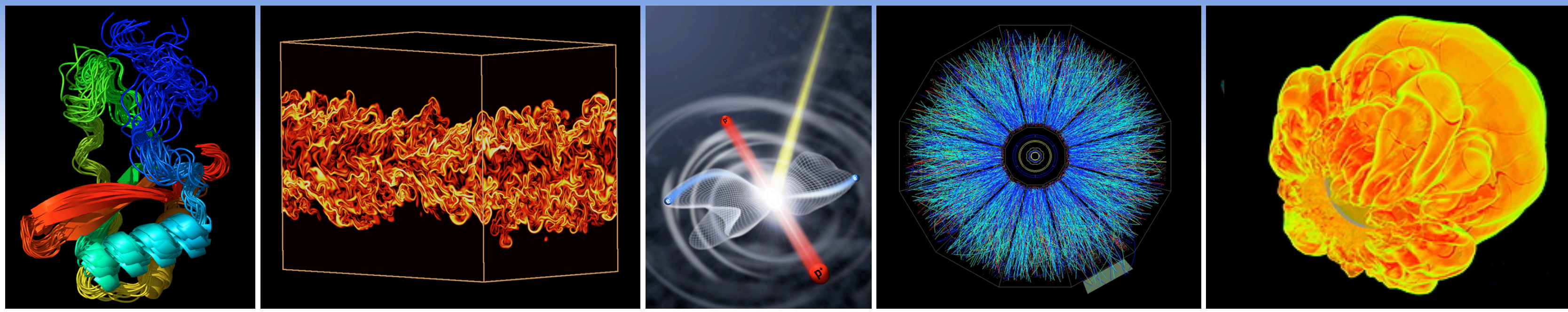


# NERSC Accomplishments and Plans

Katherine Yelick
NERSC Director
Lawrence Berkeley National Laboratory

# NERSC is the Primary Computing Facility for the Office of Science

- NERSC serves a large population

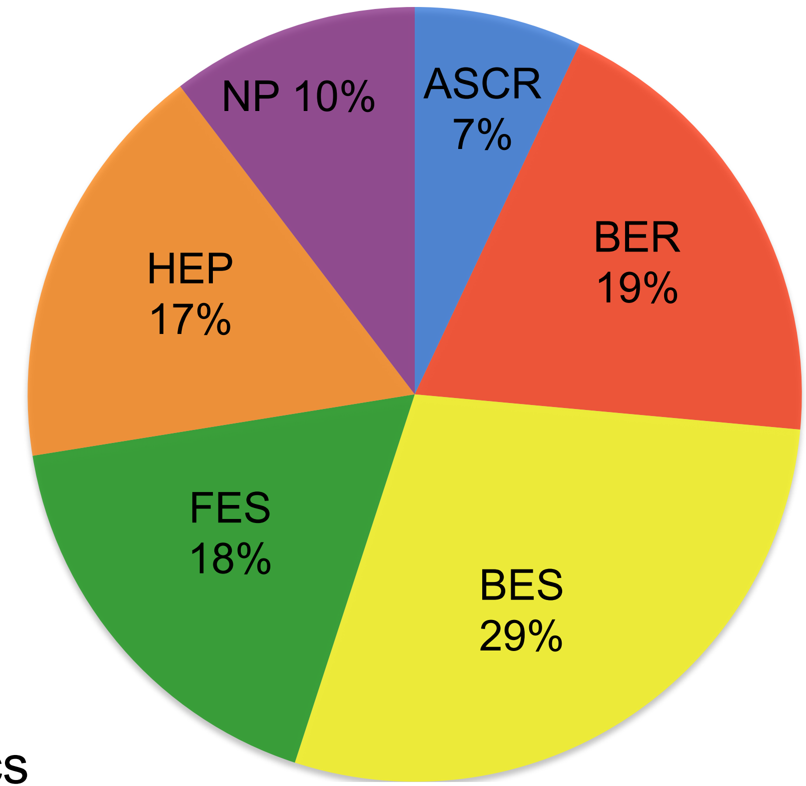
  Approximately 3000 users,

  400 projects, 500 code instances

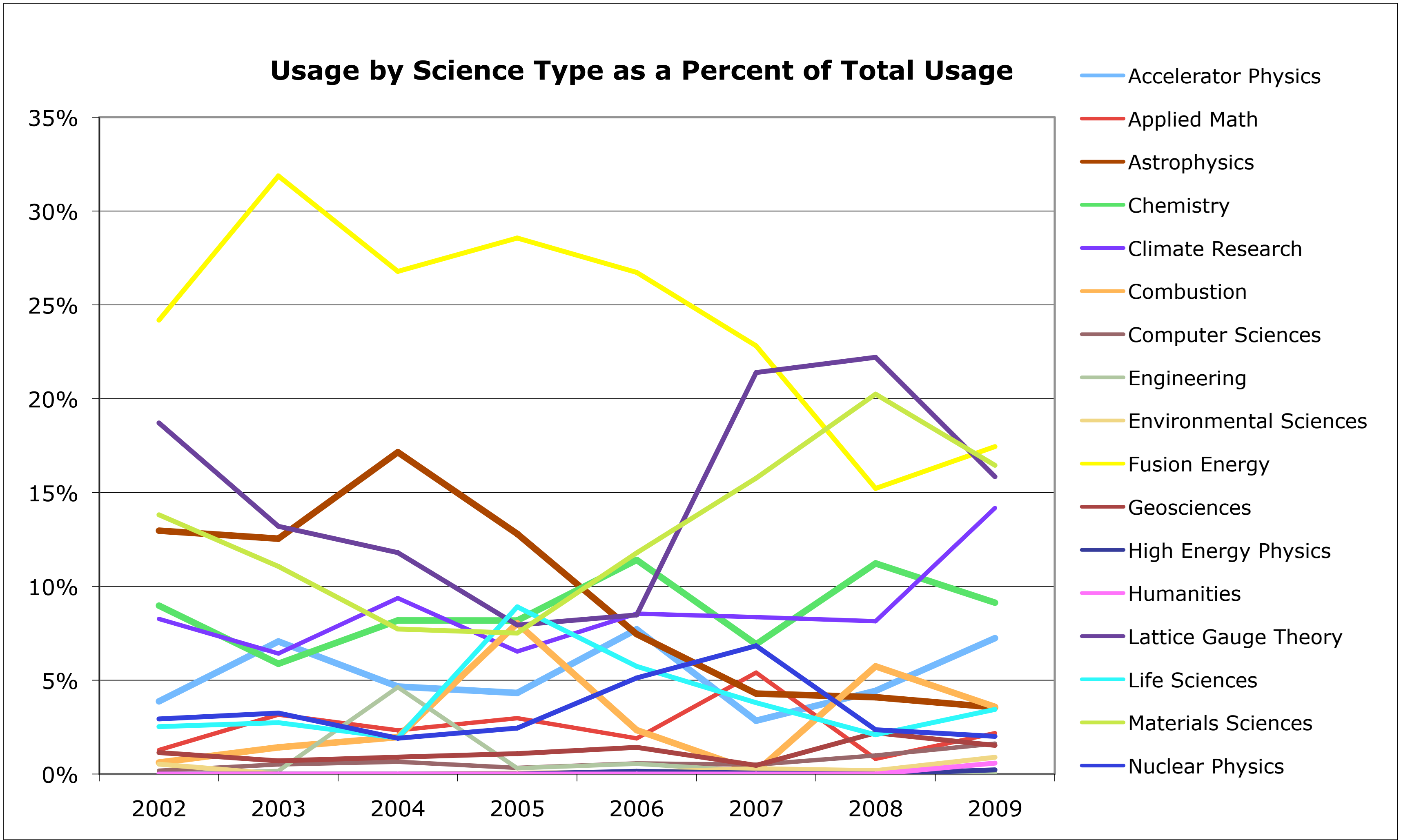
- Focus on "unique" resources
  - High end computing systems
  - High end storage systems
    - File system and tape archive
  - Interface to high speed networking

- Science-driven
  - Science problems used in machine procurements and performance metrics
  - Science services

**2009 Allocations**



# Workload Changes Over Time with DOE Priorities



# ASCR's Computing Facilities

## NERSC at LBNL

- 1000+ users, 100+ projects
- Allocations:
  - 80% DOE program manager control
  - 10% ASCR Leadership Computing Challenge*
  - 10% NERSC reserve
- Science includes all of DOE Office of Science
- Machines procured competitively

## LCFs at ORNL and ANL

- 100+ users 10+ projects
- Allocations:
  - 80% ANL/ORNL managed INCITE process
  - 10% ACSR Leadership Computing Challenge*
  - 10% LCF reserve
- Science limited to largest scale; no limit to DOE/SC
- Machines procured through partnerships

# NERSC 2009 Configuration

## Large-Scale Computing System

**Franklin (NERSC-5): Cray XT4**
- 9,532 compute nodes; 38,128 cores
- ~25 Tflop/s on applications; 356 Tflop/s peak

**Hopper (NERSC-6): Cray XT**
- Phase 1: Cray XT5, 668 nodes, 5344 cores
- Phase 2: > 1 Pflop/s peak

## Clusters

**Jacquard and Bassi**
- LNXI and IBM clusters
- Upgrading to Carver (NCS-c)

**PDSF (HEP/NP)**
- Linux cluster (~1K cores)

## NERSC Global Filesystem (NGF)

Uses IBM's GPFS

440 TB; 5.5 GB/s

## HPSS Archival Storage

- 59 PB capacity
- 11 Tape libraries
- 140 TB disk cache

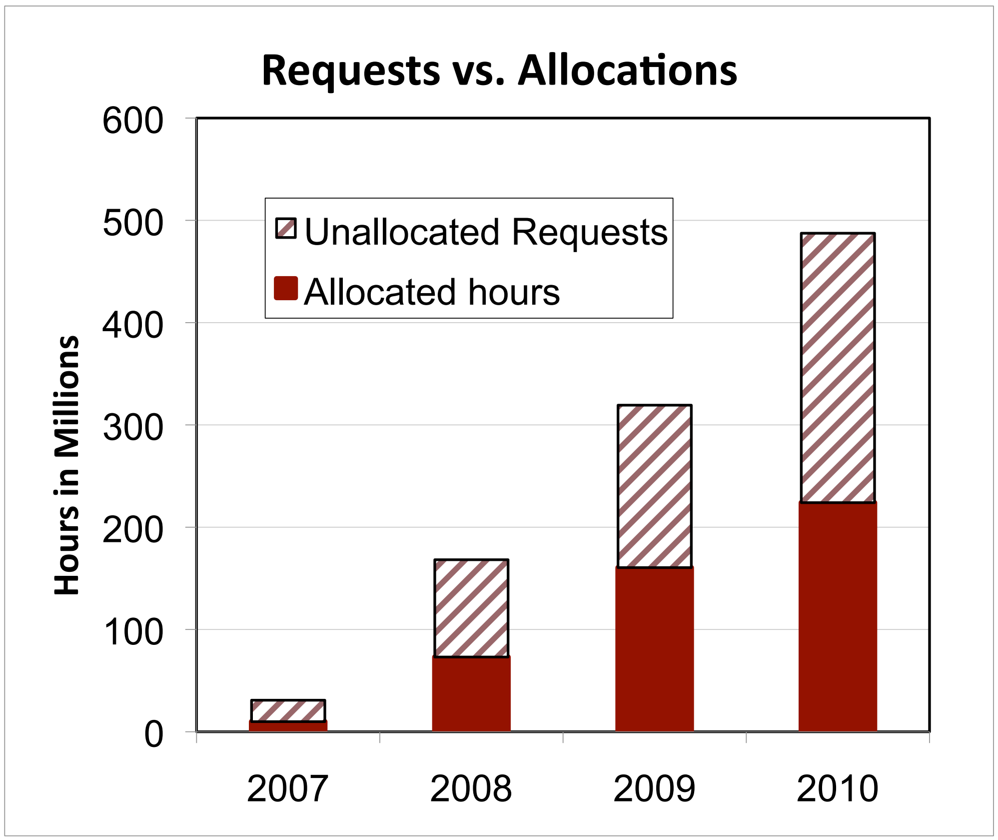
## Analytics / Visualization

**Davinci (SGI Altix)**

- Tesla testbed
- Upgrade planned

# Demand for More Computing

*Compute Hours Requested vs Allocated*

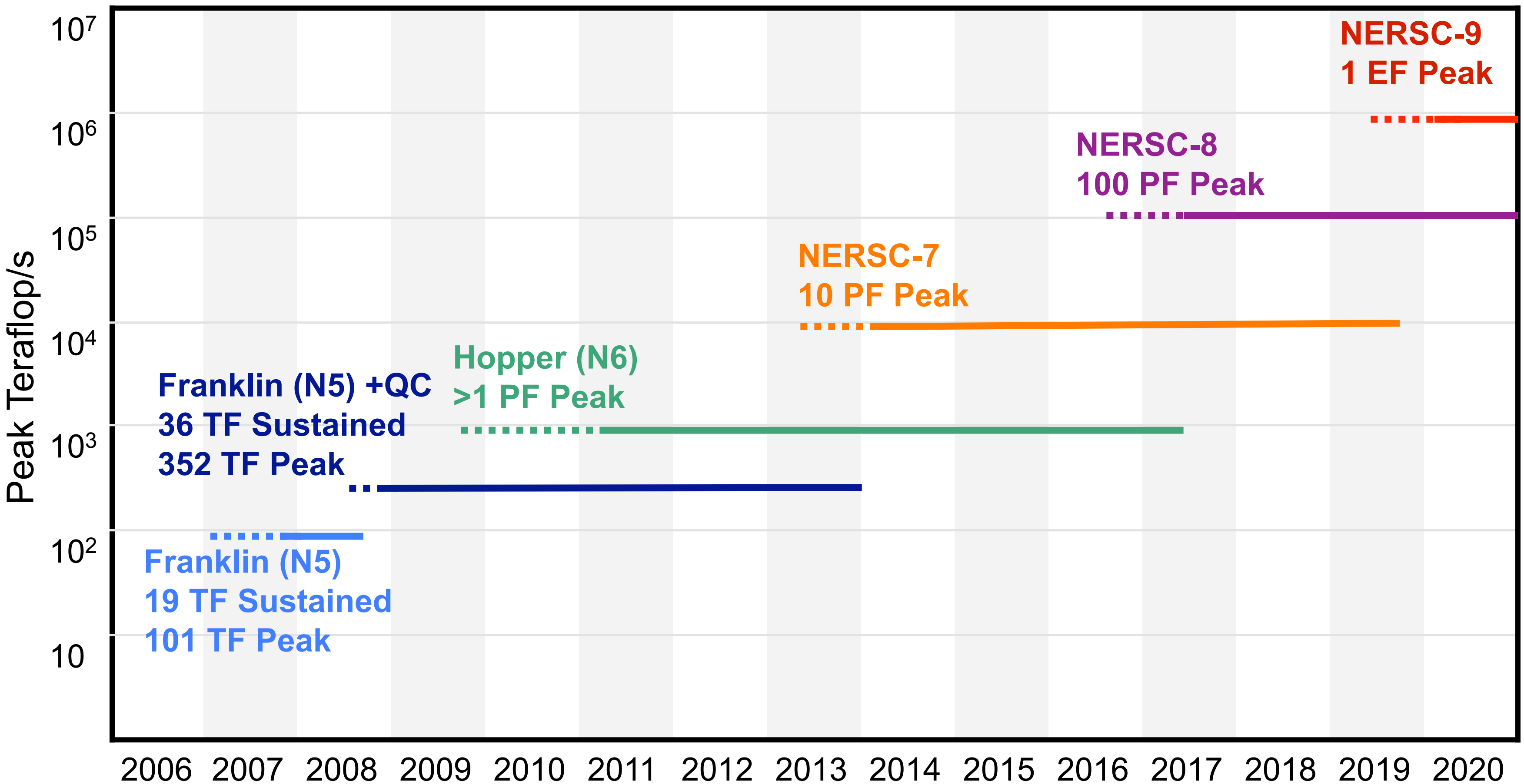



- ***Each year DOE users requests ~2x as many hours as can be allocated***
- ***This 2x is artificially constrained by perceived availability***
- ***Unfulfilled allocation requests amount to hundreds of millions of compute hours in 2010***

# NERSC Initiative for Scientific Exploration (NISE)

- For remainder of AY 2009, 10M hours available for
  - *New research problems* not covered by existing ERCAP allocation, especially high risk/high impact science
  - *New programming techniques* that take advantage of multicore compute nodes
  - *Code scaling* to higher concurrencies for codes that scale on projects limited by current allocation

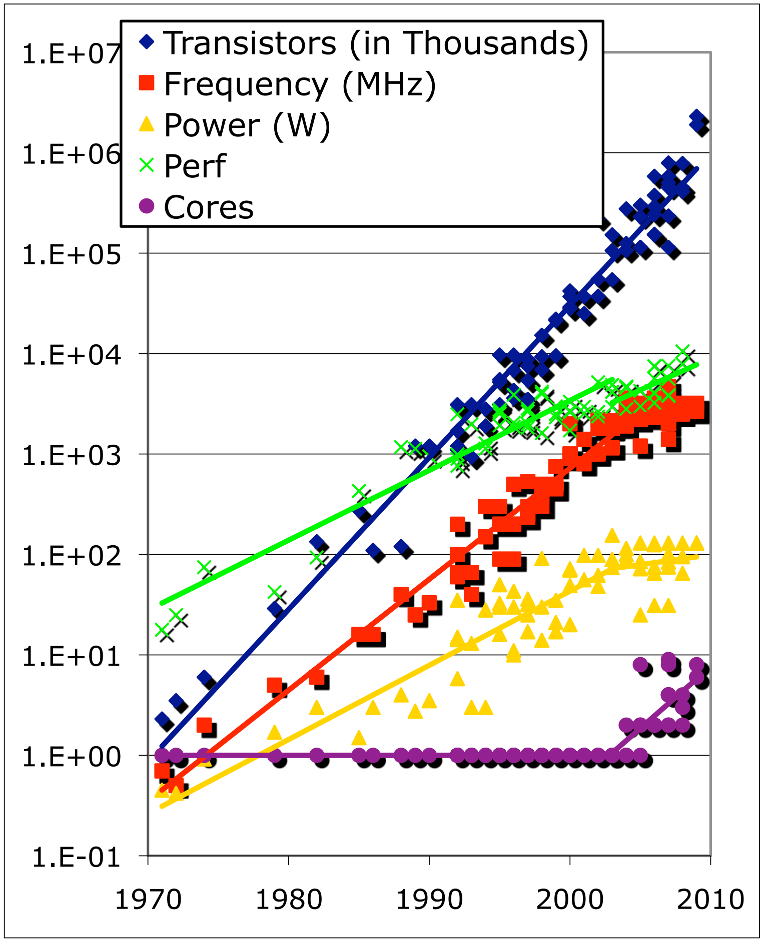
# NERSC System Roadmap




- Goal is two systems on the floor at all times
- Systems procured by sustained performance (10% of peak?)

# Center of Excellence with Cray

- NERSC/Cray "Programming Models Center of Excellence" combines:
  - Berkeley Lab strength in advanced programming models, multicore tuning, and application benchmarking
  - Cray strength in advanced programming models, optimizing compilers, and benchmarking
- Immediate question:
  - Best way to use cores in N6 node
  - MPI, OpenMP, UPC/CAF, Pthreads,…
- Long term necessity for exascale:
  - Massive on-chip concurrency necessary for reasonable power use
  - 3M for 1PF today → 3 GW for 1 EF (or 10 100PF) tomorrow?




Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanoviç

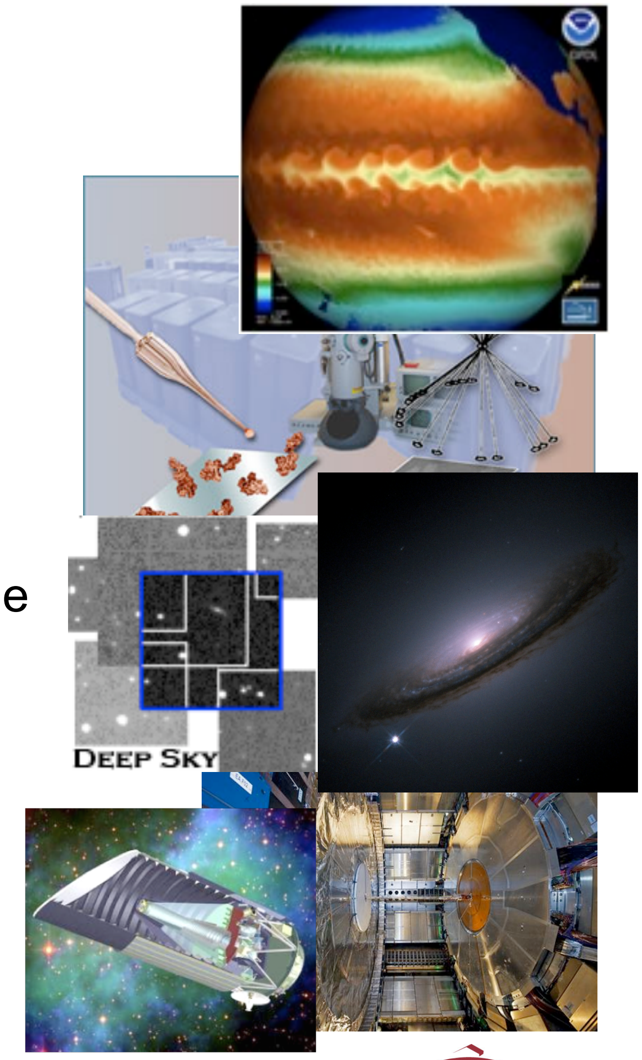
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

# DOE Explores Cloud Computing

- ASCR Magellan Project
  - $32M project at NERSC and ALCF
  - ~100 TF/s compute cloud testbed (across sites)
  - Petabyte-scale storage cloud testbed
- Cloud questions to explore on Magellan:
  - Can a cloud serve DOE's mid-range computing needs?
    → More efficient than cluster-per-PI model
  - What part of the workload can be served on a cloud?
  - What features (hardware and software) are needed of a "Science Cloud"? (Eucalyptus at ALCF; Linux at NERSC)
  - How does this differ, if at all, from commercial clouds?

# Data Driven Science

- Ability to generate data is exceeding our ability to store and analyze it
  - Simulation systems and some observational devices grow in capability with Moore's Law
- Opportunity to lead creation of scientific communities around data sets
- A *science gateway* is a set of hardware and software for remote data/services
  - Deep Sky – "Google-Maps" of astronomical image data: 36 supernovae in 6 nights
- Petabyte data sets will be common:
  - *Climate modeling:* IPCC will be 10s of petabytes
  - *Genome:* Genomes will double each year
  - *Particle physics*: LHC is projected to produce 16 petabytes of data per year

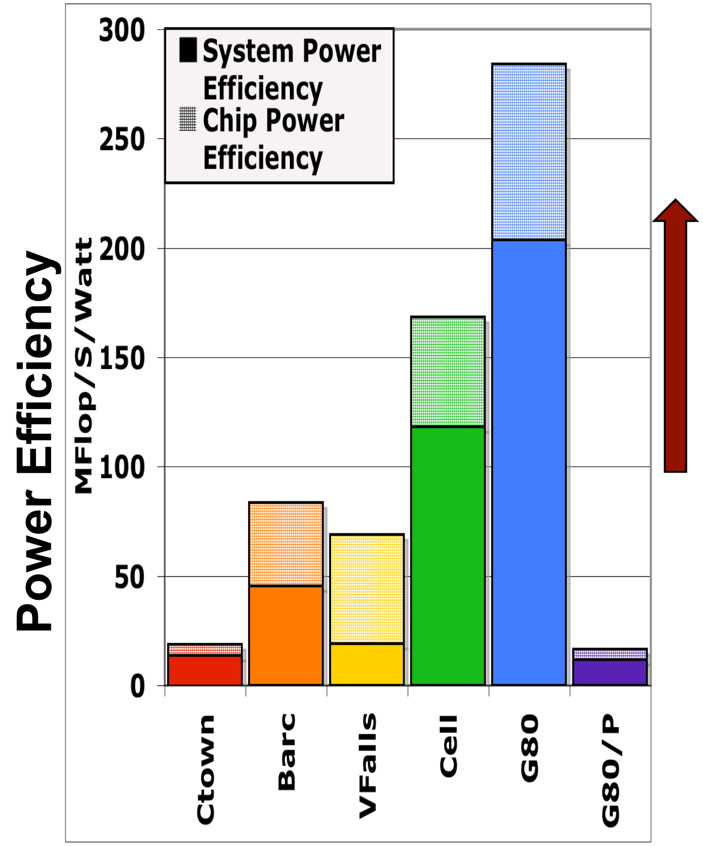


# Tesla/Turing GPU Testbed

- 2-node testbed with shared-memory GPU architecture on each node
- Goal 1: application experience
  - Can science computation use GPUs?
- Goal 2: administration experience
  - Batch queues and GPUs (GPU/CUDA, OpenGL/vis)
- Goal 3: visualization experience
  - Remote delivery of hardware-accelerated graphics/vis
- Goal 4: large memory workload

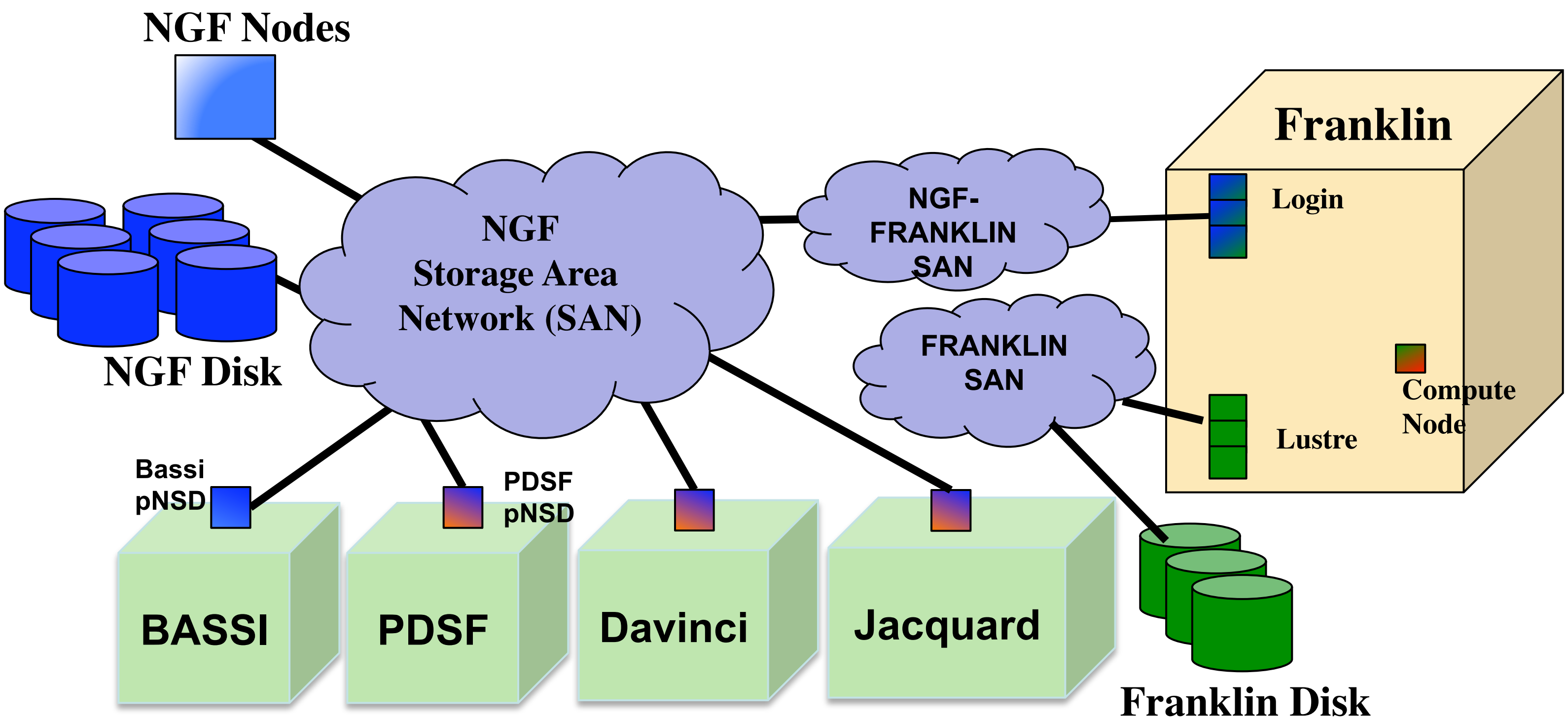
– 256 GB of shared memory

- Note: testbed, not production machine!




*Mflops / Watt of 3D Stencil*

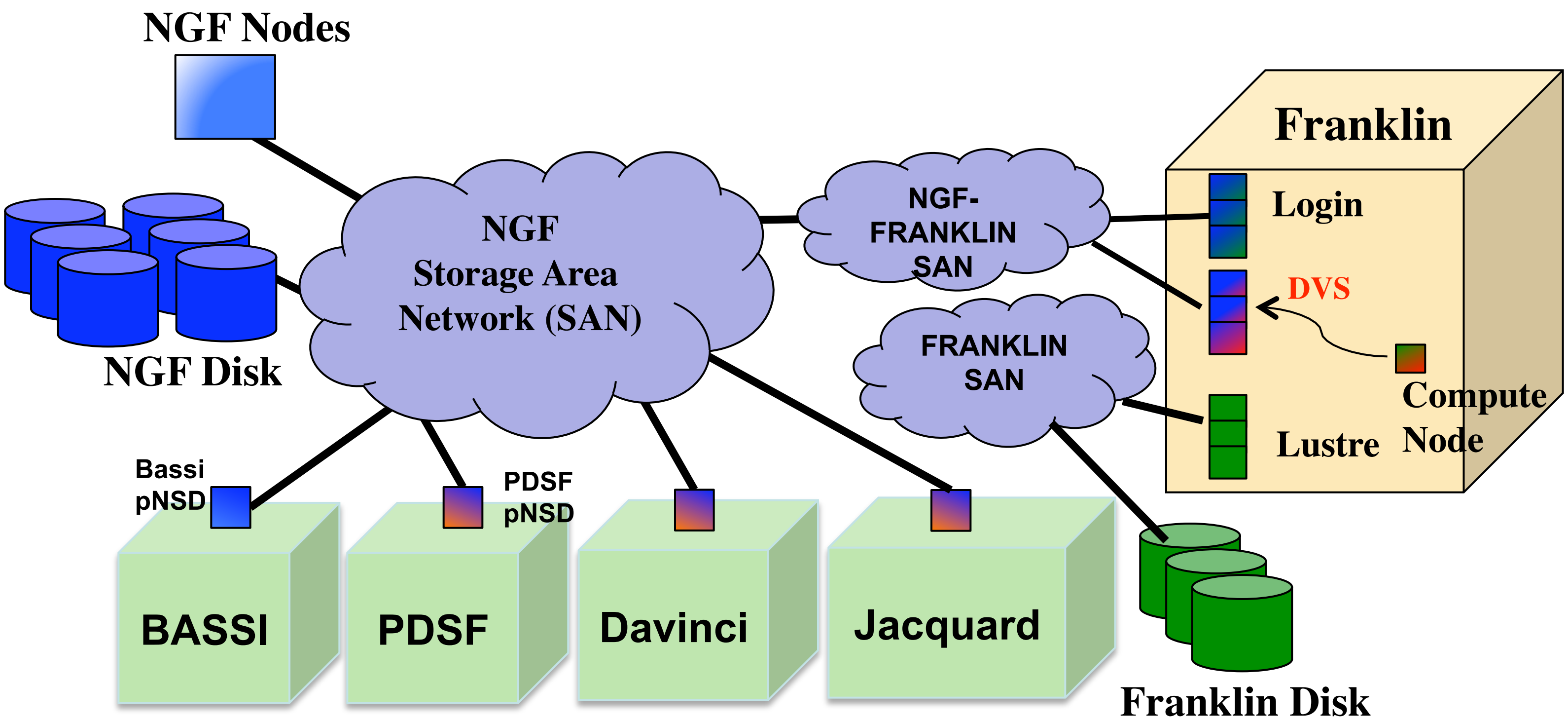
# NERSC Global File system (NGF)




- **A facility-wide, high performance, parallel file system**
  - Uses IBM's GPFS technology for scalable high performance
  - The /project file system in NGF from all NERSC systems
  - Intended for data that is shared across machines or users in a project

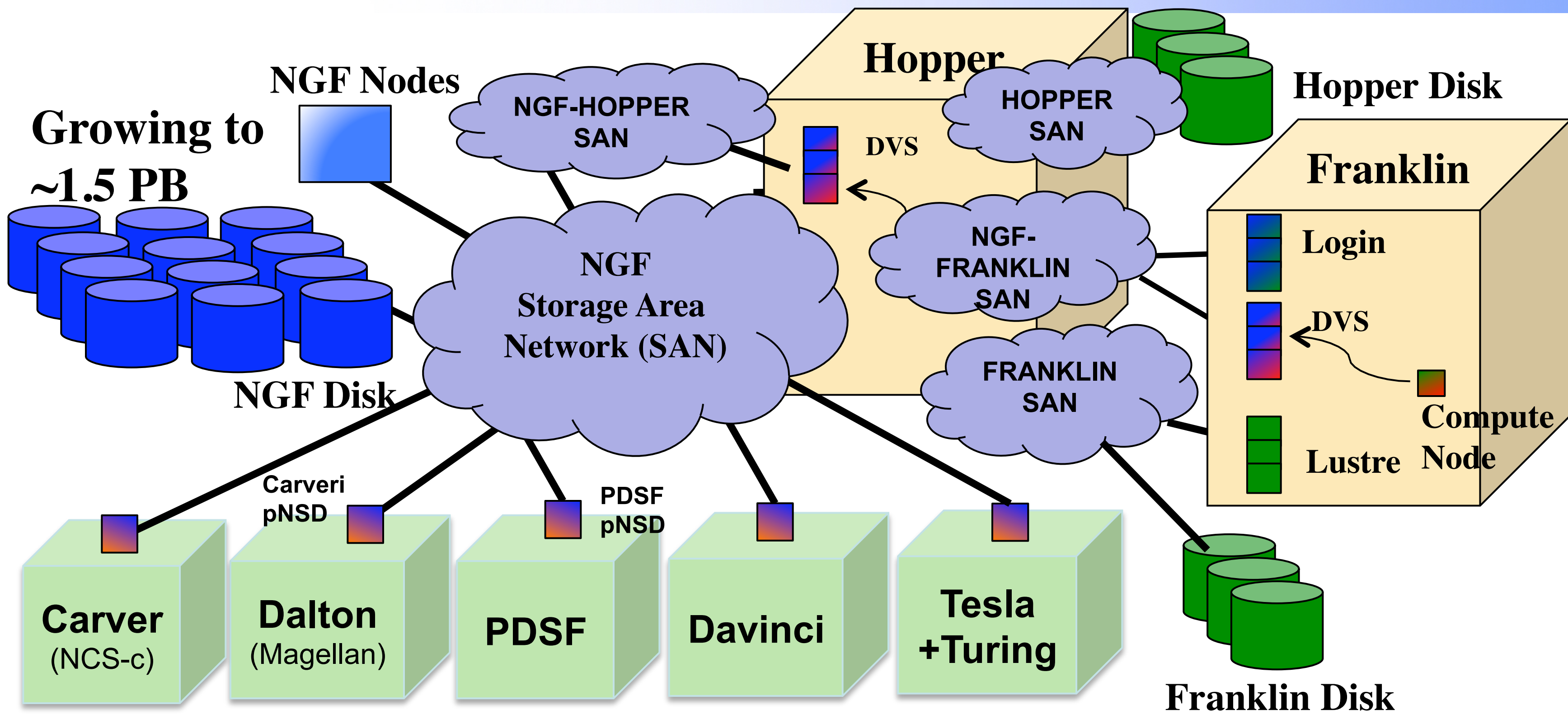
See: http://www.nersc.gov/nusers/services/proj.php

# NERSC Global File system (NGF)




- **Announcing access to NGF from Franklin compute nodes**
  - Effective immediately /project is available on Franklin compute nodes
  - Uses Cray DVS (Data Virtualization Services) software
  - Expect ~4GB/s from /project vs. ~10GB/s from /scratch or /scratch2

# NERSC Global File system (NGF)




- **Coming soon to NGF**
  - Additional storage, up to ~1.5 PB total
  - Access to NGF from new systems: Carver (replacing Jacquard and Bassi); Dalton (the Magellan testbed); Tesla & Turing (GPU testbed)

# HPSS at NERSC

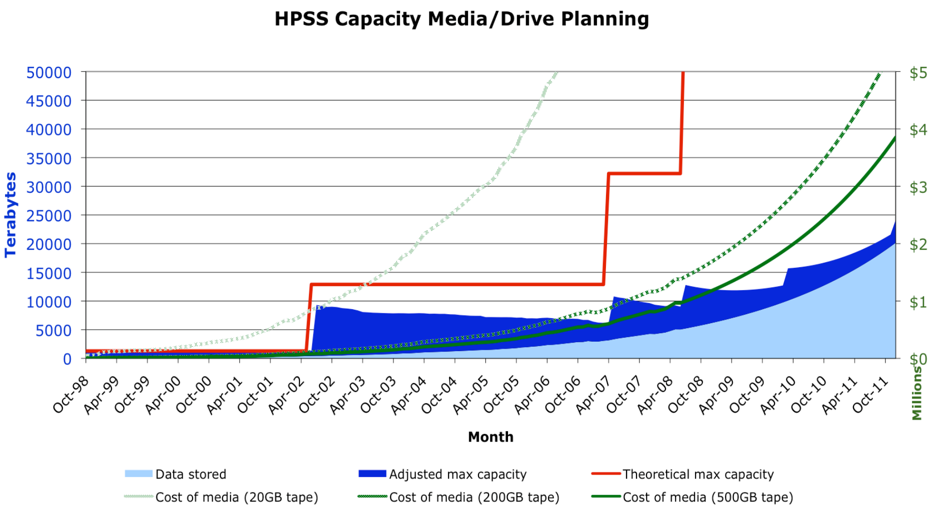
**NERSC has been archiving data with HPSS since 1998**

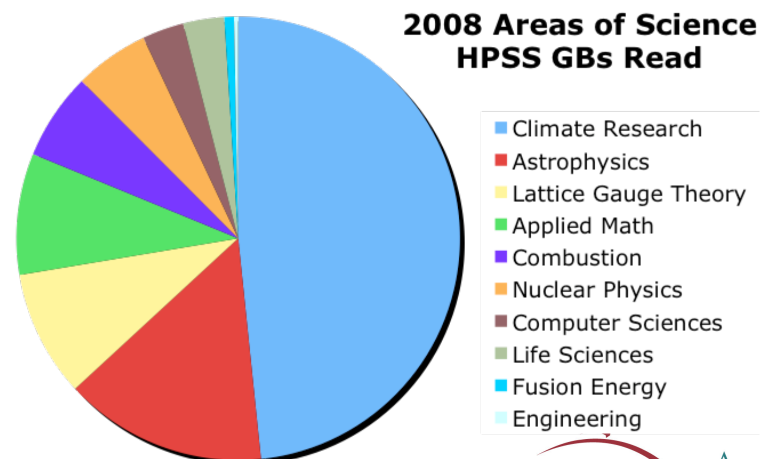
- **The total data volume increases by ~50% annually**

**NERSC has two HPSS systems:**

- **An Archive system that stores user files optimized for high-transfer rates; about 66M files in 2009**
- **A Backup system for NGF; about 12M files in 2009**

**HPSS averages 100 MB/s, with peaks to 450 MB/s**




U.S. DEPARTMENT OF ENERGY | Office of Science

# HPSS Upgrades and Plans

- Increased bandwidth
  - Franklin increased load on HPSS by 50%
  - New movers and servers; new clients on all NERSC systems
- Increased capacity through new hardware / tapes
  - 3 new storage libraries in past 2 years; 1 more in 2010
  - Currently have max capacity of 59 PB if filled with 1 TB tapes
  - 1 ½ year repack (40K tapes onto 10K 1 TB tapes) underway
- Ease of use improvements
  - Upgraded software to HPSS version 6.2
  - Integrated HPSS into NIM for account/password management
  - Improved MTBI from ~5 days in 2008 to ~9 days 2009.
- Evaluating new clients for bandwidth and functionality
  - rsynch, conditional stores, and dynamic file aggregation

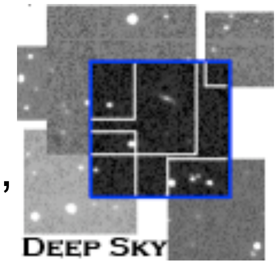
# Services for Science

# Reservations at NERSC

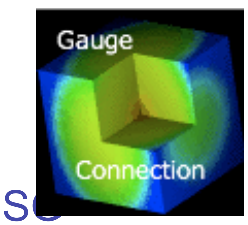
- Reservation service being tested:
  - Reserve a certain date, time and duration
    - Debugging at scale
    - Real-time constraints in which need to analyze data before next run, e.g., daily target selection telescopes or genome sequencing pipelin
  - At least 24 hours advanced notice
    - https://www.nersc.gov/nusers/services/reservation.php
  - Successfully used for IMG run, Madcap, IO benchmarking, etc.

# Science Gateways at NERSC

- Create scientific communities around data sets
  - Models for sharing vs. privacy differ across communities
  - Accessible by broad community for exploration, scientific discovery, and validation of results
  - Value of data also varies: observations may be irreplaceable




- A *science gateway* is a set of hardware and software that provides data/services remotely
  - Deep Sky – "Google-Maps" of astronomical image data
    - Discovered 140 supernovae in 60 nights (July-August 2009)
    - 1 of 15 international collaborators were accessing NGF data through the S nodes 24/7 using both the web interface and the database.
  - Gauge Connection – Access QCD Lattice data sets
  - Planck Portal – Access to Planck Data




- Building blocks for science on the web
  - Remote data analysis, databases, job submission

# Visualization Support

***Petascale visualization*: Demonstrate visualization scaling to unprecedented concurrency levels by ingesting and processing unprecedentedly large datasets.**

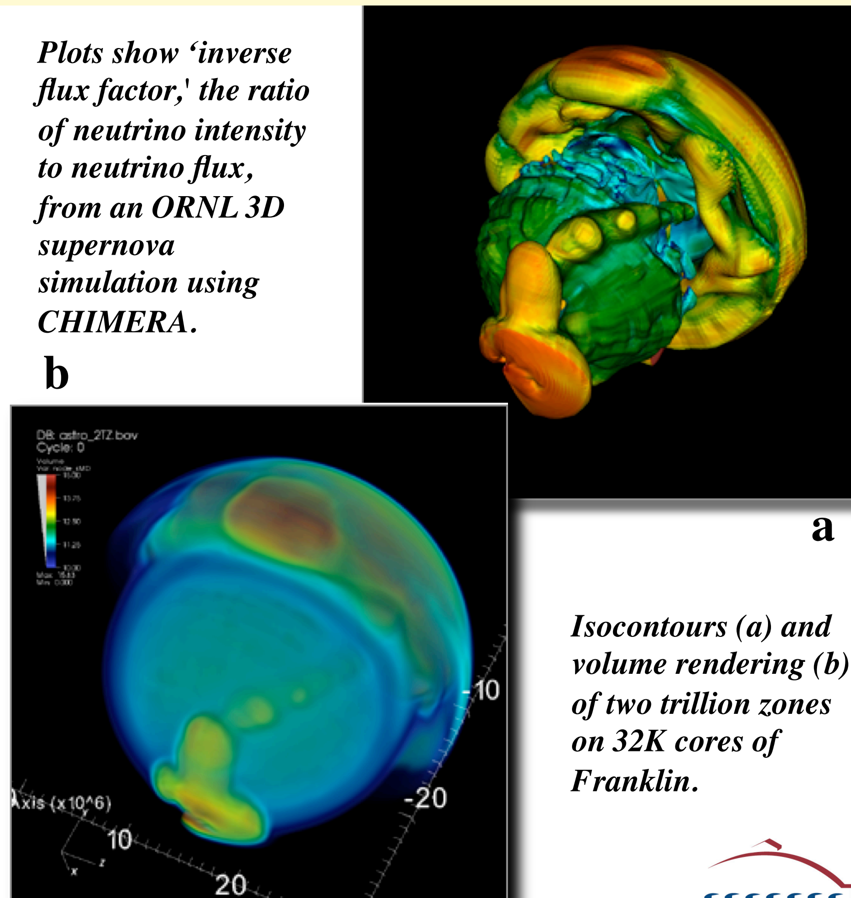
***Implications*: Visualization and analysis of Petascale datasets requires the I/O, memory, compute, and interconnect speeds of Petascale systems.**

***Accomplishments:* Ran VisIt SW on 16K and 32K cores of Franklin.**

- **First-ever visualization of two *trillion* zone problem (TBs per scalar); data loaded in parallel.**
- **Petascale visualization**

*Plots show 'inverse flux factor,' the ratio of neutrino intensity to neutrino flux, from an ORNL 3D supernova simulation using CHIMERA.*

**b**




**a**

*Isocontours (a) and volume rendering (b) of two trillion zones on 32K cores of Franklin.*

# HEP: Accelerator Modeling

***Objective***: Use INCITE resources to help design and optimize the electron beam for LBNL next-generation Free Electron Laser.

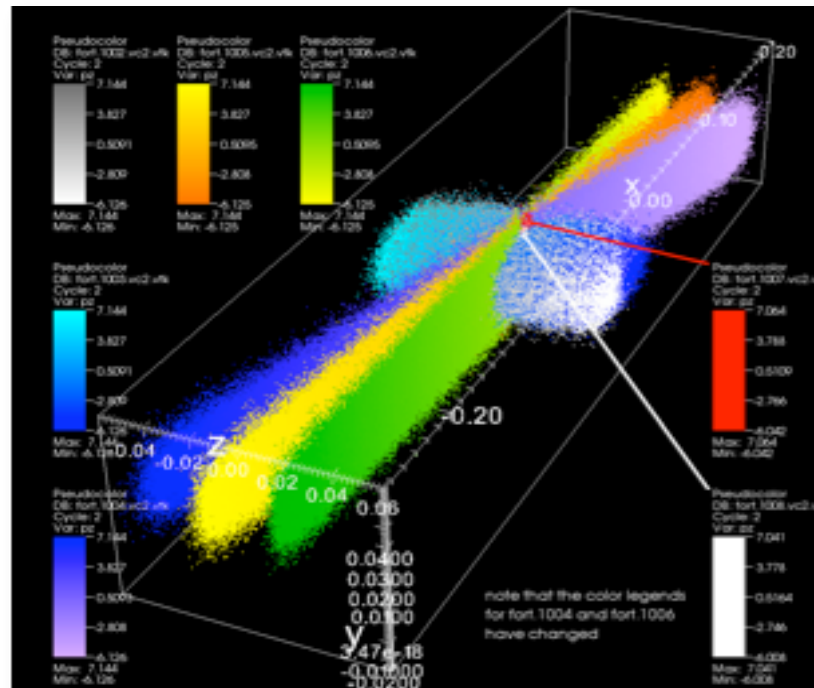
***Implications***: Numerically optimizing the beam lowers cost of design / operation and improves X-ray output, helping scientific discovery in physics, material science, chemistry and bioscience.

***Accomplishments:*** Code includes self-consistent 3D space-charge effects, short-range geometry & longitudinal synchrotron radiation wakefields, and detailed RF acceleration / focusing.

- ***Billion***-particle simulation required for details of high brightness electron beams subject to microbunching instability.
- Key NERSC visualization support.

***NERSC:***
- 400k hours used in 2009 (~50% of allocation).
- Uses IMPACT code, part of NERSC6 test suite.

## PI: J. Qiang (LBNL)




*Visualization of an electron beam bending and changing orientation as it passes through a magnetic bunch compressor.*

Proc. Linac08 Conference

# Cloud-Resolving Climate Model

**PI: D. Randall, Colo. St**

***Objective***: Climate models that fully resolve key convective processes in clouds; ultimate goal is 1-km resolution.

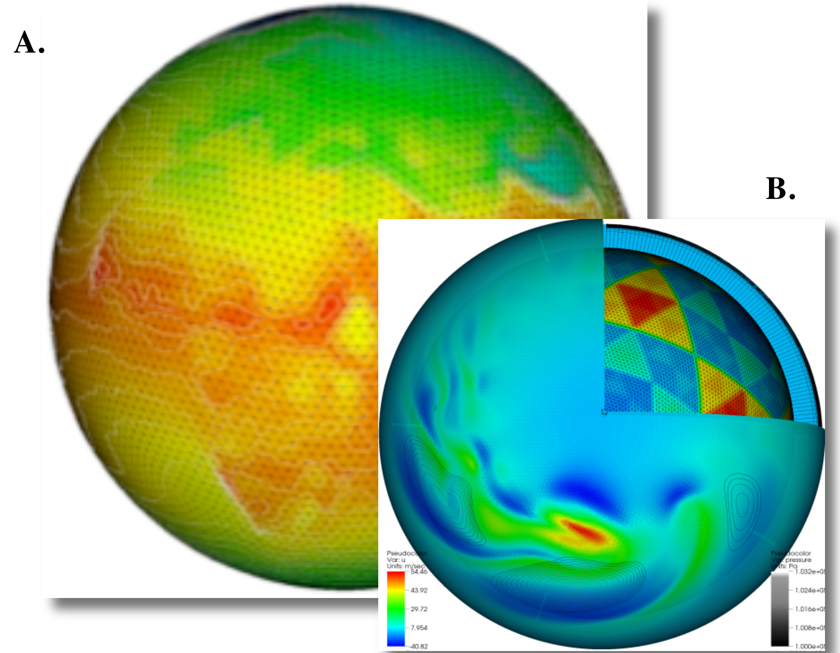
***Implications***: Major transformation in climate/weather prediction, likely to be standard soon, just barely feasible now.

***Accomplishments:*** Developed a coupled atmosphere-ocean-land model based on geodesic grids.

- Multigrid solver scales perfectly on 20k cores of Franklin using grid with 167M elements.
- Invited lecture at SC09.

***NERSC:***

- 2M hour allocation in 2009.
- NERSC/LBNL played key role in developing critical I/O code & Viz infrastructure to enable analysis of ensemble runs and icosohedral grid.




*A. Surface temperature showing geodesic grid.*
*B. Composite plot showing several variables: wind velocity (surface pseudocolor plot), pressure (b/w contour lines), and a cut-away view of the geodesic grid.*

# Material Science: Optical Data Storage

***Objective***: Explore ultrafast optical switching of nanoscale magnetic regions.

***Implications***: Potential for laser operated hard drives, 1000s of times faster than today's technology.

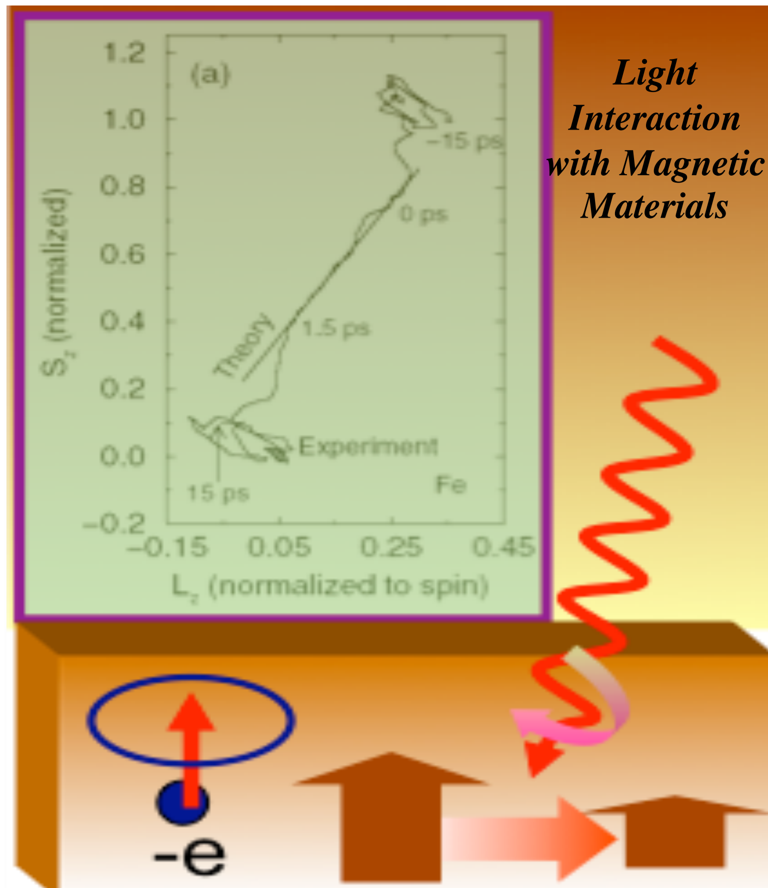
***Accomplishments:*** First-principles, time- & spin-dependent DFT study using locally-designed code on laser-irradiated Ni.

- Discovered that light leverages the crystal structure to transfer spin of electrons to higher orbit
- Study is the first to clearly demonstrate that this phenomenon is a relativistic effect connected with electron spin.
- Discovery matches experiment and can guide synthesis of new materials.

***NERSC:***

- 1.5 M hours in 2009; typically using 2,800 cores.

**PI: G. Zhang (Indiana St)**




J. Appl. Phys. (2008)

# Supernova Core-Collapse

NERSC
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

***Objective***: First principles understanding of supernovae of all types, including radiation transport, spectrum formation, and nucleosynthesis.

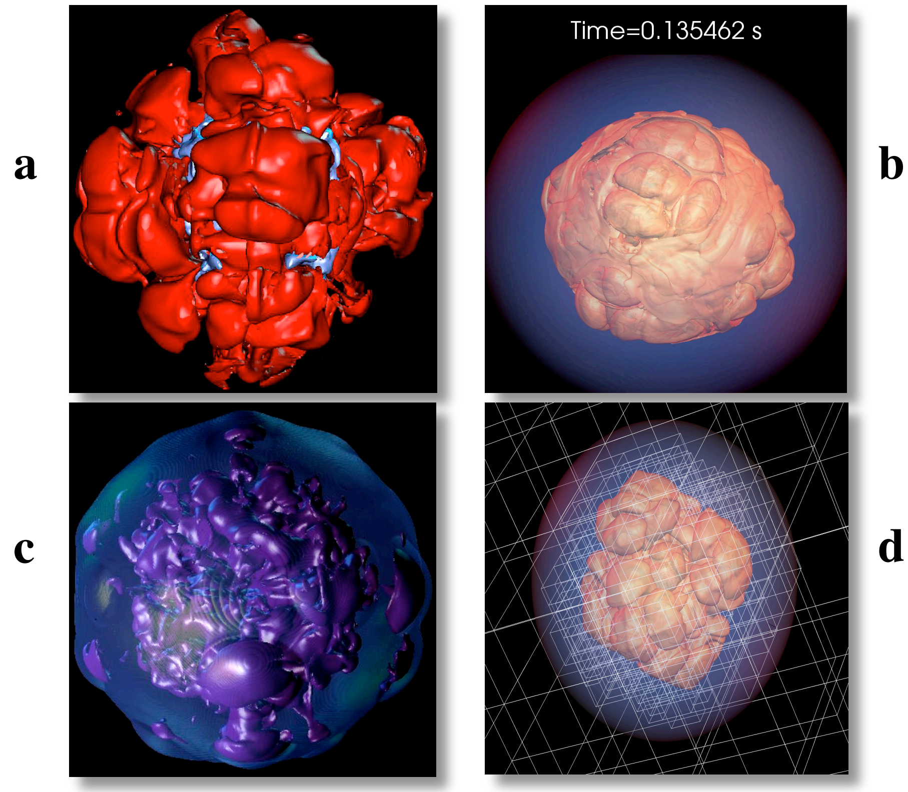
***Implications***: Will help confront one of the greatest mysteries in high-energy physics and astronomy -- the nature of dark energy.

***Accomplishments:*** NERSC runs of VULCAN core collapse explain magnetically-driven explosions in rapidly-rotating cores.

- First 2.5-D, detailed-microphysics radiation-magnetohydrodynamic calculations; first time-dependent 2D rad-hydro supernova simulations with multi-group <u>and</u> multi-angle transport.
- CASTRO, new multi-dimensional, Eulerian AMR hydrodynamics code that includes stellar EOS, nuclear reaction networks, and self-gravity.

**NERSC:** 2M hours alloc in 2009

**PIs: S. Woosley (UCSB), A. Burrows (Princeton)**




*The exploding core of a massive star. a), b), and c) show morphology of selected isoentropy, isodensity contours during the blast; (d) AMR grid structure at coarser resolution levels."*

ENERGY | Science

BERKELEY LAB

# Chemistry: Improving Catalysis

***Objective***: First-principles studies to develop better catalytic processes.

***Implications***: Improved power sources such as lithium-ion batteries, fuel cells.

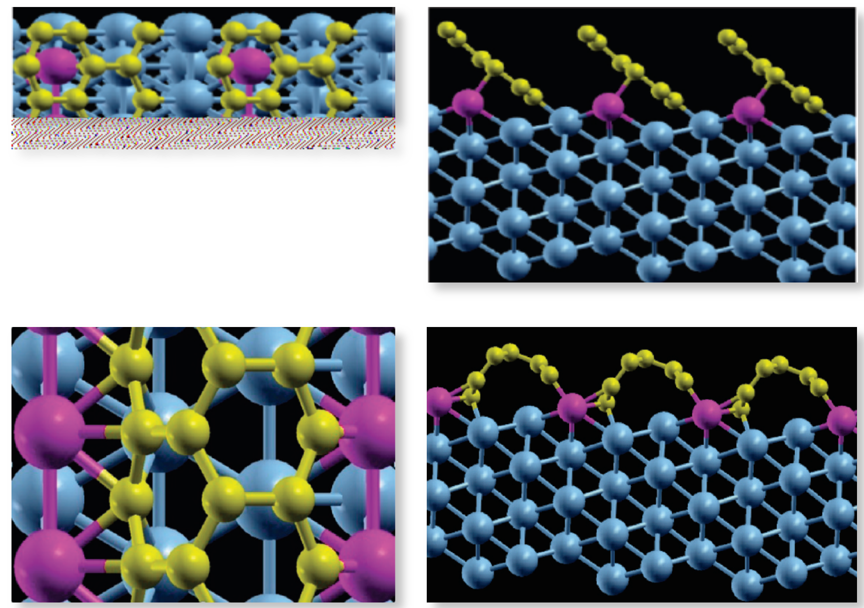
***Accomplishments:*** DFT studies of catalyzed single-walled carbon nano-tube growth on Cobalt nano-particles.

- Predict most stable adsorption sites.
- Carbon atoms form curved & zigzag chains in various orientations – some are likely precursors to graphene.
- Showed strong preference for certain metal sites.
- Next step is to investigate growth on chiral surfaces

***NERSC:***

- VASP / CPMD on Franklin; .7M hour alloc..

**PI: P. Balbuena, Texas A&M**



*Simulation showing carbon atom chains (yellow) on cobalt surfaces (blue & pink).*

**J. Phys. Chem. C, Sept, 2009 Cover Story**

# Fusion: Gyrokinetic Modeling

***Objective***: Comprehensive first-principles simulation of energetic particle turbulence and transport in ITER-scale plasmas.

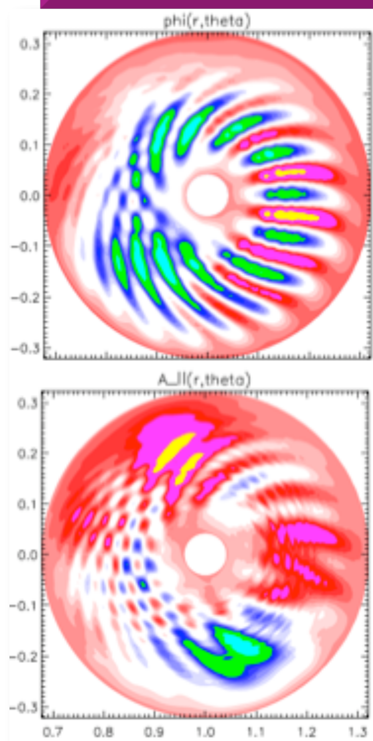
***Implications***: Improved modeling of fusion systems is essential to achieving the predictive scientific understanding needed to make fusion safe and practical.

***Accomplishments:*** GTC simulation explains measurement of fast ion transport in General Atomics DIII-D tokamak shot.

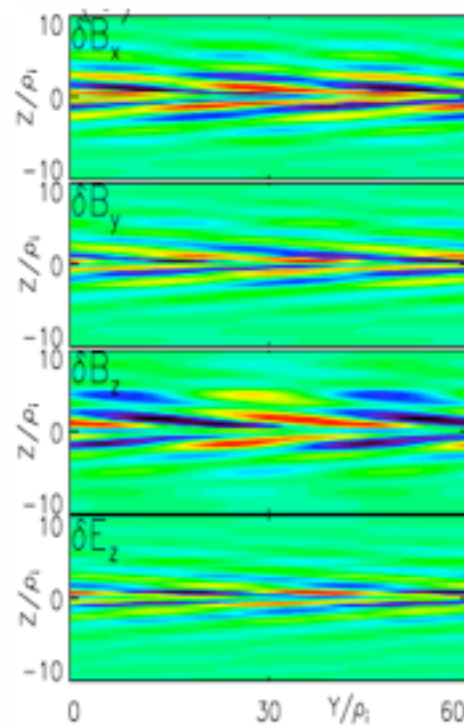
- Diffusivity decreases drastically for high-energy particles due to averaging effects of large gyroradius and banana width, and fast wave-particle decorrelation.
- 3 Fall 2009 invited talks.

***NERSC:*** 4M hours used in 2009; GTC part of NERSC6; 15-hour, 6,400-node run in March, 09

## PI: Z. Lin, UC Irvine




*Gyrokinetic simulation with kinetic electrons using a hybrid model in GTC.*




*2-D Electromagnetic field fluctuations in a simulated plasma due to microinstabilities in the current.*

Comm Comp Phys (2009)

Phys Rev Lett (2008)
Phys Plas. (2008)

# Cover Stories from NERSC Research


**NERSC is enabling new science in all disciplines, with about *1,500 refereed publications* per year**