BERKELEY LAB

LAWRENCE BERKELEY NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Computational Research at Berkeley Lab

David L. Brown, Director

dlb@lbl.gov

Computational Research Division

March 2012

❖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

People

- *3,863 FTE*
- *3,040 Employees*
- *267 Joint faculty*
- *491 Postdoctoral researchers*
- *328 Graduate students*
- *194 Undergraduates*
- *8,025 Facility users*
- *1,612 Visiting scientists and engineers*

FY10 Total Operating Costs: US $680.6M

LBNL at-a-glance

Advanced Light Source

National Energy Research Scientific Computing Center (NERSC)

88-Inch Cyclotron

Molecular Foundry

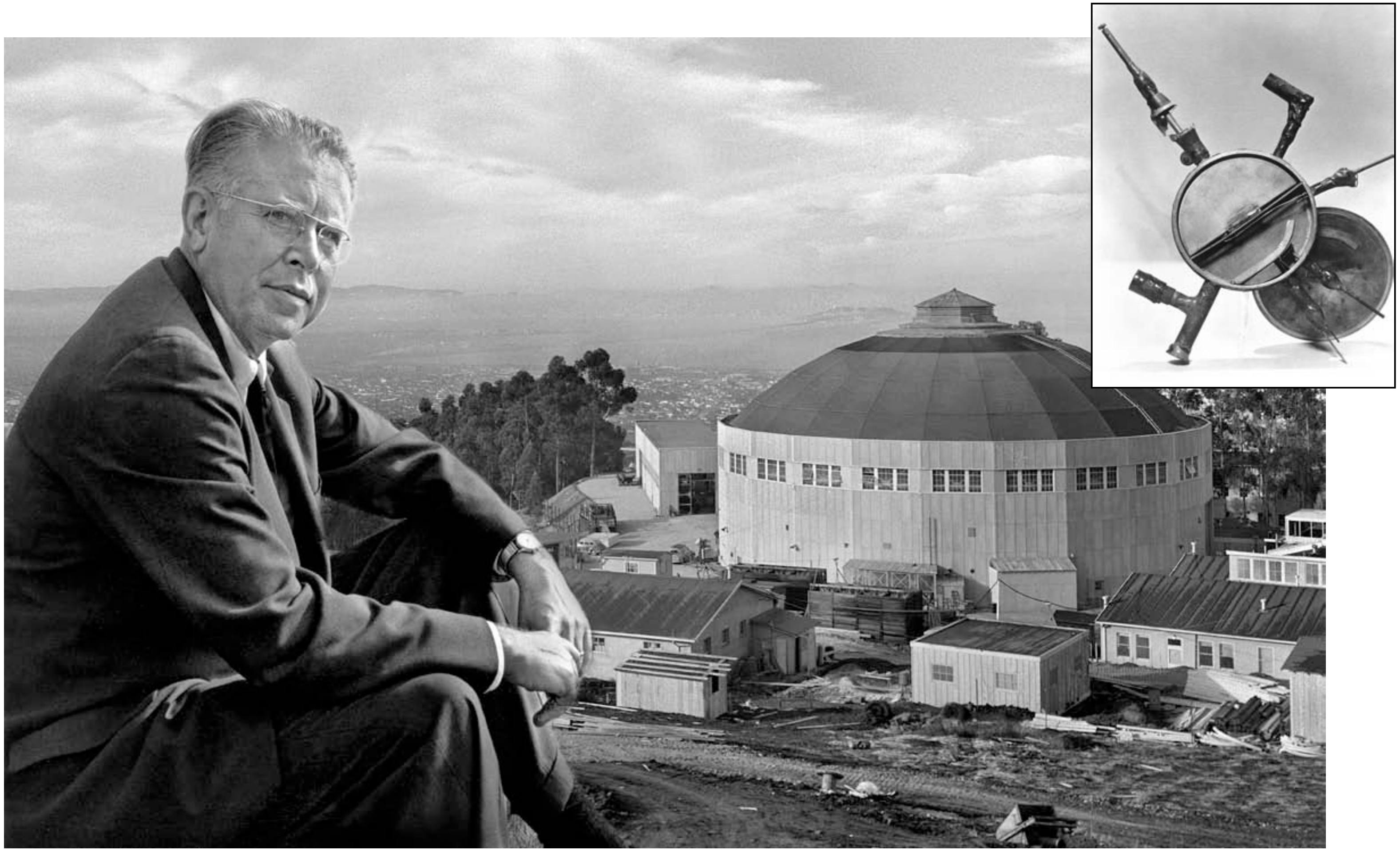
Energy Sciences Network (ESnet)

Joint Genome Institute

National Center for Electron Microscopy

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

Founded on the UC Berkeley campus in 1931, moved to the current site in 1940



Lawrence Introduces Big Team Science in 1931
LBNL: The First DOE National Laboratory

Berkeley Lab Organization

BERKELEY LAB

Ernest Orlando Lawrence Berkeley National Laboratory
University of California

OFFICE OF THE DIRECTOR

- Chief Financial Officer — **J.A. FERNANDEZ**
- Laboratory Counsel — **J.A. BLAIR**
- Chief Information Officer — **R. ALVAREZ**
- Research and Institutional Integrity Office — **M. E. MONTGOMERY**
- Institutional Assurance — **H. K. HATAYAMA**
- Chief of Staff — **C.L. YETTER**

Internal Audit — **T.L. HAMILTON**

Technology Transfer — **C.A. FRAGIADAKIS**, *Department Head*

Laboratory Director.........A.P. ALIVISATOS
Deputy Director .....................H.D. SIMON
Chief Operating Officer..... J.T. KRUPNICK

Operations
J.T. KRUPNICK, *Associate Laboratory Director and C.O.O.*
R. RISHELL, *Deputy C.O.O.*

- Facilities — **J.S. RIDGEWAY**, *Division Director*
- Information Technology — **R. ALVAREZ**, *Division Director*
- Public Affairs — **J.W. MILLER**, *Department Head*
- Environment, Health & Safety — **D.M. FLEMING**, *Division Director*
- Office of the Chief Financial Officer — **J.A. FERNANDEZ**, *CFO*
- Human Resources & Workforce Diversity — **V. POTAPENKO**, *Chief Human Resources Officer*

Computing Sciences
K. A. YELICK, *Associate Laboratory Director*

- Computational Research — **D.L. BROWN**, *Division Director*
- National Energy Research Scientific Computing Center — **K. A. YELICK**, *Division Director*
- Scientific Networking Division — **G.R. BELL**, *Division Director, Acting*

Biosciences
J.D. KEASLING, *Associate Laboratory Director*

- Physical Biosciences — **A. P. ARKIN**, *Division Director*
- Genomics — **E.M RUBIN**, *Division Director*
- Life Sciences — **G.H. KARPEN**, *Division Director*

Energy & Environmental Sciences
D.J. DEPAOLO, *Associate Laboratory Director*

- Chemical Sciences — **S.R. LEONE**, *Division Director*
- Environmental Energy Technologies — **A.J. GADGIL**, *Division Director*
- Materials Sciences — **M.B. SALMERON**, *Division Director*
- Earth Sciences — **D.J. DEPAOLO**, *Division Director*

General Sciences
T.J. SYMONS, *Associate Laboratory Director*

- Accelerator and Fusion Research — **S. GOURLAY**, *Division Director*
- Engineering — **K.E. ROBINSON**, *Division Director*
- Nuclear Science — **T.J. SYMONS**, *Division Director*
- Physics — **D.R. NYGREN**, *Division Director, Acting*

Advanced Light Source — **R.W. FALCONE**, *Division Director*

12/05/2011

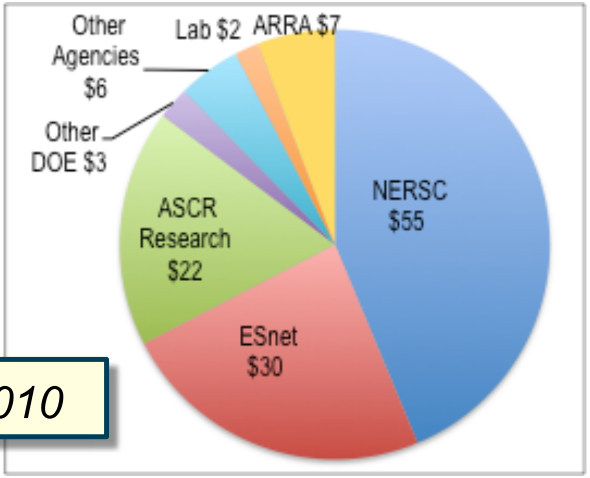
DAVID L. BROWN
Division Director

- **ADVANCED COMPUTING FOR SCIENCE DEPARTMENT** — *DEBORAH AGARWAL, Department Head*
  - **CYBERINFRASTRUCTURE DEVELOPMENT** — *DEBORAH AGARWAL, Acting Group Leader*
  - **DATA INTENSIVE SYSTEMS** — *DEBORAH AGARWAL, Group Leader*
  - **METADATA, SEMANTICS, AND ECOINFORMATICS** — *BRUCE BARGMEYER, Group Leader*
  - **SCIENCE SOFTWARE SYSTEMS** — *CRAIG TULL, Group Leader*
- **APPLIED MATHEMATICS & SCIENTIFIC COMPUTING DEPARTMENT** — *ESMOND NG, Department Head*
  - **APPLIED NUMERICAL ALGORITHMS** — *PHILLIP COLELLA, Group Leader, (FB)*
  - **COMPLEX SYSTEMS** — *DAVID BAILEY, Group Leader*
  - **SCIENTIFIC COMPUTING** — *ESMOND NG, Group Leader*
- **BIOLOGICAL DATA MANAGEMENT & TECHNOLOGY CENTER** — *VICTOR MARKOWITZ, Department Head*
  - **BIOLOGICAL DATA ANAYLYSIS** — *KEN CHU, Group Leader*
  - **BIOLOGICAL DATABASE SYSTEMS** — *AMY CHEN, Group Leader*
- **COMPUTER & DATA SCIENCES DEPARTMENT** — *JOHN SHALF, Department Head*
  - **ADVANCED TECHNOLOGIES** — *JOHN SHALF, Group Leader*
  - **FUTURE TECHNOLOGIES** — *ERICH STROHMAIER, Group Leader*
  - **SCIENTIFIC DATA MANAGEMENT** — *ARIE SHOSHANI, Group Leader*
  - **VISUALIZATION** — *E. WES BETHEL, Group Leader*
- **MATHEMATICS & COMPUTATIONAL SCIENCES DEPARTMENT** — *JOHN BELL, Department Head*
  - **CENTER COMPUTATIONAL SCIENCES & ENGINEERING** — *JOHN BELL, Group Leader*
  - **COMPUTATIONAL COSMOLOGY CENTER** — *PETER NUGENT, Group Leader, (FB)*
  - **MATHEMATICS** — *JAMES SETHIAN, Group Leader (FB)*

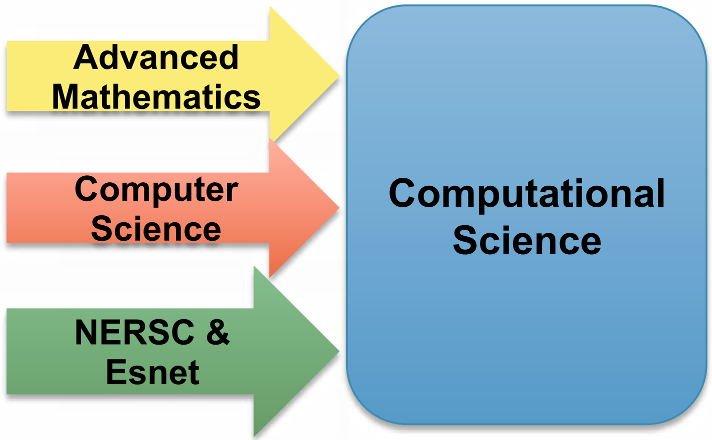
CS mission is to accelerate scientific discovery through advanced computing and mathematics

- **Through advanced research & facilities we are a leader in delivering scientific output from computing**

- **Key initiatives:**
  - *Exascale Computing:* Improve energy efficiency of computational science: more science per Watt
  - *Data enabled science:* Improve insight from experimental, observational and simulated data
  - *New Computational Science:* Bring advanced mathematics and computing to new science problems



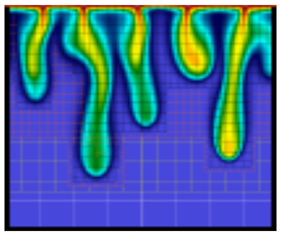
US$126M in 2010



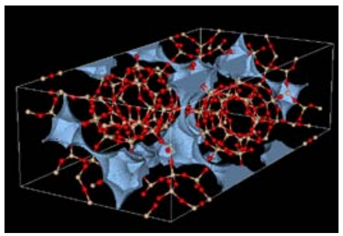
Applied Mathematics is key to new science and engineering innovations

- **New mathematical models** to improve computational science throughput
- **Architecture-aware algorithms** to exploit the changing computing hardware landscape
- **Analysis methods for large data** to supports discovery science at the major DOE science facilities
- **Multi-faceted mathematics** brings together multiple math and science domains to solve complex multi-scale multi-physics problems
- **Uncertainty quantification methods** provide rigorous confidence levels using simulation and experimental data

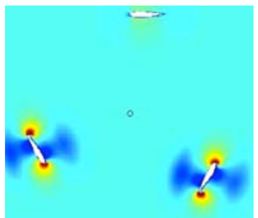
Low Mach number approximation



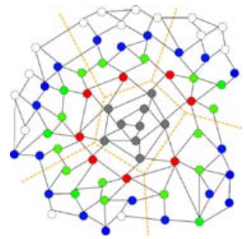
Adaptive Mesh Refinement



Level Set and Fast Marching Methods



Higher-Order Discontinuous Galerkin methods



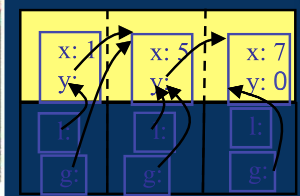
Communication-Avoiding Algorithms

State-of-art Computer Science R&D to advance scientific discovery

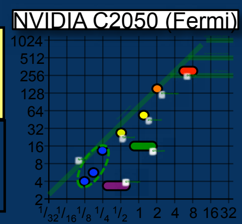
COMPUTATIONAL RESEARCH DIVISION

- **Advance computing technology:** Develop energy efficient architectures, programming models and systems software for exascale computing
- **Address the data exploration challenge:** Design algorithms and software for data analysis and exploration tools for huge data sets
- **Manage complex workflows:** Design software and workflow management for predictive simulation in secure, distributed environments
- **Build mathematical infrastructure:** Build mathematical libraries to encapsulate algorithms and methods for scientific productivity
- **Provide state-of-art networking for science:** Analyze technologies for terabit wide-area networks

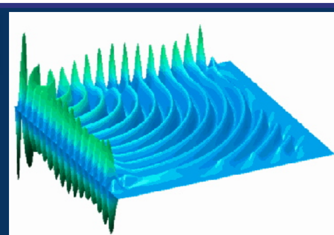
RAMP and CoDEx for hardware design



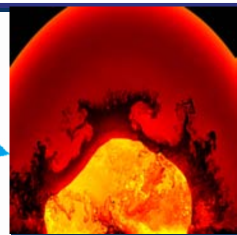
PGAS languages



Performance autotuning



SuperLU , Chombo, LAPACK

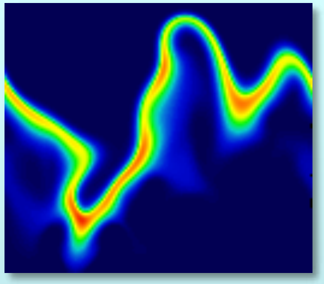


VisIt visualization

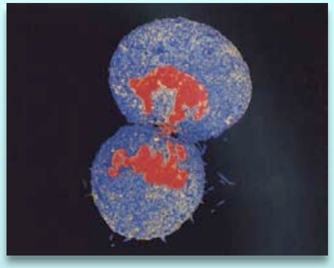
FastBit data algorithm

CRD is expanding the reach of computational science through innovative research

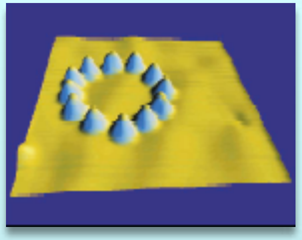
C O M P U T A T I O N A L   R E S E A R C H   D I V I S I O N

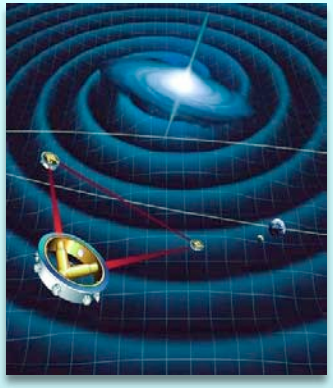

Combustion processes

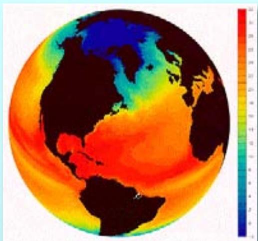

energy technology

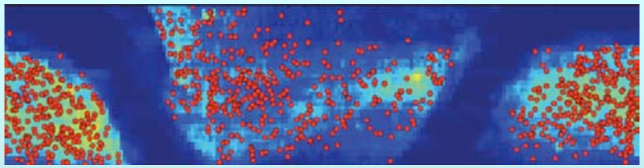

Biological systems

In addition to basic math and CS, CRD directly engages in computational science research


Nano systems


Astrophysics simulation


global climate


cosmology

NERSC provides HPC Systems for Science

- More users (4000) than any other DOE Science facility
- Over 1,500 publications per year
- Broad scientific coverage

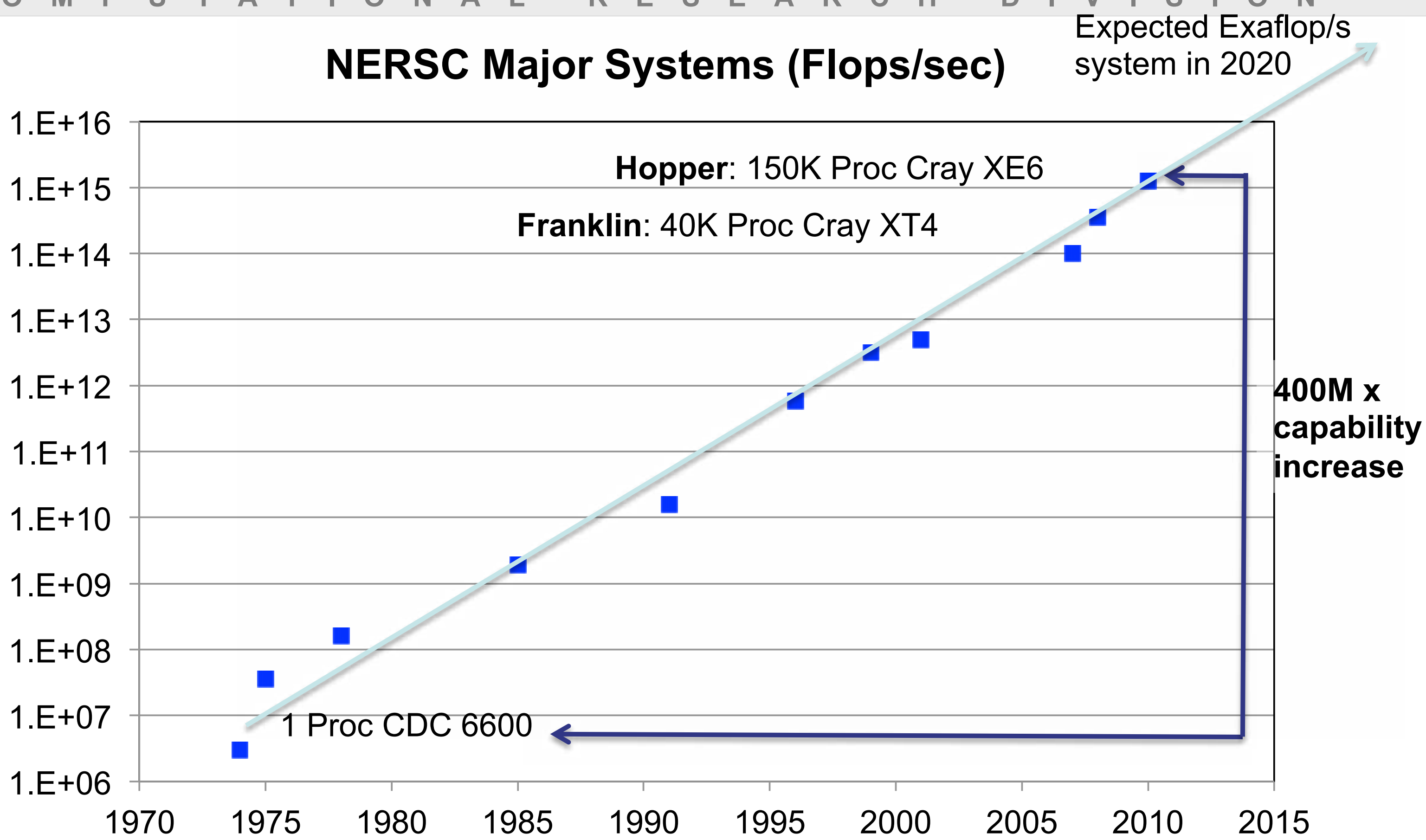
- Systems "optimized" for large science community, not peak performance

- Move the users with technology changes

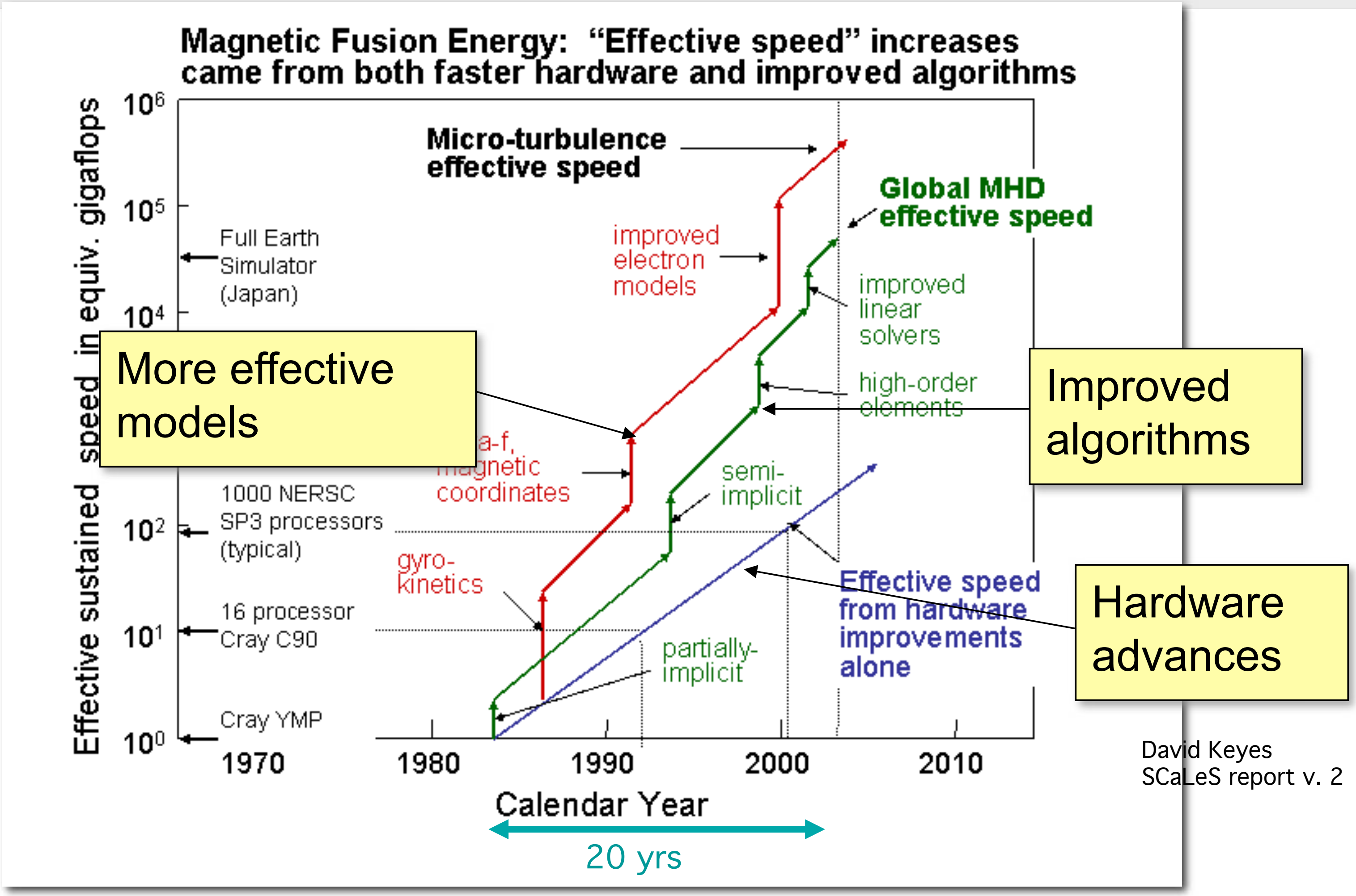
- Science is done across scales

1.25 Petaflop/s Hopper System 3 MWatts

Computing Capability at NERSC



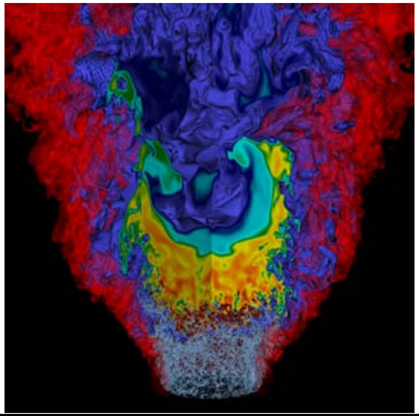
Effective HPC performance as much due to models and algorithms as to hardware advances



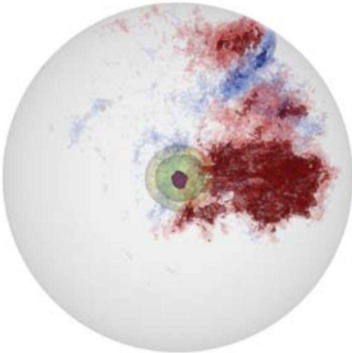
David Keyes
SCaLeS report v. 2

Transforming how we compute: "smart" math, numerics, HPC give unprecedented capability

- Low Mach number formulation
  - Exploit mathematical structure of the problem to give more efficient methods
- Advanced numerical methodology
  - Projection methodology
  - Adaptive mesh refinement– resolution where its needed
- Software framework for parallel implementation
  - Over 100x increase in throughput
- Science and technology impact
  - Simulation at previously inaccessible scales
  - New understanding of **turbulent combustion** for new combustion devices and processes
  - Understanding the conditions at **Type 1a supernova ignition**



Numerical simulation emissions in a low swirl burner fueled by hydrogen



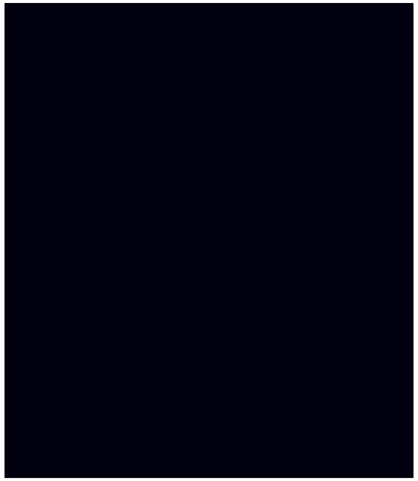
MAESTRO simulation near ignition showing flow from center of star and region of high energy generation

HPC can validate reduced-order models for multi-scale physics simulations

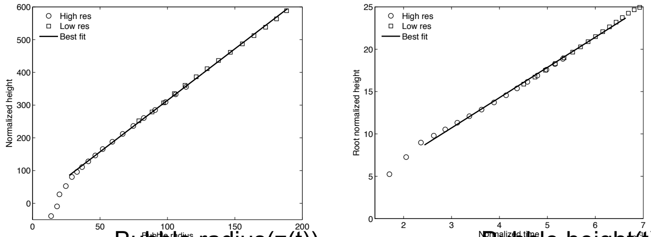
❖*Buoyant vortex ring ("thermal"):* In a Type Ia supernova a carbon fusion flame ignites near center, creating a small bubble of hot, buoyant ash surrounded by fuel. Sheer forces generate turbulence during rise, entraining fuel and increasing buoyancy.

❖PI's modified Morton-Taylor-Turner (1956) theory to include nuclear burning, assume fuel exists in two states (inside & outside the bubble) and that it burns instantaneously.

❖Validated behavior will be included into sub-grid models for propagation of expanding bubbles in full-star supernova calculations -- *otherwise impossible as initial phenomena is sub-meter size in a 1500 km white dwarf*



3D vorticity rendering: inert thermal

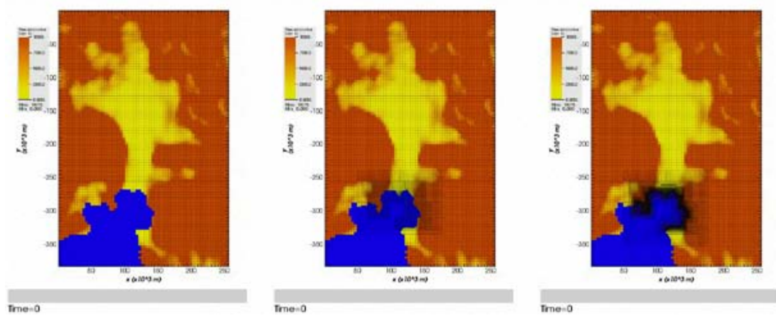


Comparison to theory agrees with remarkable accuracy, allowing inclusion of behavior in sub-grid model for full-star calculations

Collaboration with LLNL: *A. Aspden & J. Bell (LBNL), R. Hoffman (LLNL)*
Code: *MAESTRO* Run time: *~100,000 CPU hrs (each) on ATLAS*

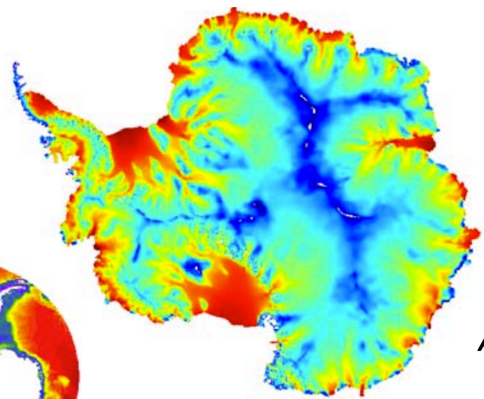
Ice sheet dynamics simulations critical to understanding climate impact on sea level

- Current models inadequate to capture interaction between warming ocean and Antarctic ice sheet
  - Results in incorrect estimates of sea level rise



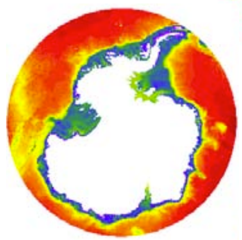
BISICLES Pine Island Glacier simulation – mesh resolution crucial for grounding line behavior.

- LBNL BISICLES ice sheet model uses AMR to resolve essential fine scales at ice-ocean interface.
  - Dynamics dominated by need for sub-km resolution at fronts
  - Leverages US DOE investments in Chombo AMR-PDE infrastructure



Antarctic ice speed (left): AMR enables sub-1 km resolution (black, above)
(Using NERSC's Hopper)

- Collaboration among BISICLES and DOE/BER-sponsored IMPACTS, COSIM will couple ice sheet and ocean models

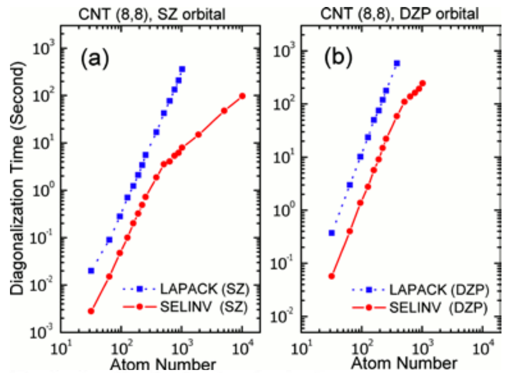


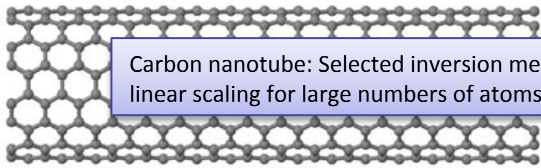
Enhanced POP ocean model solution for coupling to ice (LANL)

Dan Martin, Esmond Ng

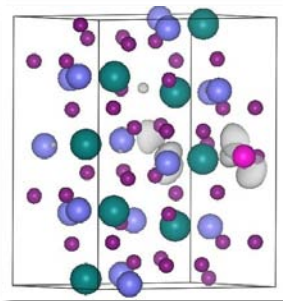
Discovery of new Materials through HPC Modeling & Simulation

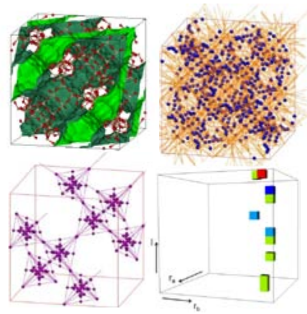
- **New math formulations and fast, reliable algorithms for electronic structure calculations** (Chao Yang)
- **Large Scale First Principles Electronic Structure Calculations to understand materials properties** (Andrew Canning)
  - High throughput approaches to discover new scintillator gamma ray detectors
  - f-electron soft X-ray spectroscopy simulation, theory, and experiment for clean energy materials
  - Large scale eigenvalue calculations in the study of electron excitation for photovoltaic materials
- **High-throughput analysis of materials** (Maciej Haranczyk)
  - Identify optimal materials for applications





Carbon nanotube: Selected inversion method shows linear scaling for large numbers of atoms





New Bright Scintillator Ba2CsI5:Eu

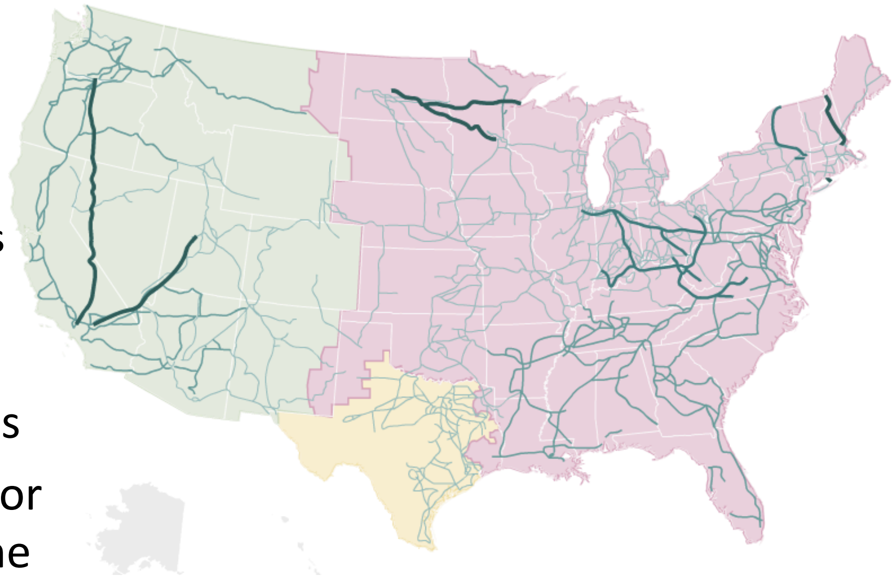
Localization of excited state (gray) on the Eu atom (pink) is one of the necessary theoretical conditions for gamma detection

Porous material (FAU zeolite)

Voronoi tesselation is used to find void space in molecules leading to interesting material properties

Combinatorial Algorithms are used to analyze the Electric Power Grid

- Research develops Optimization and Combinatorial techniques
  - Detect power-grid vulnerabilities
  - Analyze cascading outages
  - Perform resource allocation across multiple locations and times
- Combinatorial/graph algorithms identify critical line failure scenarios
- Identify and rank optimal detours for rerouting power during multiple line failures to facilitate decision makers involving many Control Area Operators



Aydin Buluç, LBNL
Juan Meza, UCM
Ali Pinar, SNL

FastBit helps scientists quickly find patterns in large datasets up to 100x faster

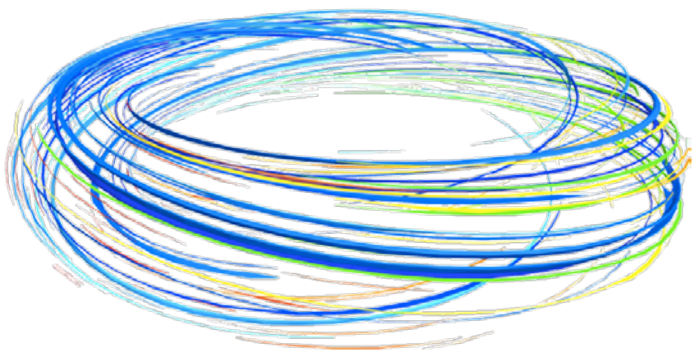
Problem

- Apply indexing methods to accelerate scientific data analysis
- Extend searching capability to find spatially coherent regions in addition to finding individual data records

Impact

- Efficiently find spatial features such as ignition kernels in combustion and hurricanes in climate modeling
- Perform in-situ index generation on hybrid hardware to achieve orders of magnitude speed up on searches for regions

	Time (ms)		Speedup
Coordinates	magnetic	Cartesian	
Query Set I	45.6	44060	**966**
Query Set II	32.0	19016	**549**



Progress and future challenges

- **FastBit is a bitmap search technology, helps users to quickly find patterns of interest in large datasets up to 10-100 times faster than other methods. Received 2008 R&D 100 Award.**
- **FastBit indexing is capable of searching for regions of interest defined on cartesian, AMR, and toroidal meshes.**
- **Example: Speed up searches for regions on toroidal meshes by nearly 1000X on a set of data from GTC**
- **Challenge: design a parallel version of FastBit that scales linearly for in-situ processing on hybrid hardware, minimizing data movement (and energy).**

US DOE has Unique Challenges for Data Intensive Science from their Facilities

Astronomy

Particle Physics

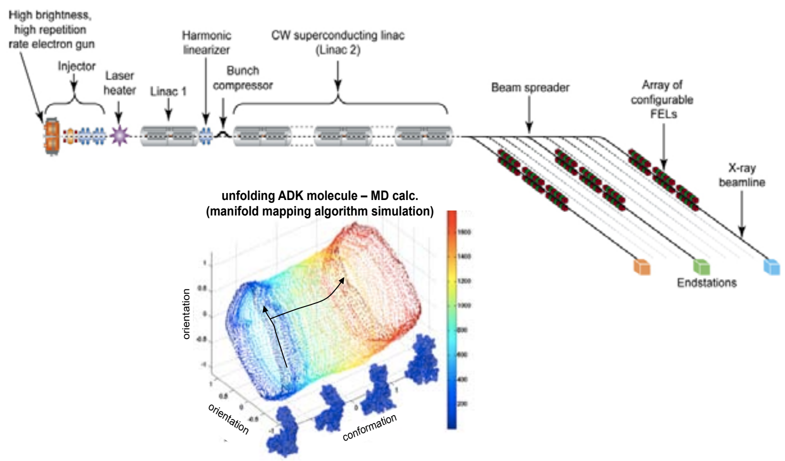
Chemistry and Materials

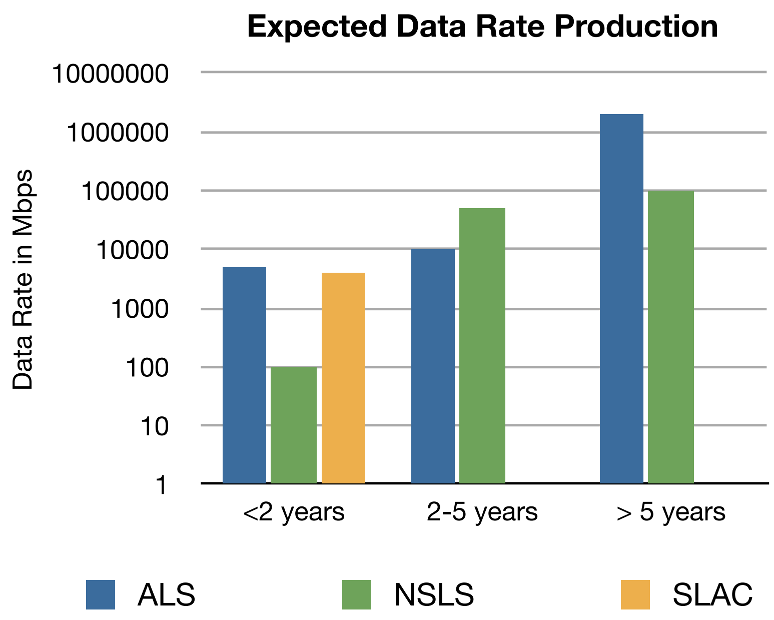
Genomics

Fusion

- Petabyte data sets today, many growing exponentially
- Processing grows super-linearly
- Exascale program addresses some data challenges from modeling and simulation, e.g., in-situ analysis, uncertainty quantification
- Experimental facility data challenges are unique

Data from Light Sources growing faster than Moore's law



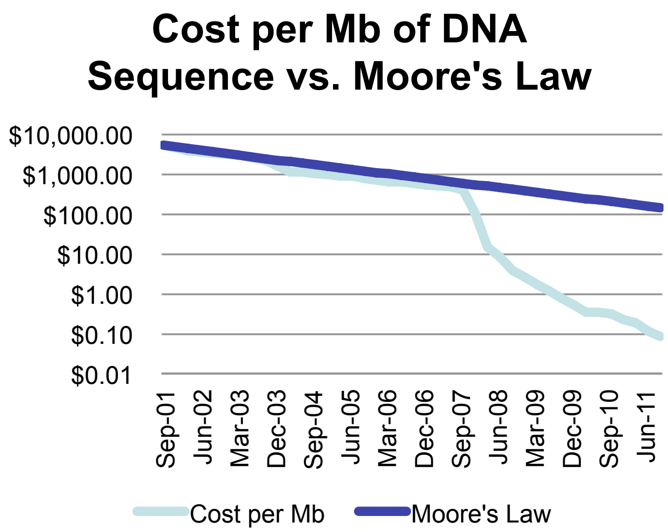


Next Generation Light Source 2025: Orientation reconstruction alone requires an exaflop/s system for pulse rate of 10^5 images/second

- **Data rates to grow 10,000x in next 10 years**
- **Fundamental challenges in algorithms for data processing**
- **Filtering likely required near/on the detectors**

Data is now the big challenge in Biology

- Joint Genome Institute output will increase 1000x over the next 10 years
  - Cost of computation and analysis will dominate
- Automated pipelines required
  - Challenges in algorithms (currently $O(n^2)$ annotation),

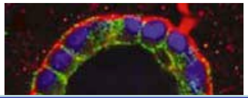
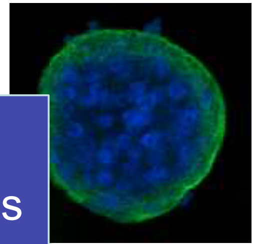
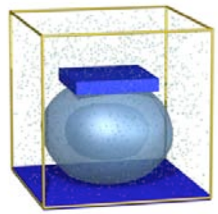
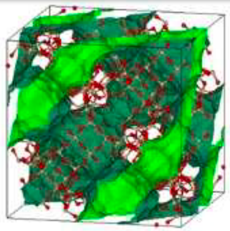
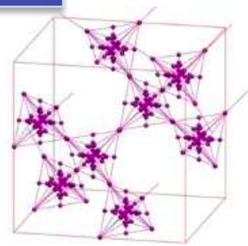
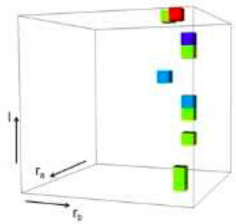


- **Complexity of data will also increase**
  - **Metagenomics to study the response of bacterial communities to natural and experimental changes**
  - **Gene expression data, collected across different environmental conditions, cell-types, etc. across multiple genomes**

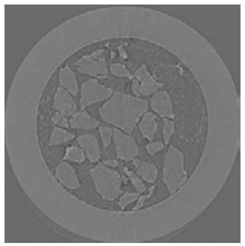
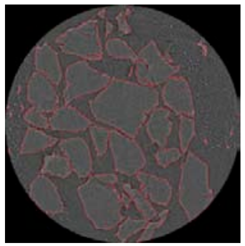
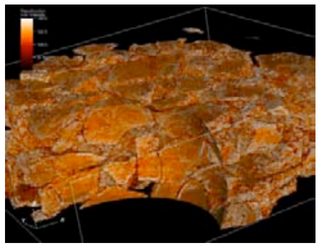
The Math Foundry accelerates the interplay between math advances and scientific discovery

- Identify new areas in science and engineering where mathematics and computation can make a major impact:
  - *Computational Materials and Chemistry*
  - *Imaging and Reconstruction in Beam Science*
  - *Biological modeling*
- Leverage significant algorithmic expertise at the Lab and Campus
  - *Adaptive mesh refinement → Multiphysics*
  - *Statistics → Large databases*
  - *Computational fluids, elasticity, mechanics*
  - *Moving interfaces*

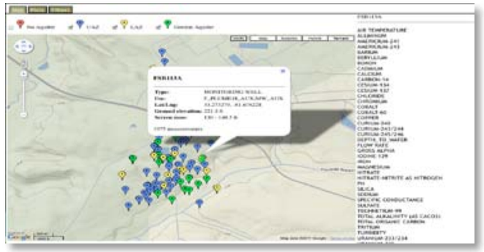
Team mathematicians with scientists

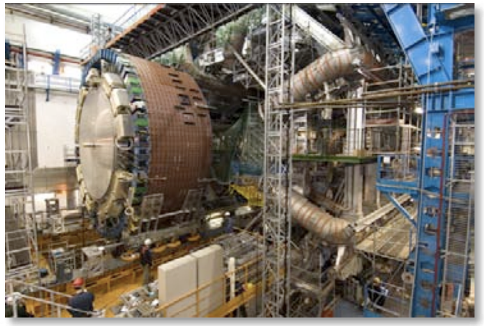
- Management Structure
  - *Director (James Sethian) 0.5 FTE*
  - *Area Leaders 0.5 FTE each*
  - *PDs, GRA partial support*
  - *On-call math staff funded through partnerships*
- We envision a $4M/year Center by 2014
  - $570K FY 2012-14 LDRD Investment
  - $2-3M FY 2013-2017 Expected ASCR Multi-faceted math center FOA
  - $500K FY 2014-2016 NSF funding for GRAs, PDs
  - Leverage other SC science calls

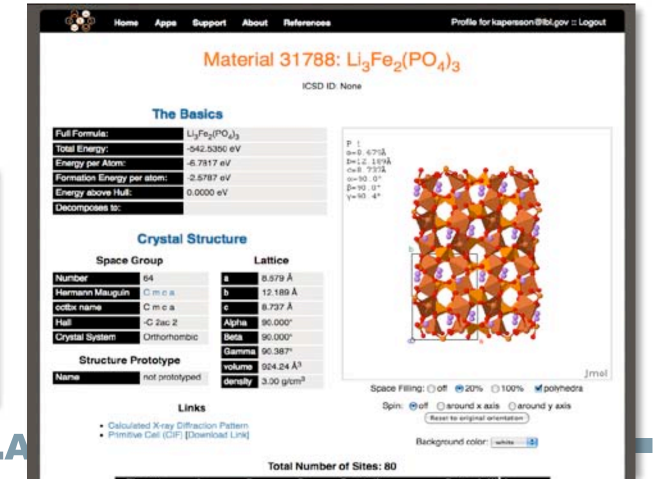




Robust software tools are essential for modern scientific discovery



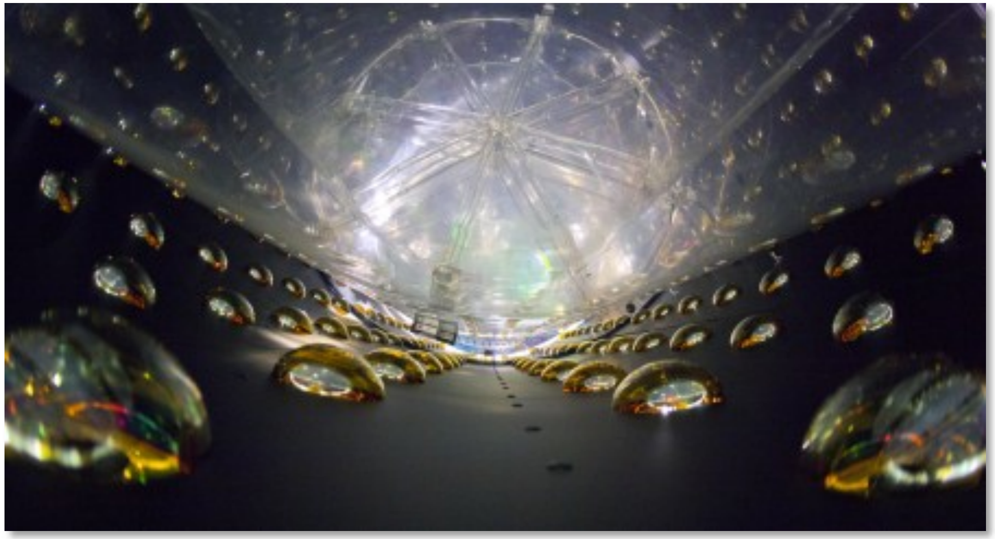
- Software Engineering:
  - Systematic, disciplined, quantifiable approach to development, operation, and maintenance of software
  - Provides scalable, reliable, maintainable software
  - Based on “best practices” and standards to improve inter-operability and long-term maintainability
- CRD has a track record of delivering effective software solutions for DOE science
  - ATLAS: LBNL leads core software platform for LHC at CERN
  - Daya Bay: Analysis tools for anti-neutrino experiment
  - CCSI: Integration framework, data management for carbon capture initiative
  - ASCEM: Data browsing, filtering and download for environmental site cleanup
  - Materials Project: Database and statistical data mining tools

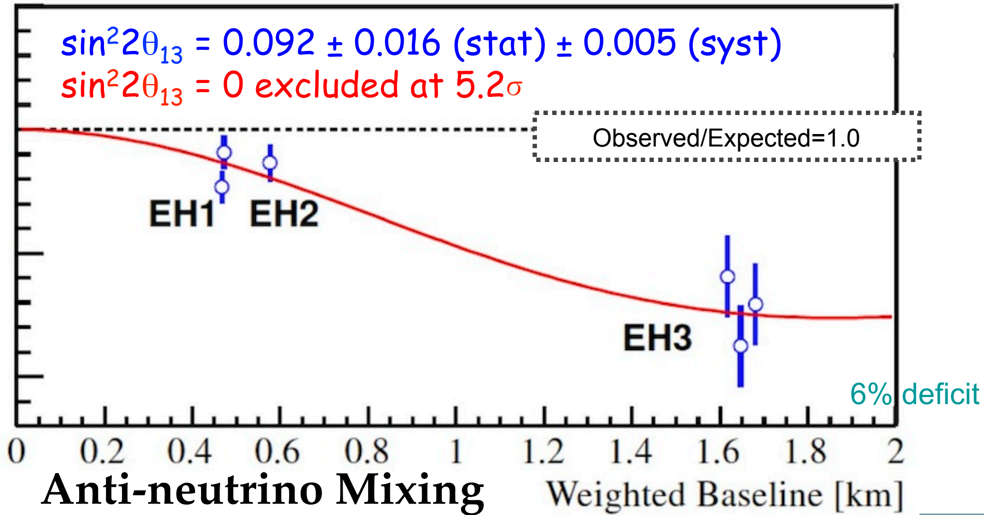




LBNL architected software for Daya Bay anti-neutrino "mixing" experiment

- Anti-neutrino disappearance experiment at nuclear power plant in Daya Bay, China.
- Craig Tull, CRD, Leader of US computing
- Physics Division & CRD experience combined with NERSC & ESNet facilities lead to rapid, high-quality science.
- All data are processed in real time and immediately available to US scientists.
- Antineutrinos seen in analysis within 24 hours.
- First results within record-breaking 75 days after beginning of 3-hall data taking (Mar. 8, 2012)





Anti-neutrino Mixing

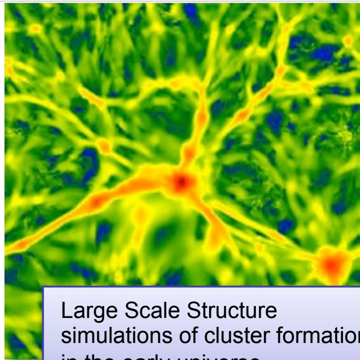
Visit us on the web:

http://crd.lbl.gov

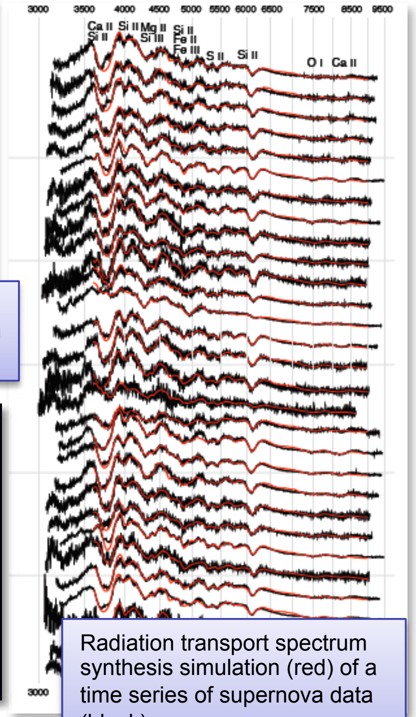
Thank-you

21st Century Astrophysics is data-driven, requires HPC

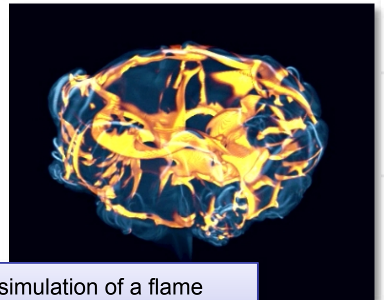
- New technologies to meet cosmological data analysis challenges through tight collaboration between astrophysicists and computational scientists
- Analysis & simulation of 100s of TeraBytes of data from ground- and space-based observations
- Modeling & simulation of supernovae and large-scale structure formation
- Science areas:
    - *Nuclear and atomic physics*
    - *Detector simulations*
    - *General relativity*
    - *Radiation Transport*



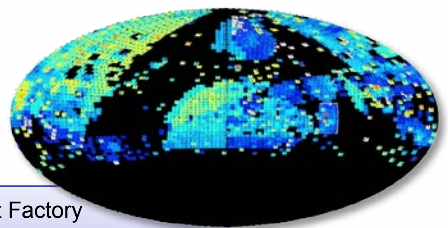
Large Scale Structure simulations of cluster formation in the early universe



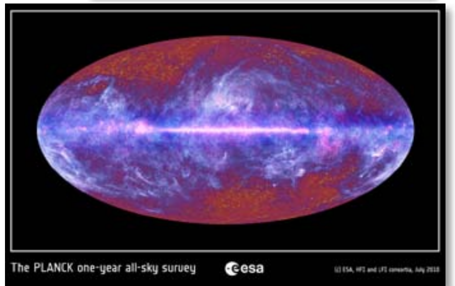
Radiation transport spectrum synthesis simulation (red) of a time series of supernova data (black)



Hydro simulation of a flame front in a thermonuclear supernova explosion

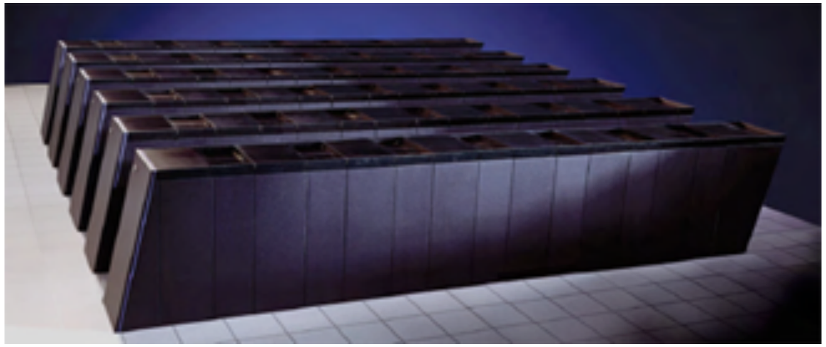


Palomar Transient Factory data-analysis sky-coverage map for the first 3 years of the project



Planck 1st year data analysis

DOE Leadership Class Facilities look to the future of HPC for breakthrough scientific discovery



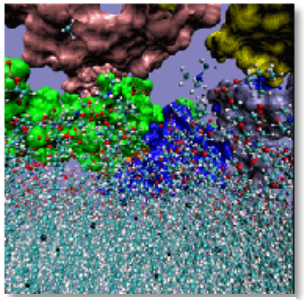
Argonne Leadership Class Facility

- Intrepid (IBM BG/P)
- Surveyor (IBM BG/P)

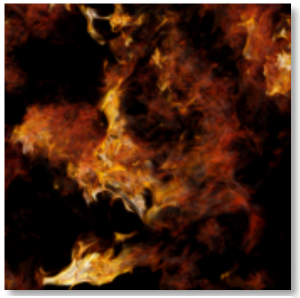
OLCF

Oak Ridge Leadership Class Facility

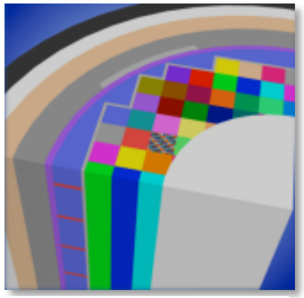
- Jaguar (Cray XT-5/XT-4)
  - *2.3 PF Peak*



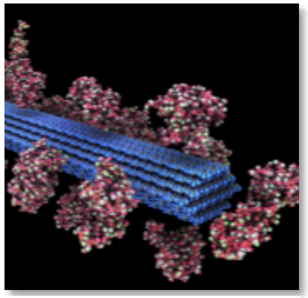
Medicine
Molecular modeling of Parkinson's disease. Research findings at the ALCF offer insight into the disease



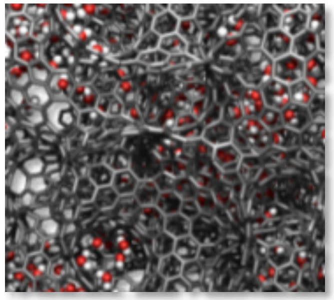
Turbulence
Understanding the statistical geometry of turbulent dispersion of pollutants in the environment.



Nuclear Energy
High-fidelity predictive simulation tools for the design of next-generation nuclear reactors to safely increase operating margins.



Biofuels
A comprehensive simulation model of lignocellulosic biomass to understand the bottleneck to sustainable and economical ethanol production.



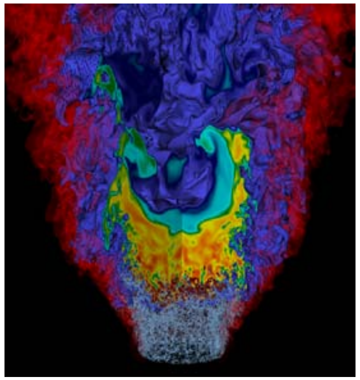
Energy Storage
Understanding the storage and flow of energy in next-generation nanostructured carbon tube supercapacitors

Combining theory, experiments and simulation leads to new scientific insights

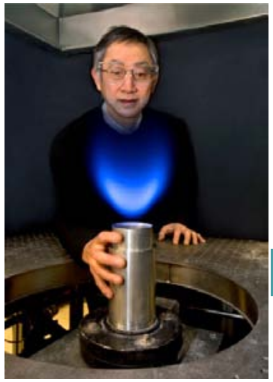
COMPUTATIONAL RESEARCH DIVISION

- *Example:* Low Swirl Burners used by Solar Turbines (Caterpillar) and Maxon Corp. (Honeywell) to improve commercial burners
  - Efficient, low-emissions, Fuel-flexible (oil, gas, hydrogen-rich fuels)
- Simulations explain combustion process resulting in improved designs
  - Modeled kinetics and chemical transport (15 species, 58 reactions)
  - Uses advanced math algorithms (AMR) equivalent to $4K^3$ mesh
  - Scales and runs in production at 20K cores

Simulations show cellular burning in lean hydrogen leads to pockets of enhanced emissions, & increasing the turbulence enhances the effect.



Simulations reveal features not visible in lab (John Bell, PI, LBNL)

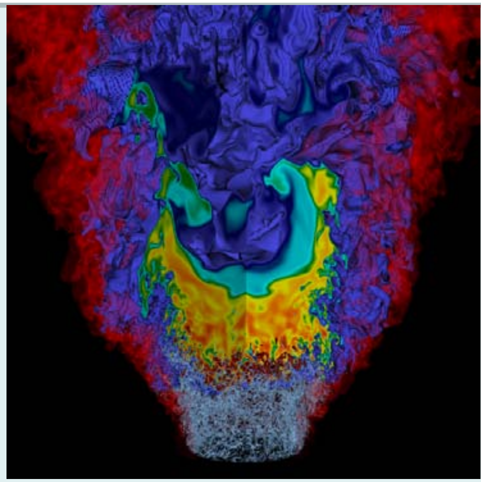


Experiments show feasibility: 50KW-50MW (Robert Cheng, PI, LBNL)

Low NOx technology licensed by industry

High-fidelity simulations key to understand fuel-flexible low-emission combustion

- *Features of next generation combustion systems:*
  - Low emissions, high pressure
  - Able to burn a variety of fuels ("fuel-flexible")
  - Operate in fundamentally different combustion regimes
- *High-fidelity computational tools required to design these systems*
  - Leverage multiscale formulation, exploiting low Mach number mathematical structure
  - Use advanced numerical methodology (AMR+high-order methods)
  - Formulation + advanced numerics results in > 200x speedup over traditional non-multiscale approaches
- *Hybrid programming model, anticipating manycore hardware*
  - Coarse-grained parallelization with MPI
  - Fine-grained parallelization with threads
  - Scales to 50K cores on Jaguarpf
  - Have identified promising reduced communication approaches for elliptic equations to extend to 100s of cores per processor



Next-generation stationary turbines may benefit from low swirl injectors:

•Simulation of NOx formation in a low swirl injector using a low Mach number formulation combined with AMR.

• Resolve internal flame structure in a laboratory-scale simulation

Multiscale formulation + advanced numerics + multicore algorithms = unprecedented ability to simulate high pressure flames