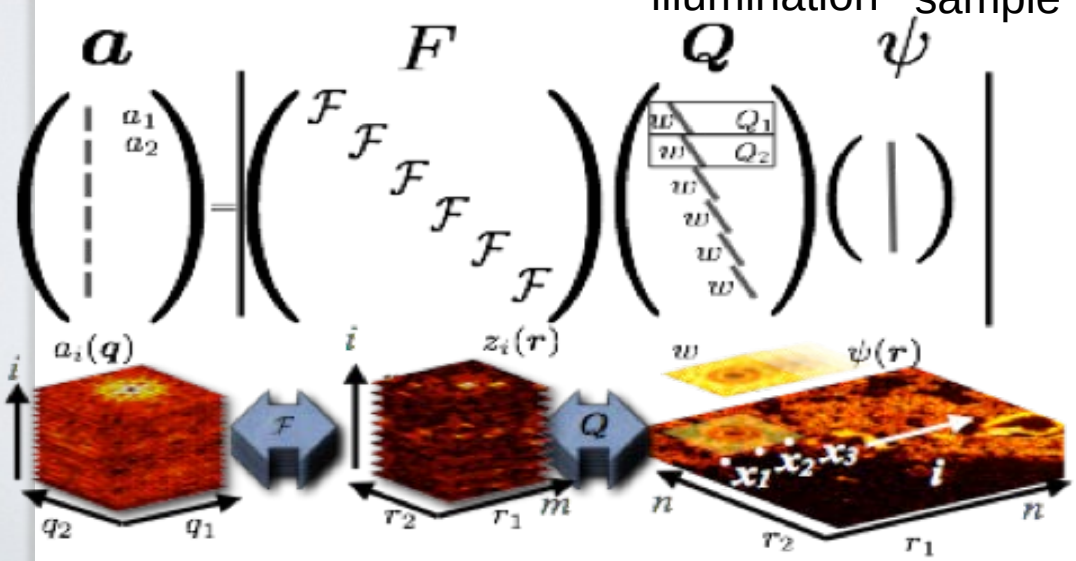
translate and illuminate

amplitude=

propagation

scanning illumination

sample



measured

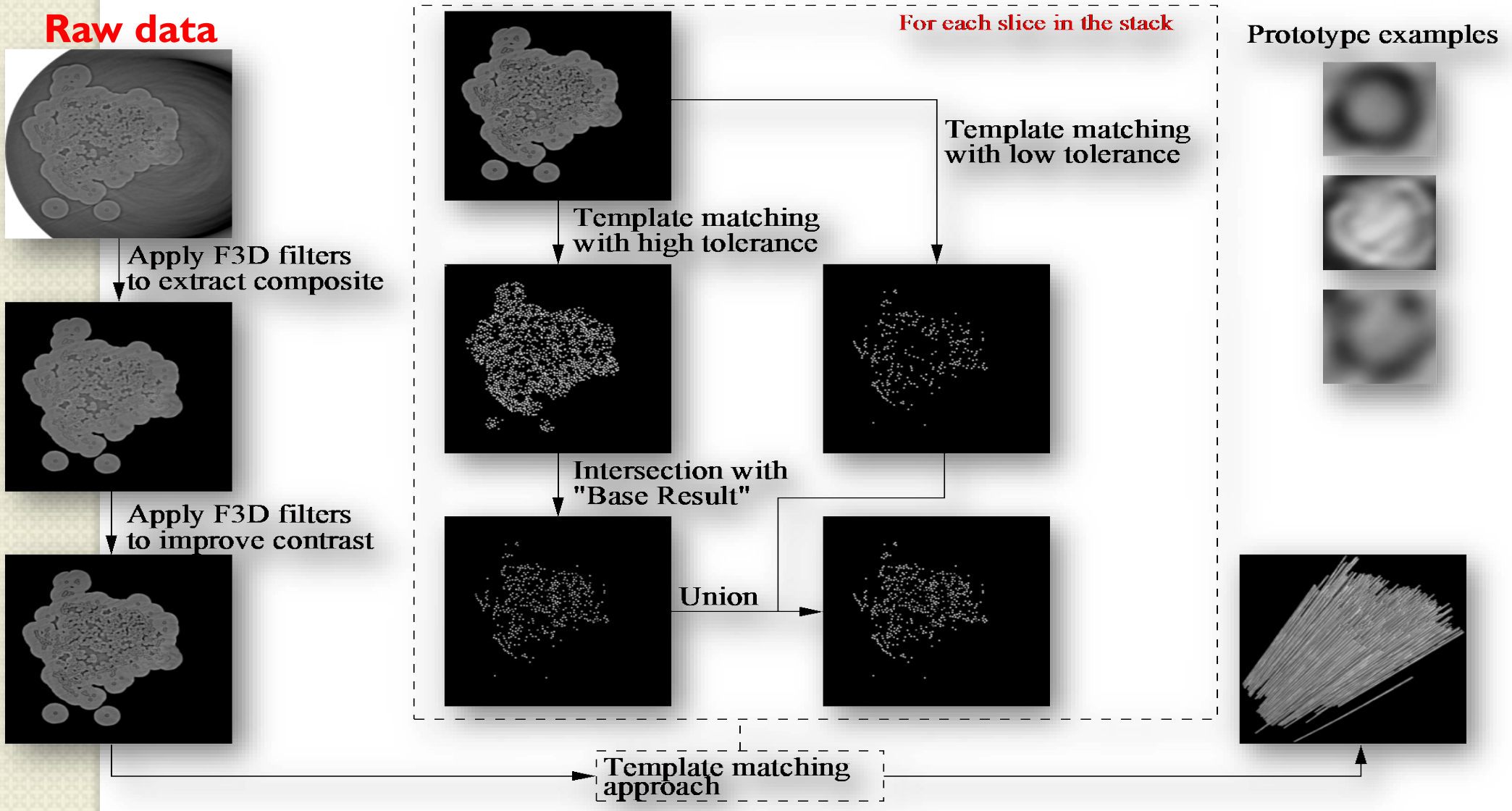
$$\begin{cases} \mathbf{a} = |\mathbf{F}\boldsymbol{z}|, \\ \boldsymbol{z} = \mathbf{Q}\boldsymbol{\psi}^{\vee}, \end{cases}$$

unknown

"frames"

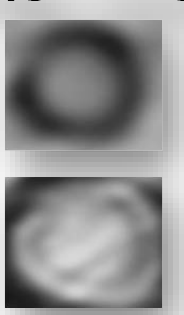
find unknown

Identification of structures



Template matching

Prototype examples



1)

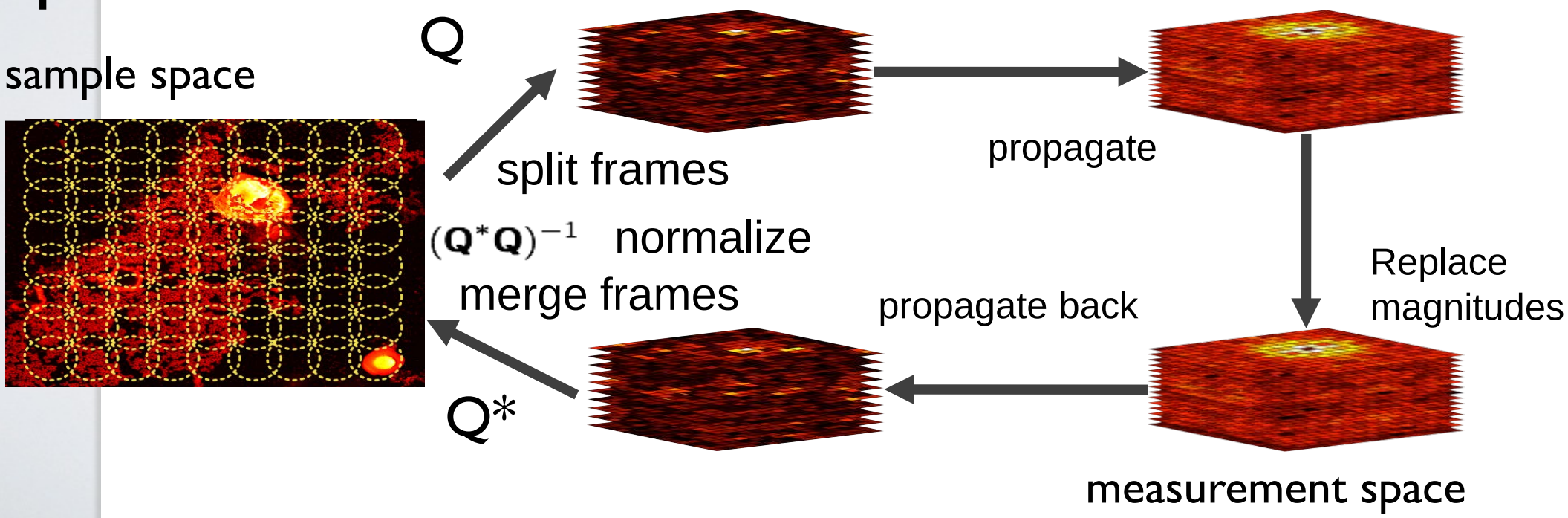
$$MSE(x,y)=\frac{1}{n}\sum_{i,j}[p(i,j)-f(x+i,y+j)]^2$$

2) Determine the best matches:

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

Alternating projections

- **project onto sample space** $P_{\mathbf{Q}} = \mathbf{Q}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*$
- **project onto measurement space** $P_{\mathbf{a}}\mathbf{z} = \mathbf{F}^* \frac{\mathbf{F}\mathbf{z}}{|\mathbf{F}\mathbf{z}|}\mathbf{a}$
- **repeat**



Least square methods

These methods are equivalent

$$\min_{\psi} \|\boldsymbol{a} - |\boldsymbol{F}\boldsymbol{Q}\psi|\|$$ minimize discrepancy with data

$$\min_{\boldsymbol{z}} \|\boldsymbol{a} - |\boldsymbol{F}\boldsymbol{z}|\| \quad \text{s.t.} \;\; \boldsymbol{z} = P_{\boldsymbol{F}\boldsymbol{Q}}\boldsymbol{z}$$

These iterative methods are equivalent

"Alternating projections"

"Error Reduction"

"projected steepest descent"

$$z^{\ell+1} = P_Q P_a z^{\ell} = z - P_Q \nabla \epsilon^2(z^{\ell})$$

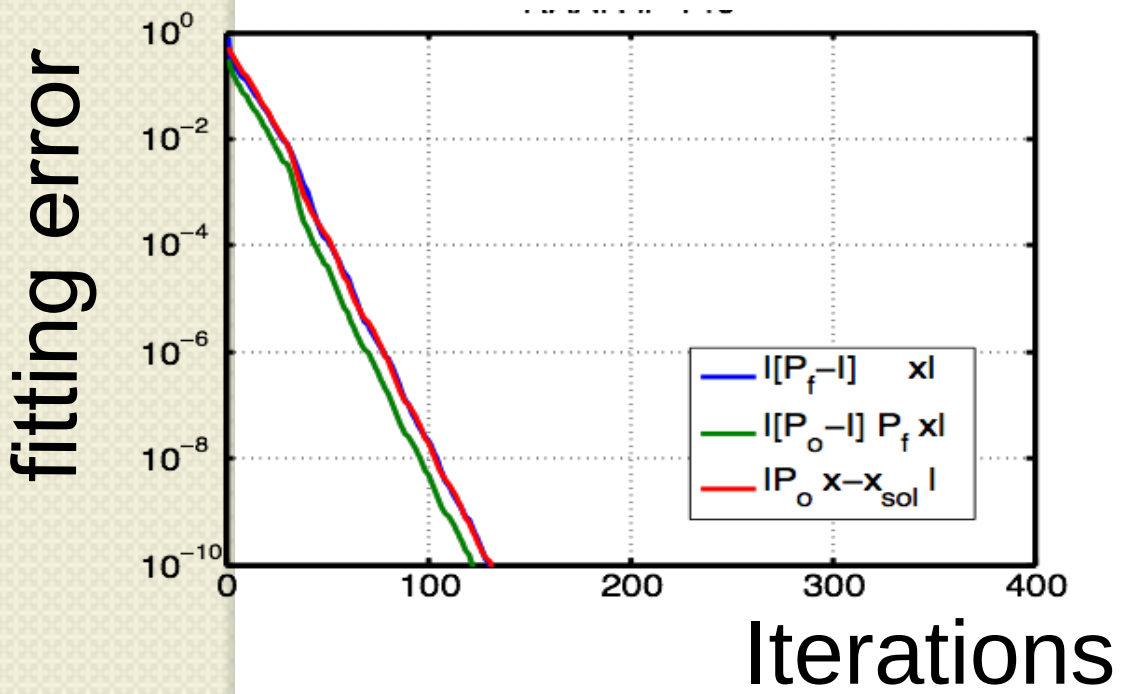
How to speed up?

- **Relaxed "Douglas-Rachford" (RAAR) (SHARP release) O(5x speedup)**
- **Conjugate directions acceleration O(10x speedup)**
  - gradient from fast projections kernels
  - line search using Newton step from implicit Hessian
- **Synchronization-Conjugate directions-line search O(20x speedup)**

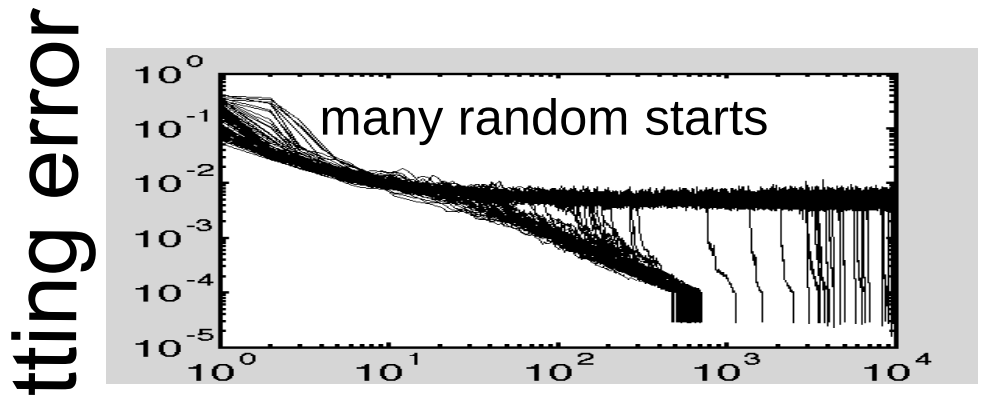
Iterative Algorithms for Ptychographic Phase Retrieval, C. Yang, J. Qian, A. Schirotzek, F. Maia, S. Marchesini, [arXiv:1105.5628] LBNL-4598E
Efficient Algorithms for Ptychographic Phase Retrieval. J. Qian, C. Yang, A. Schirotzek, F. Maia, and S. Marchesini, Contemporary Mathematics 2014.

Nearest Neighbor Overlap Enables Robust Convergence

ptychography



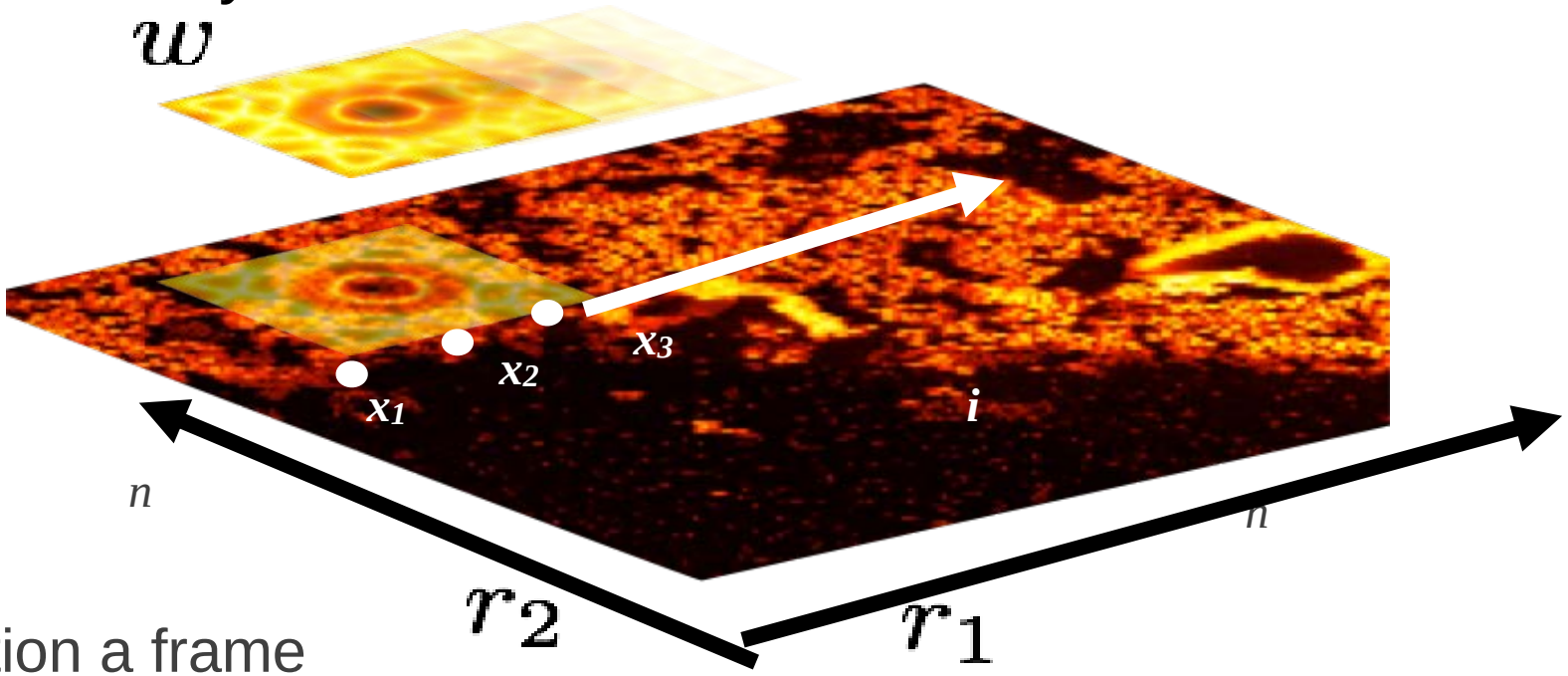
Coherent Diffractive Imaging



Numerical experiments show linear convergence rate, however…

The problem with size

Long range interactions among frames decay exponentially with distance



- at each iteration a frame only talks to neighbors
- how to achieve long range scaling?

Phase synchronization

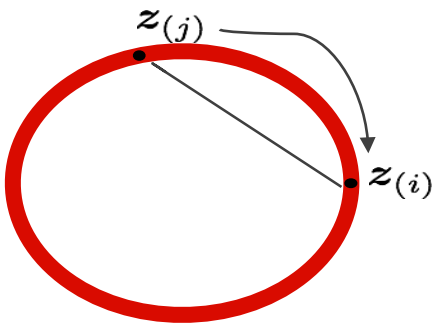
2 frames

$z_{(1)} = \xi z_{(2)}$

$z_{(2)}$

phase factor

$|\xi| = 1$



How to find common phase?

best fit

$$\min_{|\xi|=1} \|z_{(1)} - \xi z_{(2)}\|^2$$

$$\|z_{(1)}\|^2 + \|z_{(2)}\|^2 - 2\Re\left(z_{(2)}^* z_{(1)}\right)\xi^*$$

maximize product

max

Align phases:

$$\xi = \frac{\left(z_{(1)}^* z_{(2)}\right)}{\left|\left(z_{(1)}^* z_{(2)}\right)\right|}$$

Normalize dot product

what if many frames are out of phase?

$z_{(i)}$ $z_{(j)}$

phase factor

$|\xi| = 1$

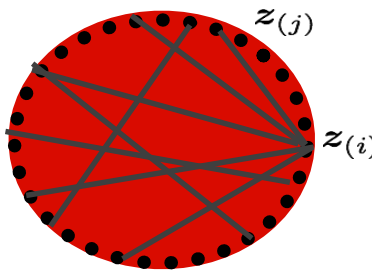
minimize all the differences

$$\min_{|\xi|=1} \sum_{i,j} \|z_{(i)}\| + \|z_{(j)}\| - \boxed{2\xi_{(i)} \left(\mathcal{H}\right)_{(i,j)} \xi_{(j)}}$$

Simplify

Dot product between frames

$$\left(\mathcal{H}\right)_{(i,j)} = \left(z^*_{(i)} z_{(j)}\right)$$



$\max_{\xi} \xi^* \left(\mathcal{H}\right) \xi$ Phase synchronization

Any meaning ?

Yes ! Spectral method

Synchronize phases by spectral methods

to align the phases, find: $\max_{\xi} \boldsymbol{\xi}^* (\mathcal{H}) \boldsymbol{\xi}$

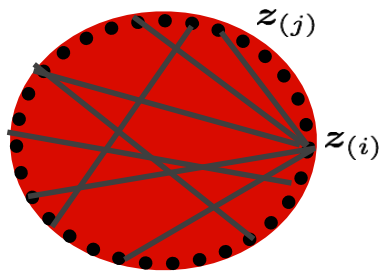
Which is equivalent to finding* largest eigenvector

*quick, scalable (e.g. by ARPACK)

$$\mathcal{H} = R^* \begin{pmatrix} \lambda_0 & & \\ & \lambda_1 & \\ & & \ddots \end{pmatrix} R$$

what does it mean?

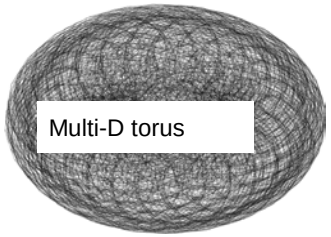
H is the "graph laplacian" of a network



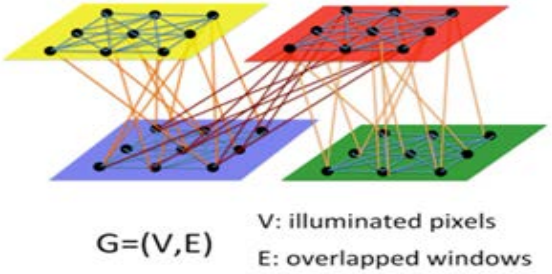
accelerate and build a better starting guess:

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

Diffraction data manifold

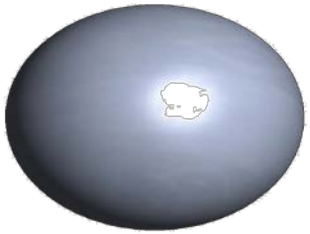


(2) Build “relationship network RN: a graph (V,E) that relates each frame to its neighbors.



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

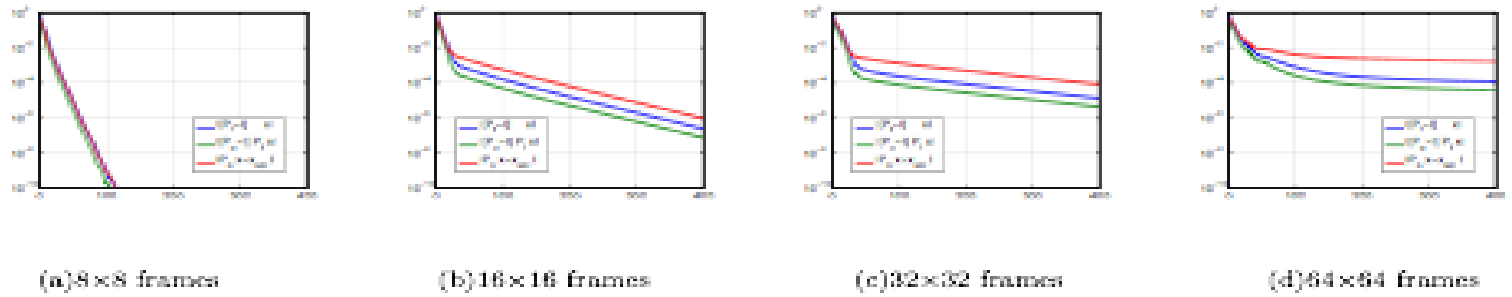
(4) The largest eigenvector of the Connection graph provides the most aligned phases encoding the (approximate) data topology.



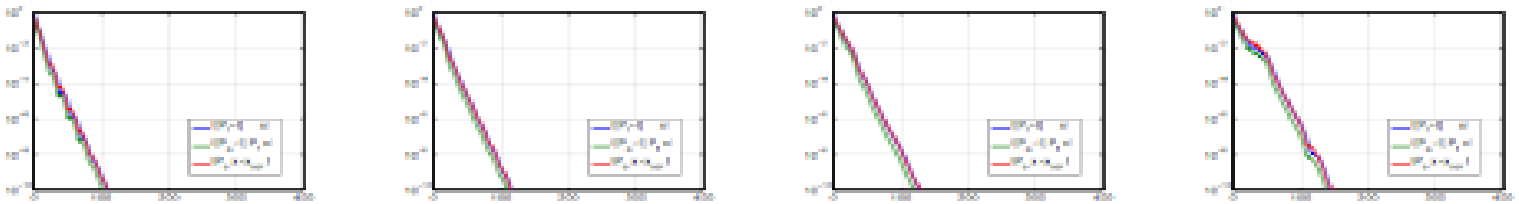
Approximate torus with ball

This provides a strong starting guess.

- RAAR



- synchro-RAAR



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. We also show that

(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.