Berkeley Lab Computational Research Update

CCS - LBL Collaborative Workshop
April 10, 2014

David L. Brown
Computational Research Division Director

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB
Lawrence Berkeley National Laboratory

Berkeley Lab at a Glance

- **Solve the most pressing and profound scientific problems facing humankind**
  - *Basic science for a secure energy future*
  - *Understand living systems to improve the environment and energy supply*
  - *Understand matter and energy in the universe*
- **Build and safely operate world-class scientific facilities**
- **Train the next generation of scientists and engineers**

- *3,863 FTE*
- *3,040 Employees*
- *267 Joint faculty*
- *491 Postdoctoral researchers*
- *328 Graduate students*
- *194 Undergraduates*
- *1,612 Visiting scientists and engineers*

FY 2011 Total
$734M core, $89M ARRA

Berkeley Lab's largest research facilities see more than 25,000 users per year

Advanced Light Source

National Energy Research Scientific Computing Center (NERSC)

88-Inch Cyclotron

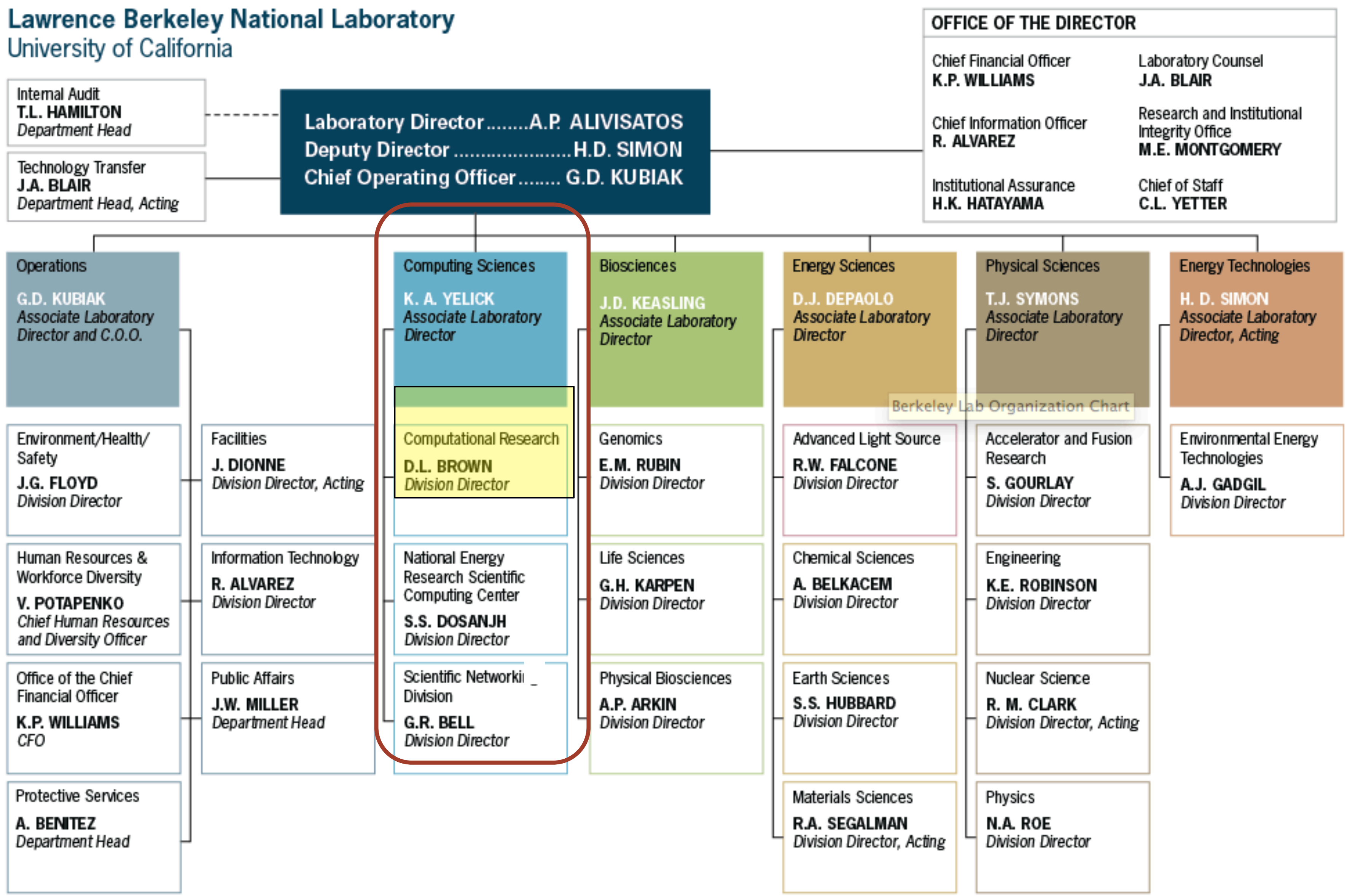
Molecular Foundry

Energy Scences Network (ESnet)

Joint Genome Institute

National Center for Electron Microscopy

We are working to enable new experimental science capabilities through advanced math and computing



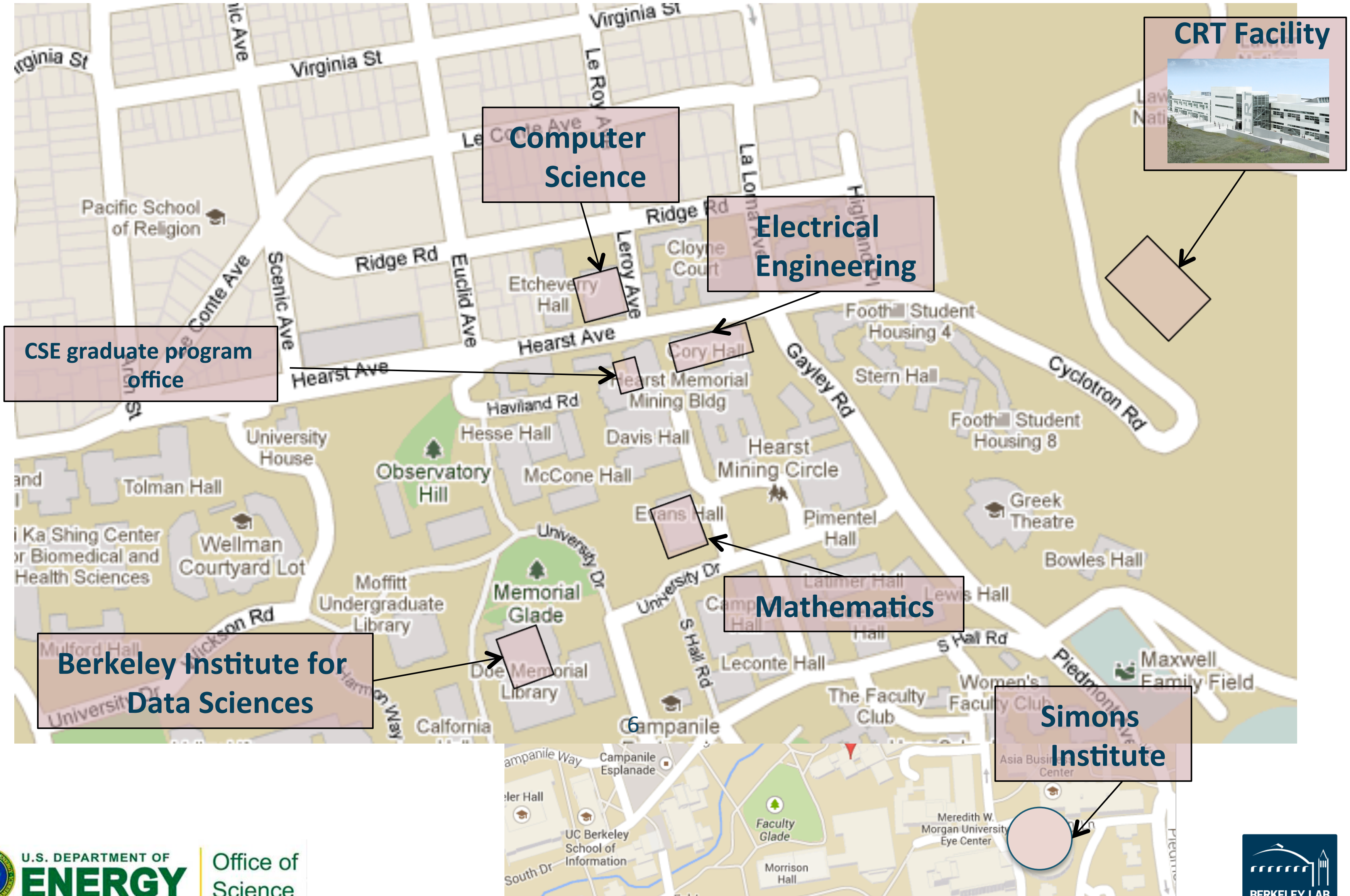
U.S. DEPARTMENT OF ENERGY | Office of Science

University of California at Berkeley

BERKELEY LAB
Lawrence Berkeley National Laboratory


UC's CRT
facility in
construction

CRT Location Supports Collaboration with University of California at Berkeley

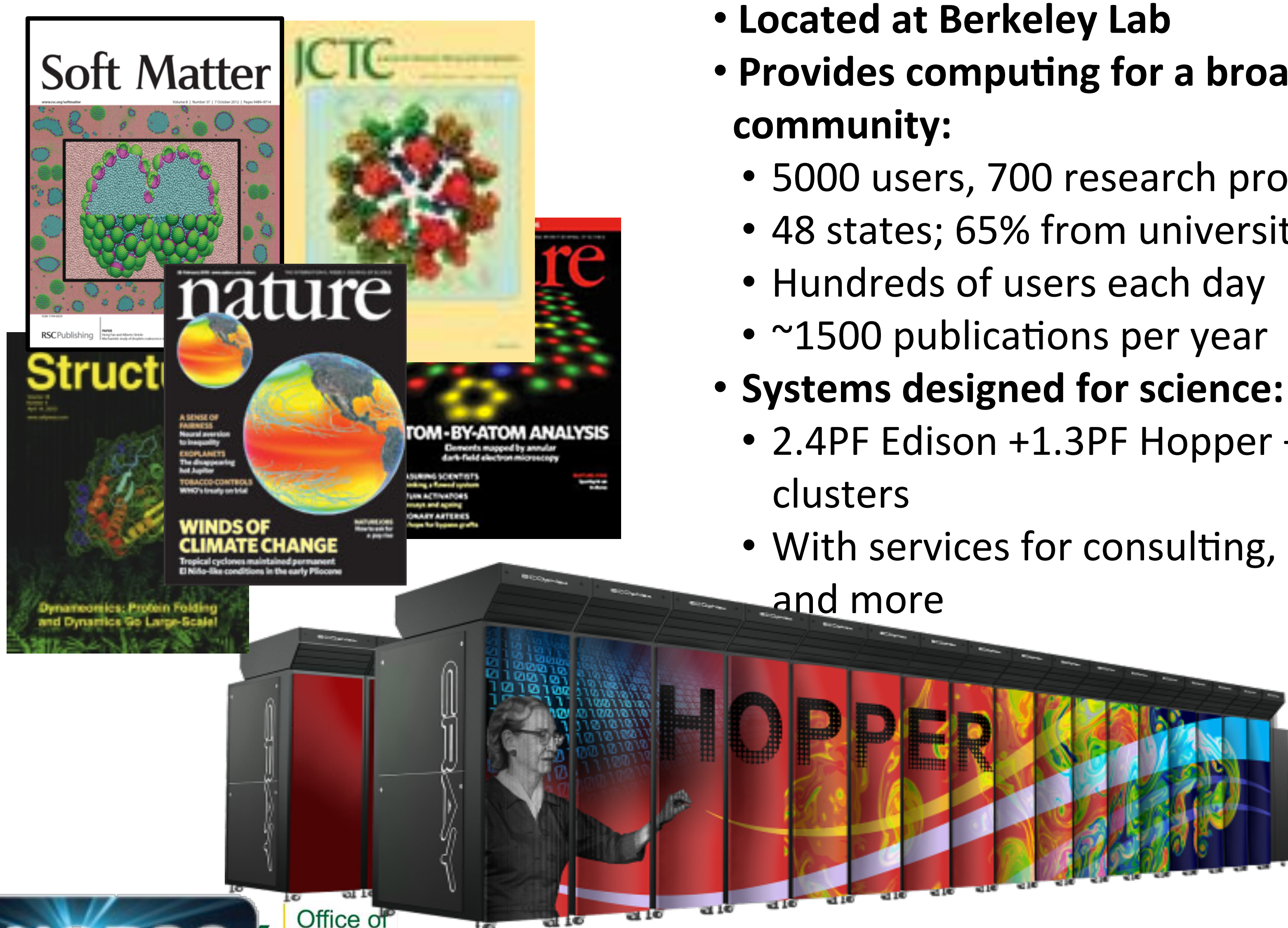


U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

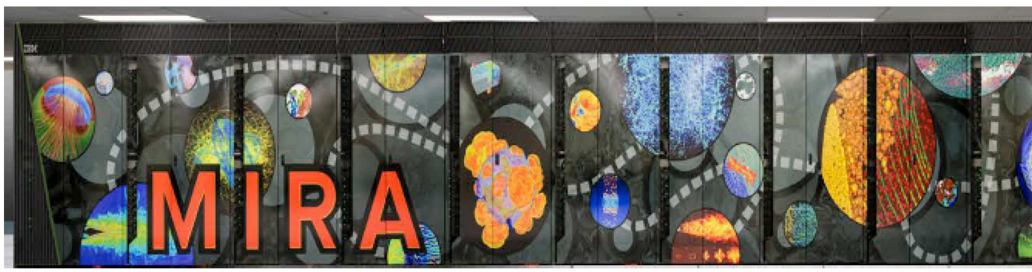
NERSC at Berkeley Lab provides HPC Systems for Science

- **Located at Berkeley Lab**
- **Provides computing for a broad science community:**
  - 5000 users, 700 research projects
  - 48 states; 65% from universities
  - Hundreds of users each day
  - ~1500 publications per year
- **Systems designed for science:**
  - 2.4PF Edison +1.3PF Hopper + .5 PF clusters
  - With services for consulting, data analysis and more



US DOE Office of Science Supercomputers at Oak Ridge, Argonne and Berkeley Labs

Titan:
- Peak performance of 27.1 Petaflops
- 18,688 Hybrid Compute Nodes
- 8.9 MW peak power



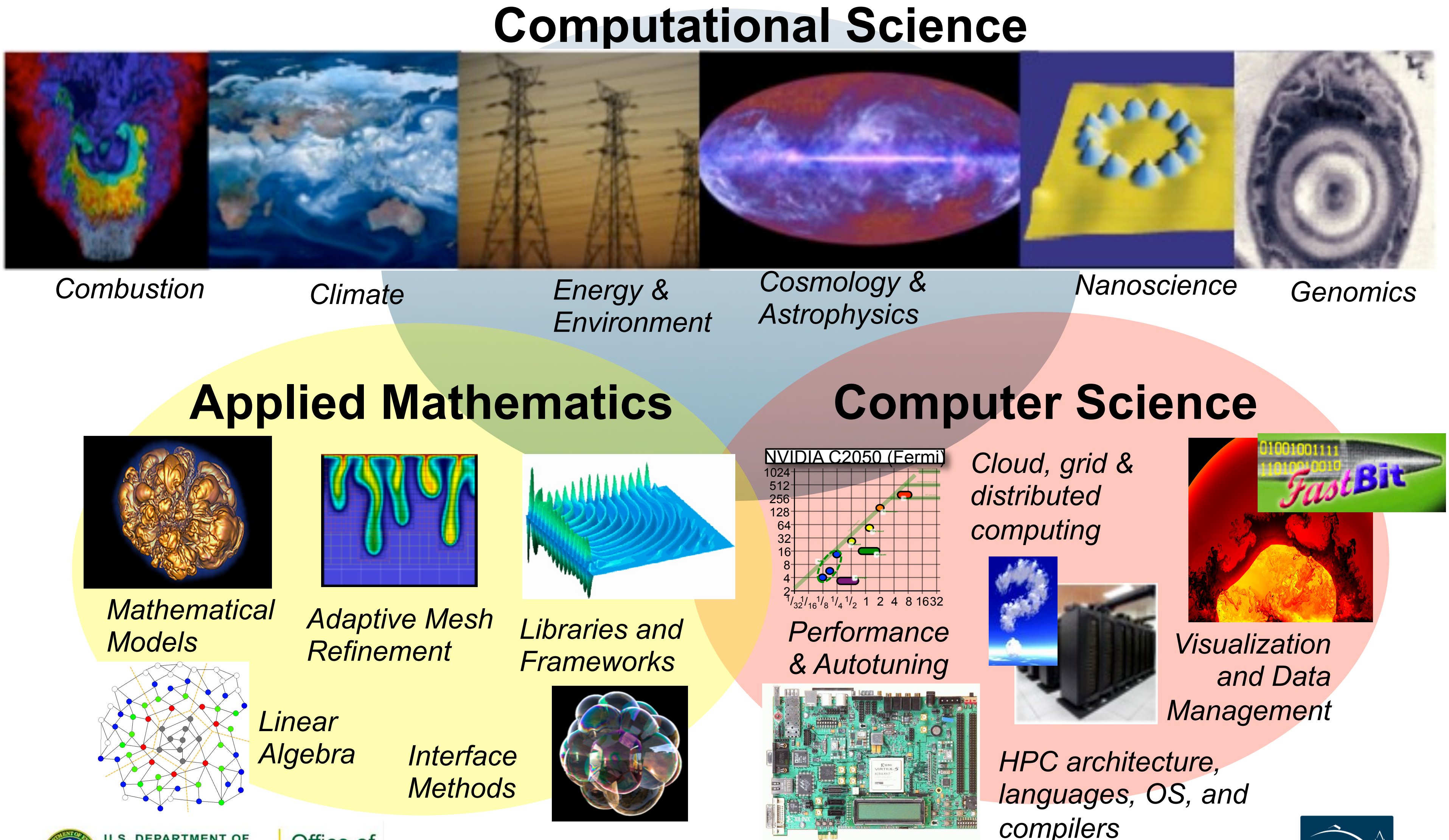
Mira:
- Peak performance of 10 Petaflops
- 49,152 Compute Nodes
- 4.8 MW peak power

Edison XC30:
- Peak performance 2.4PF
- 5,576 Compute Nodes
- 2.1 MW peak power

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Computational Research Division



U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

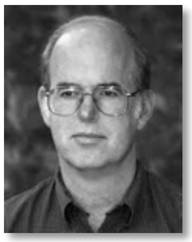
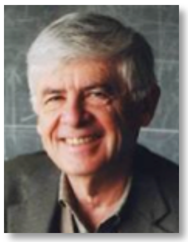
Computing Sciences Strategy Today

Extreme Computing	• **Goal:** Increase ops/Watt by 100x • Hardware, software and algorithms
Extreme Data	• **Goal:** Answer facility data deluge • Software, algorithms, and systems
Science Engagement	• **Goal:** Improve science capabilities • Mathematics and software
Computing Fundamentals	• **Goal:** Advance foundations of computing and math relevant to DOE

Applied math: broad impact from basic algorithms to scientific discovery

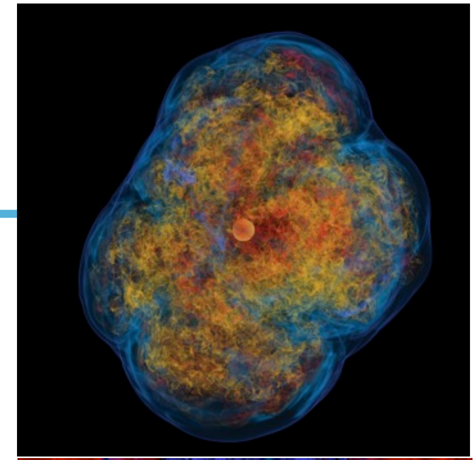
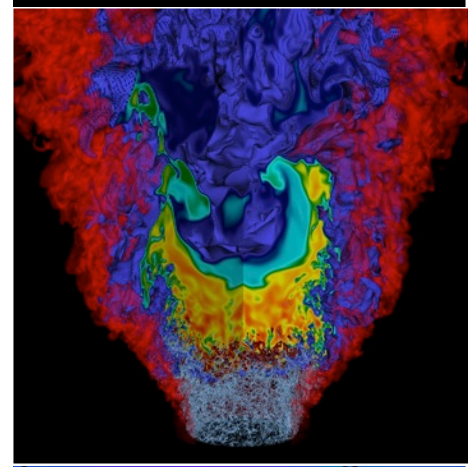
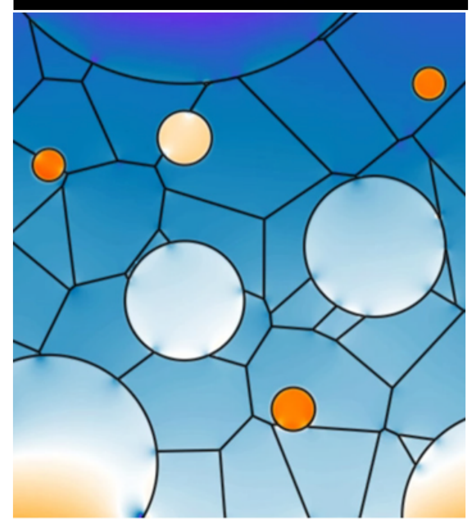
NATIONAL ACADEMY OF SCIENCES
1863–2013
Celebrating 150 Years of Service to the Nation

NATIONAL ACADEMY OF ENGINEERING

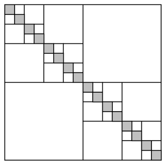


- **Pioneering methodology to solve previously inaccessible problems**
  - Level sets
  - Adaptive Mesh Refinement
  - High-order upwind schemes
  - Advanced eigenvalue solvers
  - Communication-avoiding algorithms
- **Software tools that enable science not previously possible**
  - Chombo, BoxLib, SuperLU, LAPACK …
- **Broad scientific impact**
  - *climate*
  - *combustion*
  - *chemistry and materials science*
  - *cosmology*
  - *astrophysics*
  - *environmental modeling*
  - *fusion*

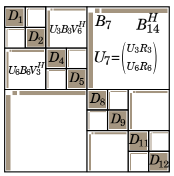




U.S. DEPARTMENT OF ENERGY | Office of Science

New numerical linear algebra techniques have impact on seismic inversion, accelerator design



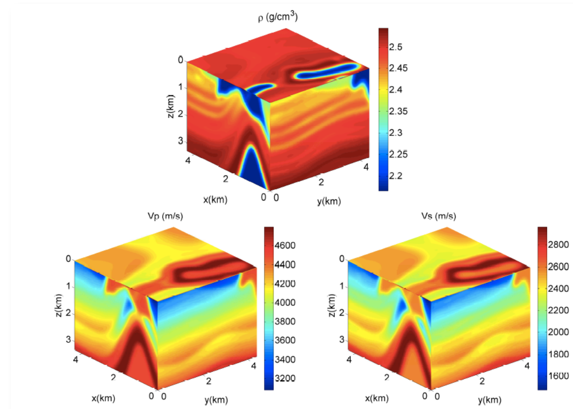
hierarchical partition



SVD with nested bases

- **Hsolver: scalable HSS-embedded low-rank sparse solvers and preconditioners**
  - For dense, but data-sparse structured matrices with rank-deficient off-diagonal blocks
  - Hierarchical partitioning
  - Nearly linear complexity

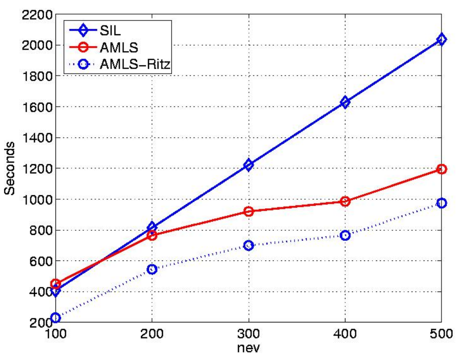
Sherry Li, Artem Napov, Francois-Henry Rouet, Jianlin Xia



3D anisotropic Helmholtz for seismic imaging
16,000+ cores, 2x faster, 1/5 memory

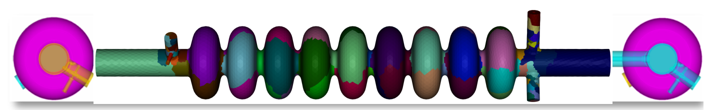
- **ASEIG: Algebraic sub-structuring eigensolver for $K\,x = \lambda M\,x$**
  - Graph partitioning divides matrix into submatrices
  - Projection space based on eigenspectra subsets
  - Compute approximate solution from projection space
  - Advantages over Shift-Invert-Lanczos: no triangular solves, orthogonalization, less memory

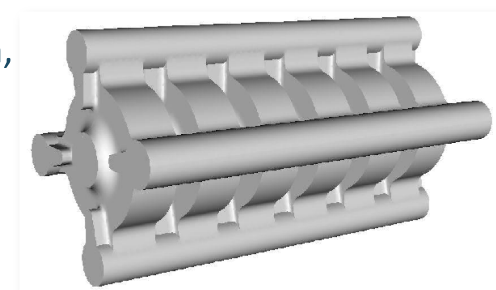
Weiguo Gao, Sherry Li, Chao Yang, Zhaojun Bai



ASEIG faster, less memory than SIL when many eigenpairs are needed;

ASEIG used in accelerator cavity design code for SLAC National Accelerator Lab



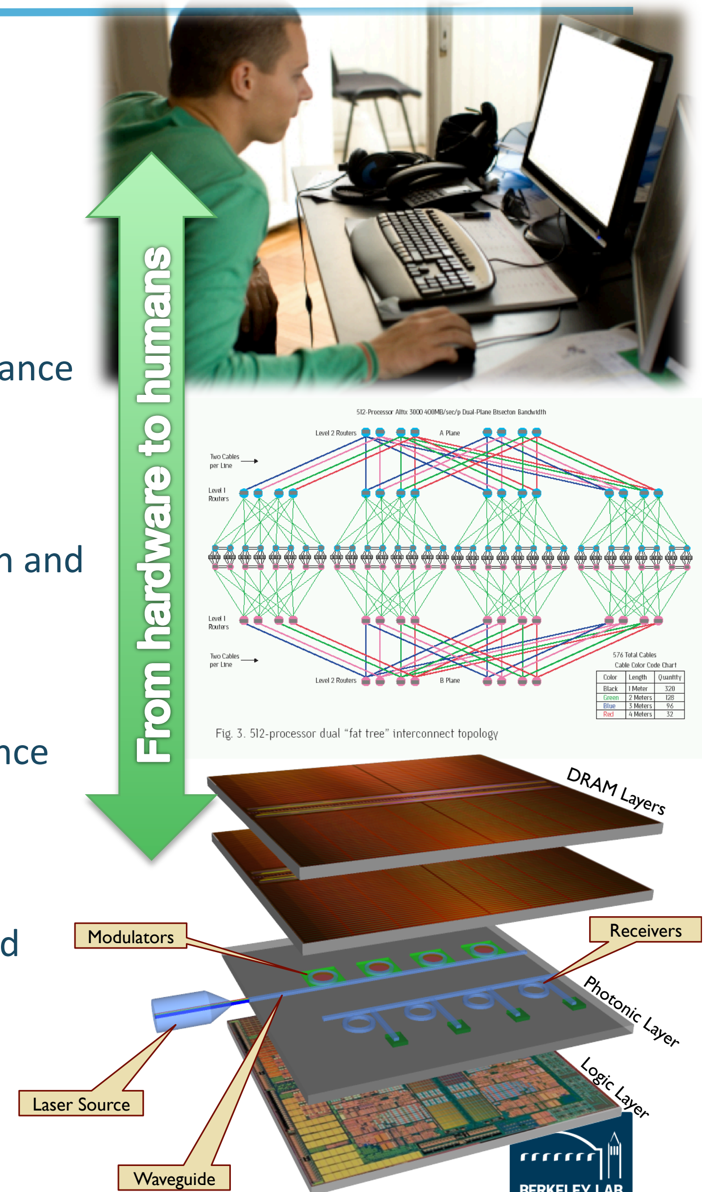


U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Our Strong CS Research Program at Berkeley Lab addresses the Challenge of Technology Disruption

- **Programming models and Languages**
  - Advanced programming environment addressing emerging technology challenges
  - Enable productive programming of exascale systems
  - *UPC, Titanium, DEGAS*
- **Auto-tuning and Performance Analysis**
  - Performance portability through automated runtime performance analysis and code transformations
  - *Roofline model, OSKI, X-Tune, SEJITS*
- **Communication Libraries and Resilience**
  - Addressing advanced communication facilities, error detection and recovery for future system s/w architecture
  - *GASnet, Berkeley Linux Checkpoint Restart (BLCR), CD*
- **Manycore Operating Systems and Runtimes**
  - Deconstructing the OS for manycore processors for performance isolation, error isolation, resource management
  - *K42, Tessellation, Speed Balancing*
- **Hardware Architecture and CoDesign**
  - Compiler-based automated performance model extraction and hardware simulation/simulation for ExaCT and CAL
  - *Design: Green Flash/Green Wave, View from Berkeley*
  - *Emulation: CAL, RAMP*

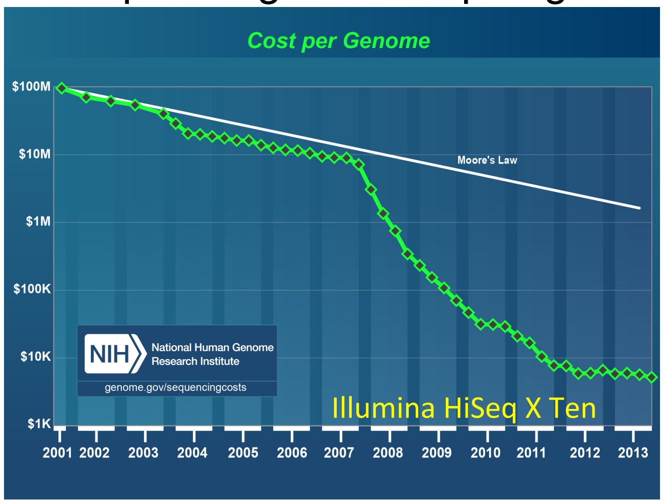


U.S. DEPARTMENT OF ENERGY | Office of Science

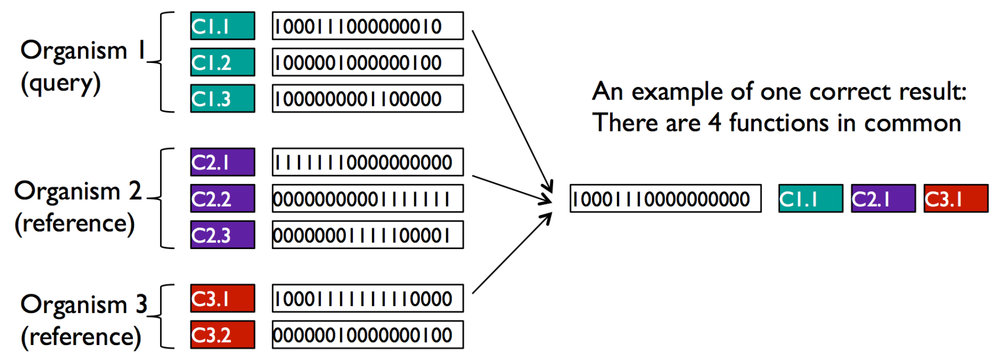
BERKELEY LAB

Data in Bioinformatics Requires New Algorithms, Software, and Systems

Sequencing Cost Drops Again



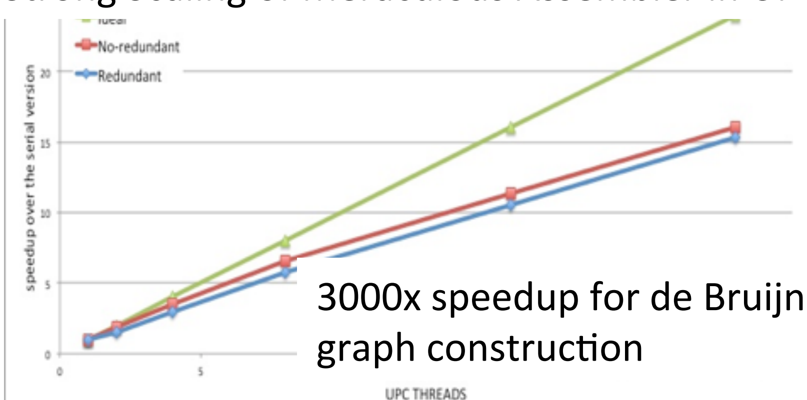
Gene Context Analysis now >1000x faster using FastBit



A. Romosan, A. Shoshani, K.Wu, V. Markowitz, K. Mavrommatis

Human: 2B reads → 20M contigs → 16K scaffolds → 23 chromosomes

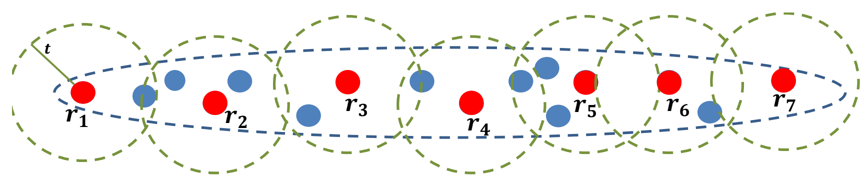
Strong Scaling of Meraculous Assembler in UPC



3000x speedup for de Bruijn graph construction

Probabilistic counting, Bloom filters, Distributed hash

New ~Linear Time Bubble Clustering algorithm



ENERGY | Office of Science

A. Buluc, V. Strnadova. J. Gilbert, E. Georganas, J. Chapmanz, D. Rokhsar, L. Oliker, and K. Yelick,, multiple papers to be submitted

BERKELEY LAB

Expansion of Computational Science at Berkeley Lab driven by challenges of data-intensive experiments and observations

- **We have a strong record of innovation in computational science, mostly modeling and simulation**
  - Major SciDAC role – including climate, astrophysics
  - Outstanding applied math
  - Visionary computer science
  - Lead DOE's data intensive computing center SDAV
  - NERSC and ESnet are major DOE computational facilities

- **But our science facilities are largely experimental in nature – data analysis has not been an HPC problem***

- **Exploding data intensive analysis demands means opportunity for significant growth in computing for science at LBNL**
  - Beamline data
  - Materials discovery - database approaches
  - Bio-image analysis
  - Genomics data - JGI
  - Modeling and Simulation to help understand observations – e.g. cosmology

Advanced Light Source

National Energy Research Scientific Computing Center (NERSC)

88-Inch Cyclotron

Molecular Foundry

Energy Sciences Network (ESnet)

Joint Genome Institute

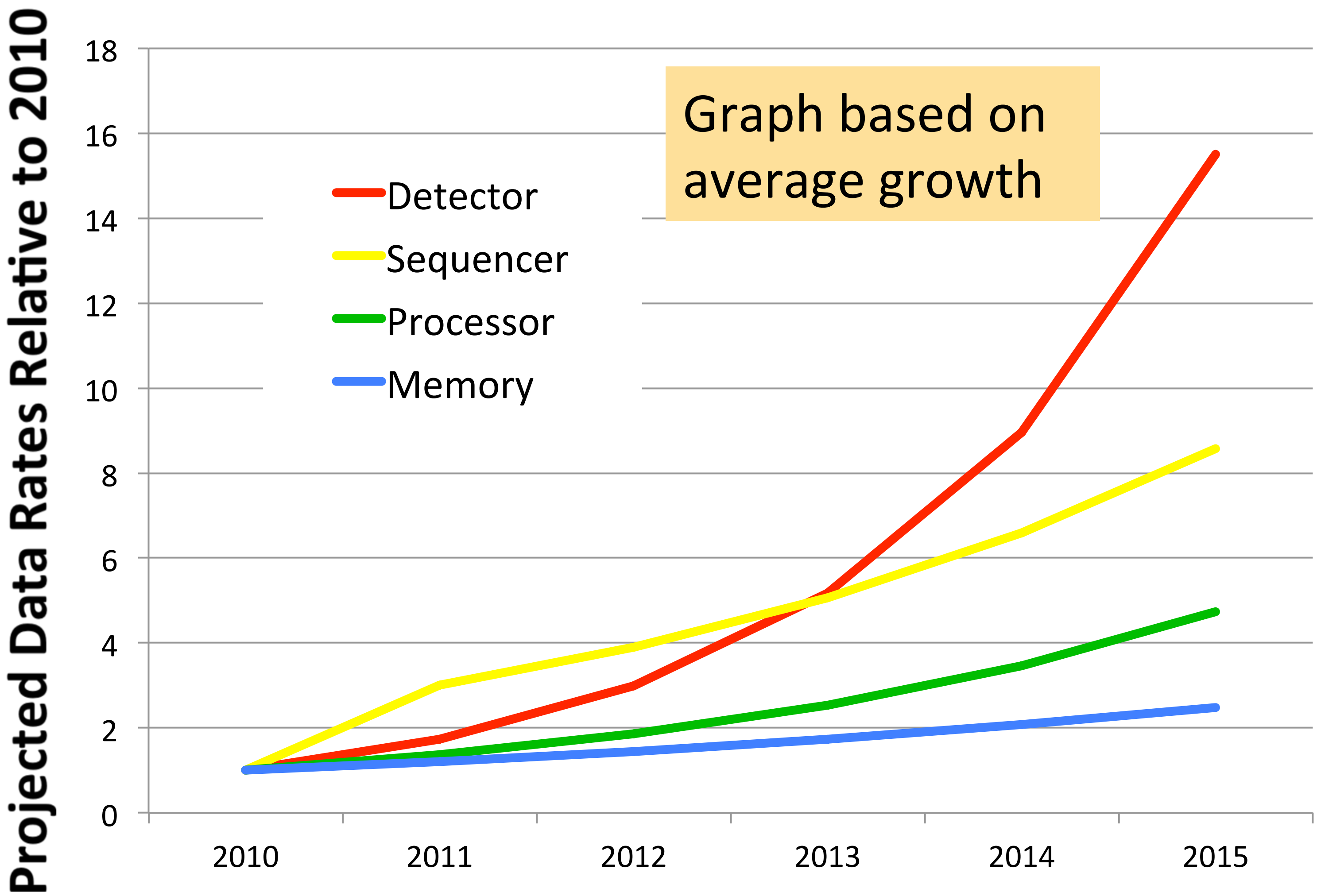
National Center for Electron Microscopy

Berkeley Lab's largest **research facilities** see more than 25,000 users per year

* Possibly excluding JGI

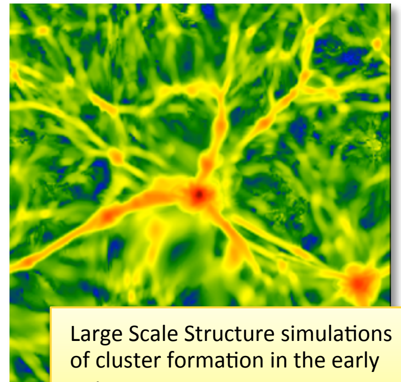
BERKELEY LAB

Data Growth is Outpacing Computing Growth

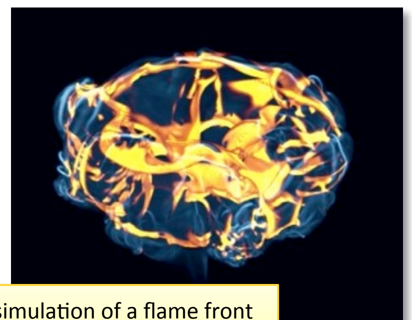


Data-driven scientific discovery requires integration of modeling, simulation, analysis, data management

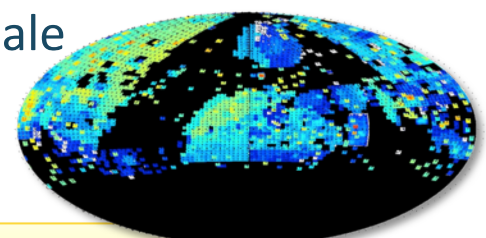
- **Example: 21st Century Astrophysics:**
  - Tight collaboration between astrophysicists and computational scientists to develop new technologies for cosmological data analysis
  - Analysis & simulation of 100s of TeraBytes of data from ground- and space-based observations
  - Modeling & simulation of supernovae and large-scale structure formation



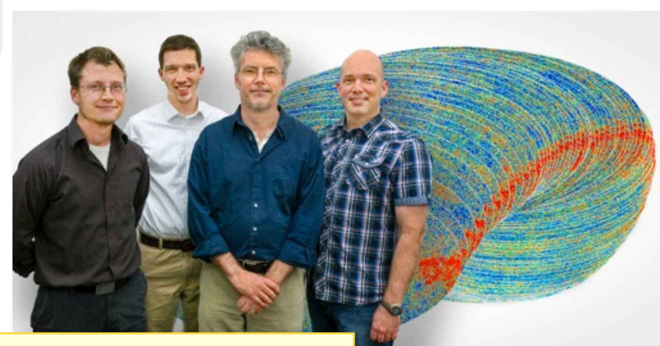
Large Scale Structure simulations of cluster formation in the early universe



Hydro simulation of a flame front in a thermonuclear supernova explosion



Palomar Transient Factory data-analysis sky-coverage map for the first 3 years of the project



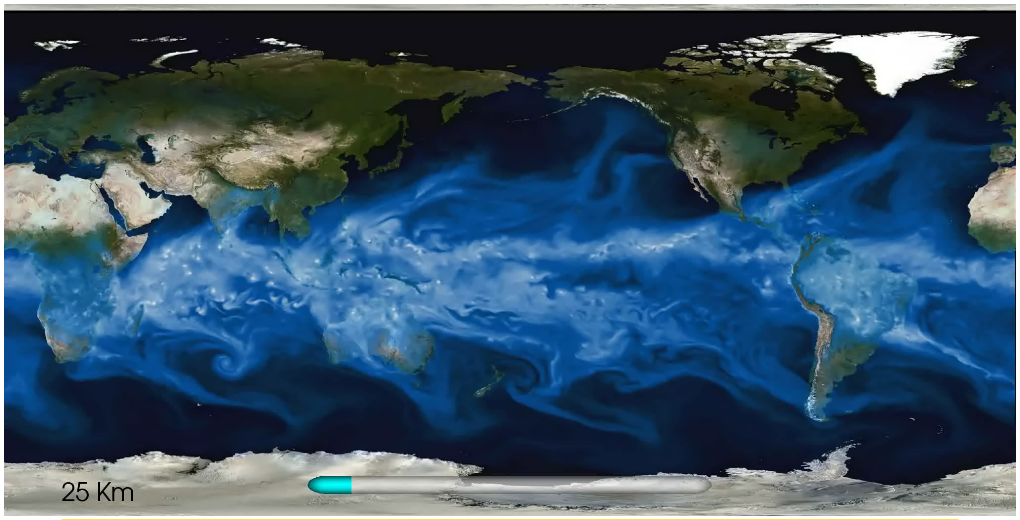
Cosmic Microwave Background Radiation data from Planck

U.S. DEPARTMENT OF ENERGY | Office of Science

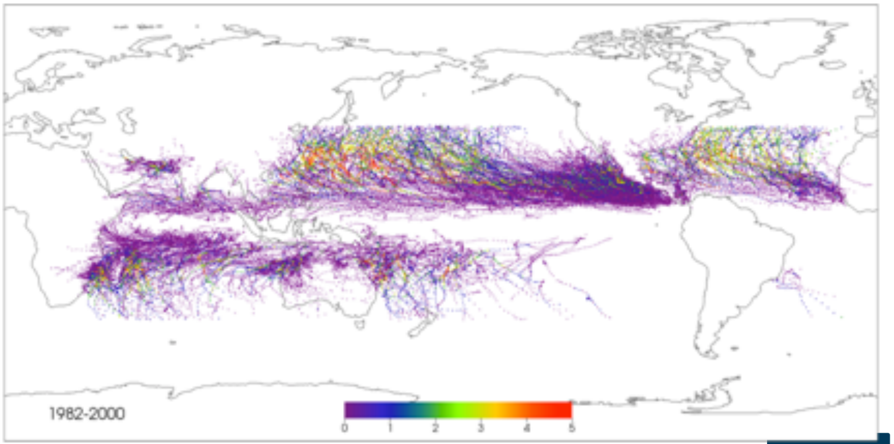
BERKELEY LAB

Advanced analytic techniques needed to identify interesting events in climate data

- New data mining, image processing, and topological analysis techniques
- Detect extreme events, e.g., hurricanes
- 10s-100s of TB of data from a single simulation
  - (The Library of Congress contains about 254 TB.)
- TECA Toolkit
  - Automatic detection of cyclones, atmospheric rivers, and more
  - Analysis time years to minutes



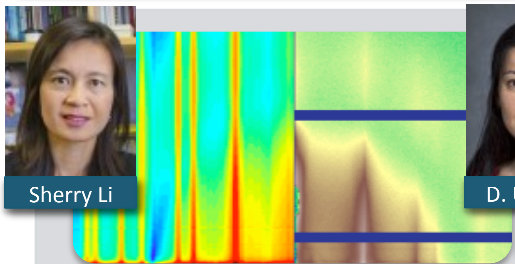
High-resolution atmospheric simulation

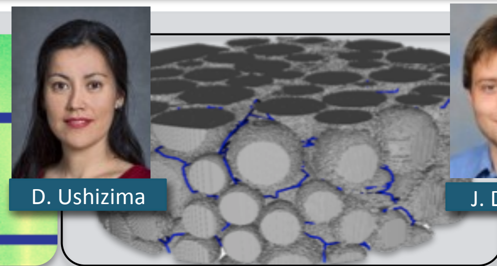


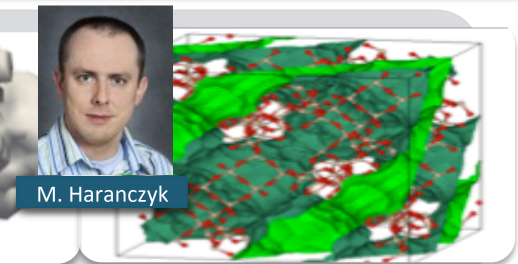
Detected cyclones

CAMERA leverages state-of-the-art mathematics to transform experimental data into understanding

CENTER FOR APPLIED MATHEMATICS FOR ENERGY RESEARCH APPLICATIONS







X-ray scattering data analysis

HipGISAXS 400-1500x faster analysis for X-ray scattering data

Now: Nonlinear optimization, genetic algorithms, pattern recognition w/ noise

Micro-CT Sample Analysis

Quant-CT provides automated quantitative analysis

Now: 3D image segmentation; pattern recognition; classification; PDE- and graph-based analysis

X-ray Nano-Crystallographic Reconstruction

Indexing ambiguity resolved [PNAS13]

Now: Image orientation, find crystal shape/size; address orientation ambiguities; data variance reduction

Designing New Materials

Designed record-breaking high-surface area materials

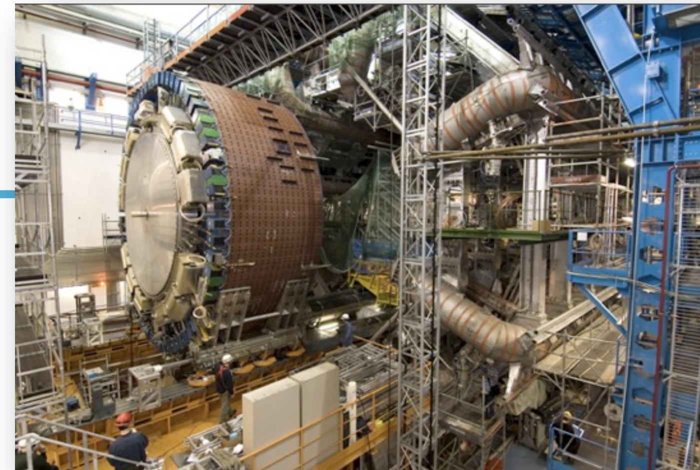
Now: 3D porous polymer model assembly; Zeo++ porosity character-ization; Optimal material designs

U.S. DEPARTMENT OF ENERGY

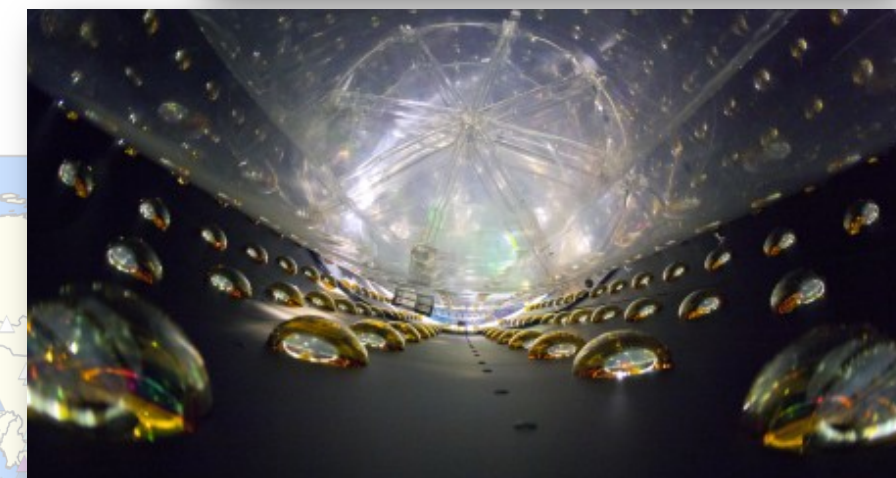
LDRD-seeded, now ASCR-supported

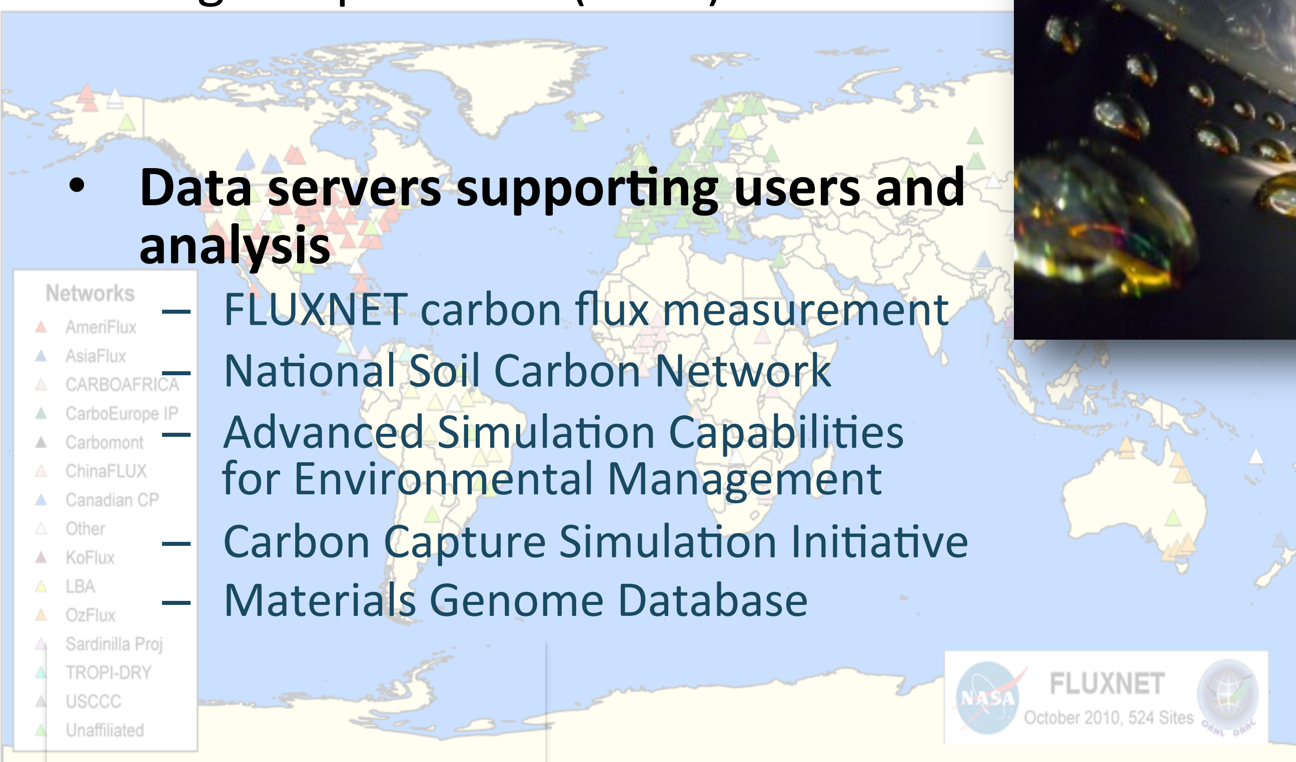
BERKELEY LAB

Robust software infrastructure for modern scientific discovery

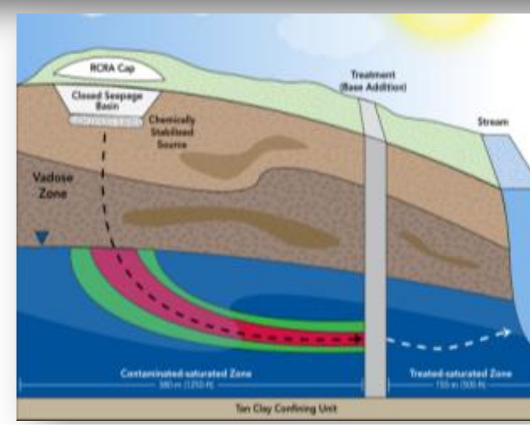


- **ATLAS (CERN):** LBNL responsible for core software and infrastructure (since 1999) – 8 FTE effort
- **Daya Bay:** LBNL architected software for anti-neutrino mixing angle experiment (2012)





- **Data servers supporting users and analysis**
  - FLUXNET carbon flux measurement
  - National Soil Carbon Network
  - Advanced Simulation Capabilities for Environmental Management
  - Carbon Capture Simulation Initiative
  - Materials Genome Database



SPOT Suite: New Usage Model for Light Sources

- **Light source data solution:**
  - Data management and storage
  - Analysis and simulation
- **Inter-disciplinary team:**
  - Computer scientists
  - light source scientists
  - materials scientists
  - ESNet & NERSC
- ***In-situ, time-resolved experiments,***
- ***Real-time analysis, algorithms,***
- ***Data sharing & collaboration***

Since March 2013: 40 experiments; 1,356 datasets; 16 TB of raw data; 346,371 NERSC jobs delivered in real time to BL 8.3.2 Micro-Tomography Beamline.

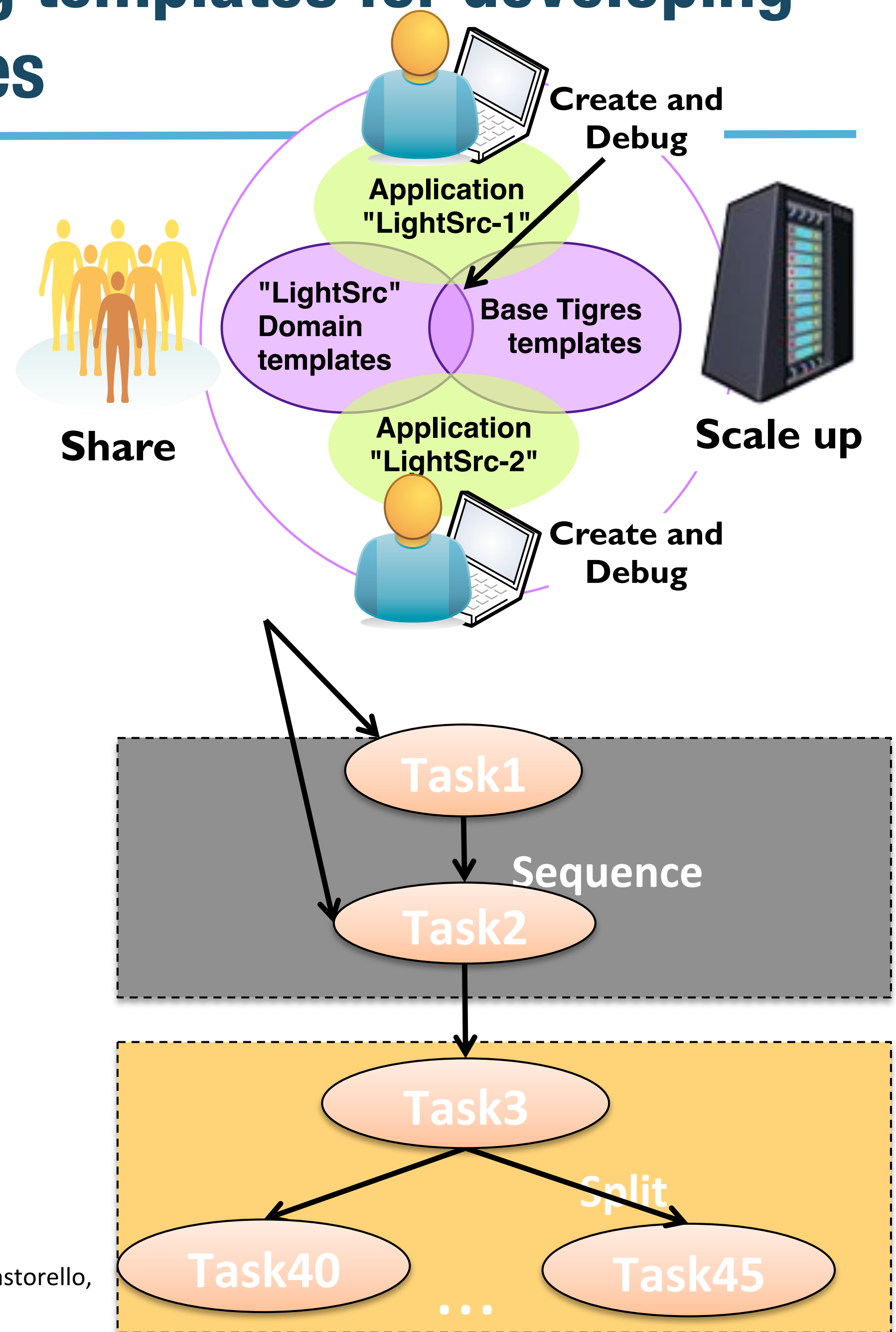
U.S. DEPARTMENT OF ENERGY | Office of Science

Craig Tull, Alexander Hexemer, Dula Parkinson

BERKELEY LAB

Tigres provides programming templates for developing scientific workflow pipelines

- **Science targets:**
  - *Image processing*
  - *Light source experiments*
  - *cosmology*
  - *DNA assembly*
  - *UQ pipelines*
- **Design Goals:**
  - Easy composition
  - Scalability
  - Lightweight execution
- **Templates:**
  - *Sequence*
  - *Parallel*
  - *Split*
  - *Merge*



L. Ramakrishnan, V. Hendrix, D. Gunter, G.Pastorello, R. Rodriguez, A. Essari , D. Agarwal

We hope to identify collaboration opportunities between Berkeley Lab and CCS during this meeting

- *Michael Wehner:* **Simulating Hurricanes and Typhoons in a Global Climate Model**
- *Esmond Ng:* **Large-scale Eigenvalue Calculations in Scientific Applications**
- *Costin Iancu:* **Dynamic Exascale Global Address Space (DEGAS) Programming Environments**
- *Casey Stark:* **The Lyman-alpha Forest in Cosmological Hydrodynamic Simulations**
- *Khalid Ibrahim:* **Analysis and Optimization of Gyrokinetic Toroidal Simulation on Emerging Architectures**