NERSC Role in Basic Energy Science Research

Katherine Yelick
NERSC Director

Requirements Workshop
Accelerate the pace of scientific discovery for all DOE Office of Science (SC) research through computing and data systems and services.

Efficient algorithms

+ flexible software
+ effective machines

→ great computational science.
NERSC is the Production Facility for DOE Office of Science

- **NERSC serves a large population**
  - Approximately 3000 users, 400 projects, 500 code instances

- **Focus on “unique” resources**
  - Expert consulting and other services
  - High end computing systems
  - High end storage systems
  - Interface to high speed networking

- **Science-driven**
  - Machine procured competitively using application benchmarks from DOE/SC
  - Allocations controlled by DOE/SC Program Offices to couple with funding decisions

**2010 Allocations**

- BES: 31%
- BER: 19%
- HEP: 15%
- FES: 19%
- NP: 11%
- ASCR: 5%
DOE Priorities for NERSC Change Over Time

Usage by Science Type as a Percent of Total Usage

- Accelerator Physics
- Applied Math
- Astrophysics
- Chemistry
- Climate Research
- Combustion
- Computer Sciences
- Engineering
- Environmental Sciences
- Fusion Energy
- Geosciences
- High Energy Physics
- Humanities
- Lattice Gauge Theory
- Life Sciences
- Materials Sciences
- Nuclear Physics

Chart showing the percentage change in usage by science type from 2002 to 2009.
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
- Introspective security

LCFs at ORNL and ANL
- 100+ users 10+ projects
- Allocations:
  - 60% ANL/ORNL managed INCITE process
  - 30% ACSR Leadership Computing Challenge*
  - 10% LCF reserve
- Science limited to largest scale; no limit to DOE/SC
- Machines procured through partnerships
- Policy-based security
NERSC 2010 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

Carver
• IBM iDataplex cluster
PDSF (HEP/NP)
• Linux cluster (~1K cores)

Cloud testbed
• IBM iDataplex cluster

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
• Euclid large memory machine
  - 512 GB shared memory
• GPU testbed
  ~40 nodes
Demand for More Computing

- 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
How NERSC Uses Your Requirements
2005: NERSC Five-Year Plan

• **2005 Trends:**
  – Widening gap between application performance and peak
  – Emergence of multidisciplinary teams
  – Flood of scientific data
  – (Missed multicore, along with most)

• **NERSC Five-Year Plan**
  – Major system every 3 years

• **Implementation**
  – NERSC-5 (Franklin) and NERSC-6 (underway) + clusters
  – **Question: What trends do you see for 2011-2015?**
    – Algorithms / application trends and other requirements
Applications Drive NERSC Procurements

Because hardware peak performance does not necessarily reflect real application performance

NERSC-6 “SSP” Benchmarks

- Benchmarks reflect diversity of science and algorithms
- SSP = average performance (Tflops/sec) across machine
- Used before selection, during and after installation
- Question: What applications best reflect your workload?
Benjamin Franklin, one of America’s first scientists, performed groundbreaking work in energy efficiency, electricity, materials, climate, ocean currents, transportation, health, medicine, acoustics and heat transfer.

NERSC-5 “Franklin”

- **Largest Cray XT4**
  - 102 cabinets
  - 9,740 Quad Core nodes
  - 38,640 CPUs (cores)
  - Novel torus network for large parallel jobs
  - Direct access to parallel filesystem

- **Performance:**
  - 25 Tflop/s of sustained application performance
  - 352 Tflop/s of Peak
NERSC-6 System “Hopper”

- Cray system selected competitively:
  - Used application benchmarks from climate, chemistry, fusion, accelerator, astrophysics, QCD, and materials
  - Best application performance per dollar based
  - Best sustained application performance per MW
  - External Services for increased functionality and availability

Phase 1: Cray XT5
- In production on 3/1/2010
- 668 nodes, 5,344 cores
- 2.4 GHz AMD Opteron
- 2 PB disk, 25 GB/s
- Air cooled

Phase 2: Cray system
- > 1 P flop/s peak
- ~ 150K cores, 12 per chip
- 2 PB disk, 80 GB/s
- Liquid cooled

Grace Murray Hopper (1906-1992)
DOE Explores Cloud Computing

- DOE’s CS program focuses on HPC
  - No coordinated plan for clusters in SC
- DOE Magellan Cloud Testbed
- Cloud questions to explore:
  - What are the costs (TCO) of clouds vs other systems?
  - Can a cloud serve DOE’s mid-range computing needs?
  - What features (hardware and software) are needed of a “Science Cloud”? Commodity hardware?
  - What requirements do the jobs have (~100 cores, I/O,...)
  - How does this differ, if at all, from commercial clouds which serve primarily independent serial jobs
- Magellan not a NERSC Program machine
  - Not allocated in ERCAP; testbed, not production
Cluster architecture

Carver
- 3200 compute cores

Magellan
- 5760 compute cores

IB Fabric
- 28 Login/network nodes
- 14 I/O nodes

10G Ethernet
- Login
- Login

Load Balancer
- HPSS

I/O
- 8G FC
- NGF
• 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 pipeline
  – 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
    • Discovered 140 supernovae in 60 nights (July-August 2009)
    • 1 of 15 international collaborators were accessing NGF data through the SG 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.

Isocontours (a) and volume rendering (b) of two trillion zones on 32K cores of Franklin.
Requirements Drive NERSC’s Long-Term Vision
• Goal is two systems on the floor at all times
• Systems procured by sustained performance
• Assume machines are well-balanced so that typical DOE application performance is 5-10% of peak
• Does DOE science justify 10x more capability by 2013? 100x by 2015?
Data Driven Science

- **Scientific data sets are growing exponentially**
  - Ability to generate data is exceeding our ability to store and analyze
  - Simulation systems and some observational devices grow in capability with Moore’s Law
- **Petabyte (PB) data sets will soon be common:**
  - *Climate modeling*: estimates of the next IPCC data is in 10s of petabytes
  - *Genome*: JGI alone will have .5 petabyte of data this year and double each year
  - *Particle physics*: LHC is projected to produce 16 petabytes of data per year
  - *Astrophysics*: LSST and others will produce 5 petabytes/year
- **Create scientific communities with “Science Gateways” to data**
Tape Archives: Green Storage

Scientific data at NERSC increases by 1.7X per year

- Tape archives are important to efficient science
  - 2-3 orders of magnitude less power than disk
  - Requires specialized staff and major capital investment
  - NERSC participates in development (HPSS consortium)

- Questions: What are your data sets sizes and growth rates?
• NERSC has a modest number of “commodity systems”
• Mostly specialized science systems for compute, disk storage (parallel filesystems), and tape archives
Since 2007, NERSC is a net data importer. In support of our users, it is important that we take on a lead role in improving intersite data transfers.

- Systems and software typically tuned only within a site
- Technical, social, and policy challenges abound:
  - High performance transfer software has too many options → hard to use.
  - Systems designed for computation can have bottlenecks in data transfers
  - Systems at different sites often have incompatible versions of transfer software.
  - Trying to maintain security exceptions (firewall holes) for all the systems and software at each site was impossible.
- … and the list goes on.
- NERSC established Data Transfer Nodes (DTNs).
  - Reduced transfer time of 30 TB from 30+ days to 2 days
  - We formed a working group with experts at the three labs and ESNet

http://www.nersc.gov/nusers/systems/DTN
http://fasterdata.es.net
Energy Efficiency is Necessary for Computing

- Systems have gotten about 1000x faster over each 10 year period
- 1 petaflop ($10^{15}$ ops) in 2010 will require 3MW
  $\Rightarrow$ 3 GW for 1 Exaflop ($10^{18}$ ops/sec)
- DARPA committee suggested 200 MW with “usual” scaling
- Target for DOE is 20 MW in 2018
Moore’s Law is Alive and Well

Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović
But Clock Frequency Scaling Replaced by Scaling Cores / Chip

Slide Source: Kathy Yelick. Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović
Performance Has Also Slowed, Along with Power (Root Cause)

Slide Source: Kathy Yelick. Data from Kunle Olukotun, Lance Hammond, Herb Sutter, Burton Smith, Chris Batten, and Krste Asanović
• Computational scaling changed in 2004
• Problems also for laptops, handhelds, data centers
• Parallelism on-chip brings algorithms, programming into question
• **NERSC: Programmable, usable systems for science**
  1) Energy efficient designs
  2) Facilities to support scale for both high and mid scale
Parallelism is “Green”

• Concurrent systems are more power efficient
  – Dynamic power is proportional to $V^2 f C$
  – Increasing frequency ($f$) also increases supply voltage ($V$) → cubic effect
  – Increasing cores increases capacitance ($C$) but only linearly

• High performance serial processors waste power
  – Speculation, dynamic dependence checking, etc. burn power
  – Implicit parallelism discovery

• Question: *Can you double the concurrency in your algorithms and software every 2 years?*
Technology Challenge

Technology trends against a constant or increasing memory per core

- Memory density is doubling every three years; processor logic is every two
- Storage costs (dollars/Mbyte) are dropping gradually compared to logic costs

**Question:** *Can you double concurrency without doubling memory?*

*Source: David Turek, IBM*
Hardware and Software Trends

• **Hardware Trends**
  – Exponential growth in explicit on-chip parallelism
  – Reduced memory and memory bandwidth per core
  – Heterogeneous computing platforms (e.g., GPUs)
  – As always, hardware is largely driven by non HPC markets

• **Software Response**
  – Need to express more explicit parallelism
  – New programming models on chip: MPI + X
  – Increased emphasis on strong scaling
  – No more serial code scaling from hardware

• **What we want**
  – Understand your requirements and help craft a strategy for transitioning to a hardware and programming environment solution
Basic Energy Science at NERSC
**Objective:** Develop Quantum Monte Carlo (QMC) methods to stochastically solve many-body electronic structure problems.

**Implications:** Accurately predict or explain chemical phenomena where other methods fail or aren’t applicable.

**Accomplishments:** Developed hybrid QMC / Molecular Mechanics formalism.
- Obtained interaction energy of a 2-water cluster treating one H\textsubscript{2}O quantum mechanically and other classically; prelude to effort to find much sought-after electron binding energy in (H\textsubscript{2}O\textsubscript{n}).
- Studied series of Li clusters in different charge states to obtain energies for cluster growth, charge, and discharge in interactions with graphene.
- ZORI scales to 32K cores on Franklin
**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
- Modest parallelism

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

*J. Phys. Chem. C, Sept, 2009 Cover Story*
**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.
- Used over 1.5M hours in 2009; 2,800 cores
Finding Hidden Oil / Gas Reserves

Objective: Apply new, highly rigorous, massively parallel, 3-D imaging techniques to create geophysical maps of hydrocarbon reservoirs in unprecedented levels of detail.

Implications: New detection abilities and exploration savings by revealing where hydrocarbon deposits reside, even when covered by ocean over a mile deep and several more miles of rock below the ocean. Can also be used for locating potential sites for CO₂ sequestration.

Accomplishments: Has already provided insight into complex geology of Campos Basin, a petroleum rich area near Brazil.

NERSC: Code developed on Franklin.

- Algorithms can run on O(10,000) cores; designed to scale well beyond. Runs on Franklin routinely use 4,000-8,000 cores.

http://escholarship.org/uc/item/0qh3p22m
Graphene as the Ultimate Membrane for Gas Separation

Objective: Study permeability, selectivity of graphene with custom sub-nanometer pores using *ab initio* DFT and vdW DF.

Implications: Potentially lower energy costs for purification and production of key industrial gases such as H₂ and methane.

Accomplishments: Such pores exhibit extremely high selectivity, presenting a formidable barrier for CH₄ but easily surmountable for H₂.
- Results suggest that graphene may be superior to traditional membranes.
- Could have widespread impact on numerous energy and technological applications, including carbon sequestration, fuel cells and gas sensors.

Nitrogen-functionalized pore in graphene (a), electron density isosurface (b), and snapshots of H₂ diffusing through the pore from NERSC first principles molecular dynamics simulations.

*D. Jiang (PI), V. Cooper, S. Dai (ORNL)*

*Nano Lett.*, 9, 4019 (2009)
Restructuring Catalyst Surfaces

**Objective:** Use simulation to understand the ability of surfaces to restructure under the influence of gaseous adsorbates.

**Implications:** Revealing the arrangement of metal atoms that form at active sites will yield increased understanding of heterogeneous catalysis mechanisms.

**Accomplishments:** DFT studies at NERSC show that CO molecules bind to small Pt nanoclusters on the catalyst surface.

- The nanoclusters seem to maximize bonding of more CO molecules.
- VASP reveals the stabilization energy gained by cluster formation and suggests the atomic arrangement.
- Formation of small metallic clusters opens a new avenue for understanding catalytic activity under high pressures.

*(top) Starting geometry of CO and Pt atoms. (bottom) After relaxation to minimize energy in the DFT calculation, two (3 × 3) clusters form. Dark blue circles represent Pt atoms in the slab layers; light blue circles represent Pt atoms at the surface. Red and gray circles represent oxygen and carbon atoms, respectively.*

**Objective:** Study small metal clusters supported on nanoparticles to understand heterogeneous catalysis; help design improved catalysts.

**Implications:** Better hydrodesulphurization in power plants; possible conversion and use of non-conventional fuels, e.g., MeOH.

**Accomplishments:** DFT calculations and state-of-the-art cluster beam studies provide insight into the reaction mechanism of catalytic activity of molybdenum-sulfur clusters on gold surfaces.

- Help identify intermediates along the catalytic reaction pathway.
- Used over 700K hours in 2009
- Special ORNL/NERSC version of VASP

Potential energy profile for the interaction of Carbonyl Sulfide (OCS) on Mo$_6$S$_8$/Au(111)

\[ \Delta E (eV) \]

\[ TS1: OCS \rightarrow S_{(ads)} + CO_{(ads)} \]
\[ TS2: Migration of S onto the Au(111) surface \]

**Objective:** design, simulate and help realize nanoscale molecular transport systems.

**Implications:** Possible use in drug delivery, advanced sieves, desalination.

**Accomplishments:** Simulations of nanomotors, nanotubes, micelles, and custom-designed nanopores using Molecular Dynamics.

- Showed that
  - Electron tunneling can drive nano-scale motors used in nanopropellers.
  - Functionalized graphene-based nanopores can serve as ionic sieves.
  - Nanodroplets can be dragged on the surface of carbon nanotubes.
- NAMD on Franklin; >250K hour in 2009

Left: Nanomotor rotates in presence of electric field; Right: two example highly-selective nanosieves – only certain ions pass across.

Objective: Electronic structure studies of complex ceramic materials with outstanding thermal & electrical properties.

Implications: Connection of atomic-scale characteristics with engineering mechanics and elucidation of properties not available by any other method.

Accomplishments: VASP DFT study of mechanical response and failure behavior of intergranular glassy films (IGFs) in Silicon Nitrides.

- Stress/strain relationship explained by fundamental electronic structure of the model.
- May be used to guide future material designs that enhance selective properties.
- Used >2.5M hours on Franklin in 2009

IGF model at five different levels of strain (left) and stress vs. strain curve (right) showing a kink in the rising stress before failure, followed by a step-like post-failure behavior at higher strains.
**Objective:** Accurate structural studies of contaminants in solution.

**Implications:** Predict long-term viability of nuclear waste containment strategies.

**Accomplishments:** Two different NWChem *ab initio*-DFT analyses of Uranium Oxide ion (UO$_2^{2+}$), one with 64 H$_2$O molecules for 22 ps and one with 122 waters for 9 ps.

- Extremely-demanding simulations due to large # of H$_2$O molecules and long integration times.
- Results help explain X-Ray spectra but also reveal additional structural features.
- Also provides NWChem to other NERSC users

*First-principles molecular dynamics simulation of 2nd hydration shell surrounding UO$_2^{2+}$ with 3 intermediate dissociative structures.*

*J Chem Phys (2008)*
Catalysis for Higher Fuel Cell Efficiency

Objective: Identify and evaluate catalytic surfaces aimed at improving the efficiency of Direct Methanol Fuel Cells (DMFCs).

Accomplishments:
- Used DFT to develop an electrochemical model to evaluate catalytic surfaces for methanol oxidation.
- Model helps identify properties of an ‘ideal’ catalyst and allows screening of novel systems that may be better and cheaper than current technology.
- Significant effort by NERSC Services for special VASP version used here

Implications: Lower power, more efficient and economical DMFCs have potential applications in powering mobile phones and laptop batteries and as an alternatives to current hydrogen fuel cell technology.

This figure shows the potential determining steps from the DFT calculations. It helps predict the lowest possible potential of a fuel cell, which is directly related to the efficiency of the catalyst.

NISE: NERSC Initiative for Scientific Exploration

- INCITE program was created at NERSC:
  - Focused computing and consulting resources
  - Became the allocation mechanism for the LCFs
- NISE program at NERSC (started in 2009)
  - Programming techniques for multicore and scaling in general
  - Science problems near breakthrough (encourage high risk/payoff)
  - [http://www.nersc.gov/nusers/accounts/NISE.php](http://www.nersc.gov/nusers/accounts/NISE.php)
- AY09 (10/1/09 through 01/11/10) ~20M Franklin hours made available to existing NERSC projects; ~6.9M to BES
  - Elasticity of b-DNA Models; Metallic Alloy Fatigue Fracture; Minimum Free Energy Paths/Profiles of Protein Conformational Changes; Diamondoid-Nanoparticle Enhanced Organometallic Surfaces; Organic Photovoltaics, AIMBP'T; MeOH Reformation; Nanocluster Energy Landscape; Graphene Bilayer Electronic Device; Sec Translocase Transmembrane Channel

- See also ASCR’s ALCC Program:
Conclusions

• NERSC requirements
  – Qualitative requirements shape NERSC functionality
  – Quantitative requirements set the performance
    “What gets measure gets improved”

• Goals:
  – Your goal is to make scientific discoveries
    • Articulate specific scientific goals and implications for broader community
  – Our goal is to enable you to do science
    • Specify resources (services, computers, storage, …) that NERSC could provide with quantities and dates
NERSC is enabling new science in all disciplines, with about 1,500 refereed publications per year.
Low swirl burner combustion simulation. Image shows flame radical, OH (purple surface and cutaway) and volume rendering (gray) of vortical structures. Red indicates vigorous burning of lean hydrogen fuel; shows cellular burning characteristic of thermodiffusively unstable fuel. Simulated using an adaptive projection code. Image courtesy of John Bell, LBNL.

Hydrogen plasma density wake produced by an intense, right-to-left laser pulse. Volume rendering of current density and particles (colored by momentum orange - high, cyan - low) trapped in the plasma wake driven by laser pulse (marked by the white disk) radiation pressure. 3-D, 3,500 Franklin-core, 36-hour LOASIS experiment simulation using VORPAL by Cameron Geddes, LBNL. Visualization: Gunther Weber, NERSC Analytics.

Numerical study of density driven flow for CO₂ storage in saline aquifers. Snapshot of CO₂ concentration after convection starts. Density-driven velocity field dynamics induces convective fingers that enhance the rate by which CO₂ is converted into negatively buoyant aqueous phase, thereby improving the security of CO₂ storage. Image courtesy of George Pau, LBNL.

False-color image of the Andromeda Galaxy created by layering 400 individual images captured by the Palomar Transient Factory (PFT) camera in February 2009. NERSC systems analyzing the PTF data are capable of discovering cosmic transients in real time. Image courtesy of Peter Nugent, LBNL.

The exciton wave function (the white isosurface) at the interface of a ZnS/ZnO nanorod. Simulations performed on a Cray XT4 at NERSC, also shown. Image courtesy of Lin-Wang Wang, LBNL.

Simulation of a global cloud resolving model (GCRM). This image is a composite plot showing several variables: wind velocity (surface pseudocolor plot), pressure (b/w contour lines), and a cut-away view of the geodesic grid. Image courtesy of Professor David Randall, Colorado State University.