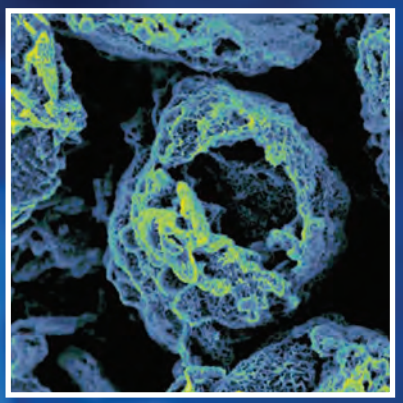
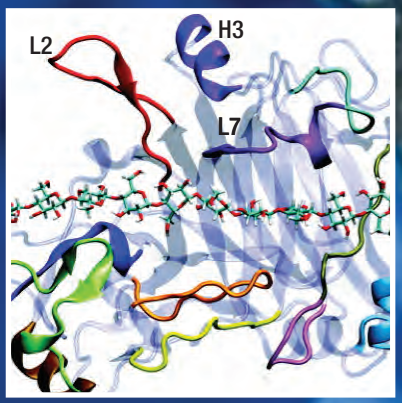
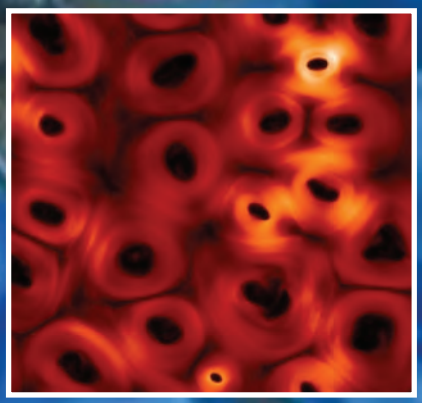






National Energy Research Scientific Computing Center

2011 Annual Report

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, Berkeley, CA 94720-8148

This work was supported by the Director, Office of Science, Office of Advanced Scientific Computing Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

National Energy Research Scientific Computing Center

2011 Annual Report

Table of Contents

The Year in Perspective

The year 2011 was unprecedented for scientific accomplishments at NERSC. Taking center stage was our support of cosmology. In October, Berkeley Lab's Saul Perlmutter was awarded the Nobel Prize in Physics for his research that found that the Universe is continuing to expand and that the rate of expansion is gradually accelerating. A key component of this work was the simulation of 10,000 supernovae on NERSC supercomputers a decade ago. And Peter Nugent, who performed those simulations, made headlines around the world in August when he discovered the closest supernova to Earth in 25 years, just as the star was beginning its explosive demise.

Our center again proved its leadership in scientific computing by moving Hopper, our first petaflops system, into production and supporting a broad range of research programs. Meanwhile, plans for the NERSC-7 system got underway, and NERSC staff, in collaboration with DOE program managers and users, completed the last of six workshops to map out future computing requirements for the program offices in the Office of Science. The bottom line indicates that demand for NERSC computing is 10 times greater than the resources we expect to have available in the next few years.

One way that NERSC is leveraging its resources to serve a broader community is by building "science gateways" to share both results and raw data with other researchers. Of particular note in 2011 was the launch of the Materials Project, which is a "Google of material properties" and is being shared with industry and academia. The Materials Project database currently contains the structural and energetic properties of more than 28,000 inorganic compounds, and up to hundreds more are added every day. This gateway opens the door for more research in support of the Materials Genome Initiative announced by the White House in June 2011.

The year ended with the preparation of an in-depth report analyzing the results of our Magellan research effort on cloud computing, which showed that NERSC's HPC systems substantially outperform standard commercial clouds in both runtime and cost while offering unique software and expertise for scientific computing.

While these selected highlights illustrate the breadth of the NERSC Center's systems and support expertise, running large-scale simulations day in and day out is the bread and butter of NERSC computing. Over 40 percent of the time used by jobs are spread over at least one eighth of the machine. Such capability jobs take advantage of the unique networks in these systems. At the same time, NERSC offers this singular capability to university researchers, who make up 65 percent of our users and typically don't have access to similar resources at their own institutions.

Large-scale simulation is the foundation of computing at NERSC, but we are seeing a move toward scientific workflows with very large numbers of jobs and data analysis involving massive datasets. To help people understand NERSC's role in these ongoing and complex changes, I've recently been discussing three trends: "science at scale," "science through volume" and "science in data." And as described below, these three trends are closely intertwined. 2011 saw huge growth in science through volume and science in data. This annual report provides examples of NERSC's contributions to all three trends.

Science at Scale

Science at scale is what supercomputers are best known for—using thousands of processor cores to perform massive calculations that simulate complex phenomena. These simulations can lead to new understanding and even the discovery of new phenomena.

Large-scale simulations can also challenge long-held assumptions. One example is modeling of an event that coincided with the abrupt climate change that took place about 8,200 years ago, when a massive ice sheet that covered much of Hudson Bay in arctic Canada melted and broke through an ice dam, suddenly dumping thousands of cubic kilometers of fresh water into the Atlantic Ocean. Geoscientists had assumed that this flood covered the surface of the Labrador and Greenland Seas, slowing the Gulf Stream and ocean convection. But a new model, which has 10 to 20 times higher resolution than previous models and runs on 1,800 computer cores, suggests the freshwater flood actually flowed much farther south on the continental shelf into the subtropical North Atlantic. This finding, bolstered by observational evidence, challenges not just our understanding of ancient climate change but also the assumptions built into ocean and sea ice models that we use to project our future climate.

An example of how science at scale can guide experimentation is the discovery of fluorine-14, an exotic, short-lived nucleus that was first predicted in simulations and then detected experimentally, increasing our understanding of the strong force. The simulation of fluorine-14 was made possible by decades of code development and optimization of the code to improve performance and mathematical precision on petascale computers.

Today, petascale computers—capable of 1 quadrillion (10^{15}) floating point operations per second (flops)—are state of the art. But NERSC is collaborating to research the path to exascale—1 quintillion (10^{18}) flops—which we hope to see within the next decade. NERSC will help vendors transition to exascale system architectures that will be useful for more than 600 scientific applications; and we will help more than 4,000 users make the transition to the new technologies, just as today we are helping users run their codes effectively on Hopper's multi-core nodes. Berkeley Lab's Computational Research and Theory Facility will be the most energy efficient computing facility in DOE, capable of supporting two exascale systems.

Science through Volume

Unlike science at scale, which involves small numbers of heroically large simulations, science through volume involves large numbers—up to tens of thousands—of smaller jobs, sometimes referred to as high-throughput computing. At NERSC, research groups in several science domains are using high-throughput computing to create large databases that serve as public resources for researchers. As noted above, one such database is used for screening properties of materials.

New materials are crucial to building a clean energy economy—for everything from batteries to photovoltaics to lighter weight vehicles—but today the development cycle is too slow: around 18 years from conception to commercialization. To cut this gap in half, researchers from Berkeley Lab and MIT teamed up to develop a new computational screening tool, called the Materials Project, which was launched in October 2011 using NERSC's science gateway infrastructure. The team is using NERSC systems to characterize the properties of inorganic compounds, such as their stability, voltage, capacity and oxidation state. The results are then organized into a database with a user-friendly web interface that gives the entire research community free and easy access and search capabilities. The purpose is not designing materials, but picking the interesting ones for more detailed research.

We are also seeing increased interest in improving uncertainty quantification in numerical simulations, though this is not a new idea. One of NERSC's early breakthroughs was the

discovery of dark energy in 1998, for which Saul Perlmutter of Berkeley Lab shared the 2011 Nobel Prize in Physics. His research team is believed to have been the first to use supercomputers to analyze and validate observational data in cosmology. To analyze the data from 40 supernovae for errors or biases, Perlmutter's team simulated 10,000 exploding supernovae at varying distances under varying circumstances. These were then plotted and compared with the observed data to detect any biases affecting observation or interpretation. NERSC's Cray T3E supercomputer was also used to check and recheck their work by resampling the data and running calculations that helped determine the reliability of their measurements thousands of times. This melding of computational science and cosmology sowed the seeds for more projects, establishing Berkeley Lab and NERSC as centers for this emerging field.

Science in Data

As a national resource for scientific computing with a focus on productivity, NERSC has long been both a producer and a consumer of big data—producing massive datasets through simulations, and analyzing and storing massive datasets from experiments and detectors. A single simulation can now produce tens to hundreds of terabytes, comparable to the amount of information found in the Library of Congress print collection. The petascale (and soon exascale) datasets produced by ensembles of simulations or large-scale experiments are too big to analyze manually, so new data mining, statistical learning, visualization, and topological analysis techniques are being developed to aid discovery.

NERSC has developed and deployed a variety of systems and tools to meet the needs of data-intensive workloads. We have deployed dedicated data servers to support data analysis, as well as data transfer nodes—servers dedicated to performing transfers between storage resources at NERSC and other sites—to speed up data sharing between collaborators. NERSC has developed a new data tracking system that reveals data storage and usage patterns, enabling us to speed up data access for users and make well-informed plans for future storage hardware and software systems. NERSC's new Parallel Incremental Backup System provides daily backups that protect users' data efficiently without interfering with their workflows.

As one of example of a data-analysis workload, the DOE Joint Genome Institute (JGI)—one of the most productive genomics user facilities in the world—began to merge its computing operations and staff into NERSC to take advantage of our computing and storage systems and our broad expertise. Using 20,000 to 75,000 processors on Hopper, NERSC can analyze 100 million genes in a few days, a task that used to require weeks at JGI. The transferred JGI staff were originally a separate genomics group; but we quickly realized that JGI's needs are not unique, and that the tools needed to manage the genomics workload could be used by other data-intensive research groups as well. The transition began in 2010, and at this point the JGI computing staff are fully integrated into previously existing NERSC teams.

NERSC has worked with users in the genomics community to understand their workloads and developed tools to help lower the barrier to entry for these users. A few examples of the innovations we have developed include a task farmer that simplifies running high-throughput workloads, tools to create a private virtual cluster, and file caching utilities for data-intensive applications that reduce the stress on the file system. With this support, JGI's Integrated Microbial Genomes (IMG) database, which serves as a community resource for comparative analysis and annotation of all publicly available microbial genomes, crossed the 1 billion gene milestone in December 2011 and has since grown to over 3 billion.

NERSC resources are also used to develop some of the latest mathematical techniques in data analysis, such as

an application of the "compressed sensing" method to be used to determine energy potentials in protein–DNA interactions, revealing where genes switch on and off at the atomic level. This method has tremendous promise not just in biology and medicine, but in many physics and chemistry applications.

Another example of science in data is finding extreme weather events such as hurricanes in multi-decade climate simulation datasets. A multi-laboratory and university collaboration is developing an automated data mining system to identify and quantify extreme weather phenomena in the large datasets generated by today's climate models. The results of this work will help answer the question of how climate change will affect the frequency of storms, floods, and droughts.

Eighteen examples of the heaviest antiparticle ever found, the nucleus of antihelium-4, have been made in the STAR experiment at the Relativistic Heavy Ion Collider at Brookhaven National Laboratory. Identifying those 18 particles required sifting through the debris of a billion heavy-ion collisions—a good example of science in experimental data. That sifting was done on the PDSF high energy physics computing cluster hosted at NERSC, using 1 petabyte of data stored on NERSC's HPSS system.

For a growing number of research groups, science in data involves automated workflows that screen data and select a small number of candidates for human follow-up. The value of automated workflows was demonstrated by the Palomar Transient Factory's (PTF's) August 2011 discovery of an unusually close Type 1a supernova only 11 hours after it exploded. The speed of this discovery enabled the researchers to calculate the actual moment of the explosion to within 20 minutes, and to gather unprecedented data about the early stages of the explosion, including the first direct evidence that a supernova started as a white dwarf, evidence that the companion star was not much bigger than our sun, and surprising details about the non-uniform, "clumpy" nature of the explosion. To make this possible, NERSC's data transfer nodes accept 100 gigabytes of automated telescope data every night, which is analyzed immediately on the Carver cluster and posted to the database on PTF's science gateway web portal for instant worldwide access. To date, this system has enabled the PTF to discover 1,500 supernovae, including 1,000 Type 1a supernovae, with archived data on over 800 million candidate detections.

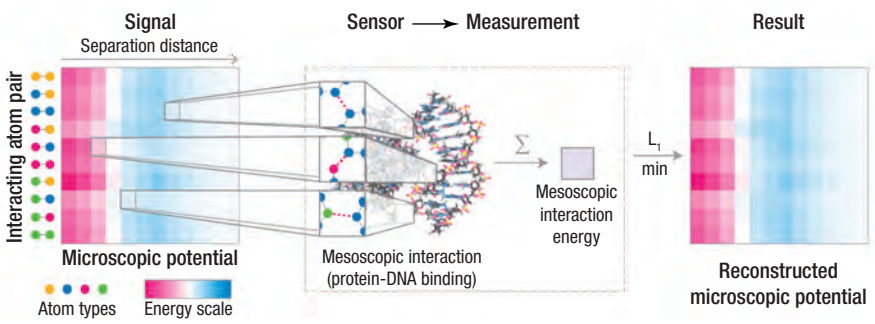
Looking to the Future

The demands from larger and more detailed simulations, massive numbers of simulations, and the explosion of experimental data sets mean the there is no end in sight to the need for NERSC resources. Combined with the challenges of clock speed scaling, memory balance, power and resilience all mean that the next decade will rely, more than ever, on the creativity and expertise of the NERSC staff in collaboration with vendor partners and the user community. Staying ahead of the technological curve, anticipating problems, and developing solutions that are effective for the broad science workload are part of NERSC's culture.

With all the advanced technologies we deploy, it is still people who make the difference. I am grateful to our DOE Office of Science sponsors for their continued endorsements, to our users who inspire us with their scientific results, the vendors who work with us to improve the technology, and to the NERSC staff for their dedication to the NERSC mission.

Katherine Yelick
Associate Laboratory Director for Computing Sciences

New Mathematical Method Reveals Where Genes Switch On or Off



Project Name
Computational Prediction of Transcription Factor Binding Sites

Project Leader
Harley McAdams, Stanford University

NERSC Resources
3.6 M hours on Carver

DOE Office
Biological and Environmental Research

Developmental biologists at Stanford University, using computing resources NERSC, have applied a new mathematical method called *compressed sensing*—previously used in signal processing—to biochemistry, using it to reveal the atomic-level details of protein–DNA interactions with accuracy approaching experimental measurement. They hope compressed sensing will speed up research into where genes are turned on and off, and they expect it to have applications in many other scientific domains as well.

Transcription factors are proteins that bind to specific DNA sequences to either activate or block the expression of a gene. They are involved in many important processes, including cellular differentiation, responses to environmental stimuli, cell growth and death, the onset of certain diseases, especially cancers, and responses to certain drugs. Understanding exactly where transcription factors bind with DNA could help answer many crucial questions in biology. But determining this experimentally requires significant labor and financial resources, so a computational solution is highly desirable. Analyzing biochemical phenomena at the molecular level involves determining the *energy potential*, a mathematical description of the energy of every possible interaction in a molecular system. Currently it's not feasible to calculate energy potentials for large molecular systems from first principles, even on supercomputers—the calculations are simply too large.

Energy potential determination as an application of compressed sensing. The potential is represented by a heat map of microscopic interaction energies ranging from repulsive (dark pink) to attractive (dark blue). *(Mohammed AlQuraishi and Harley McAdams)*

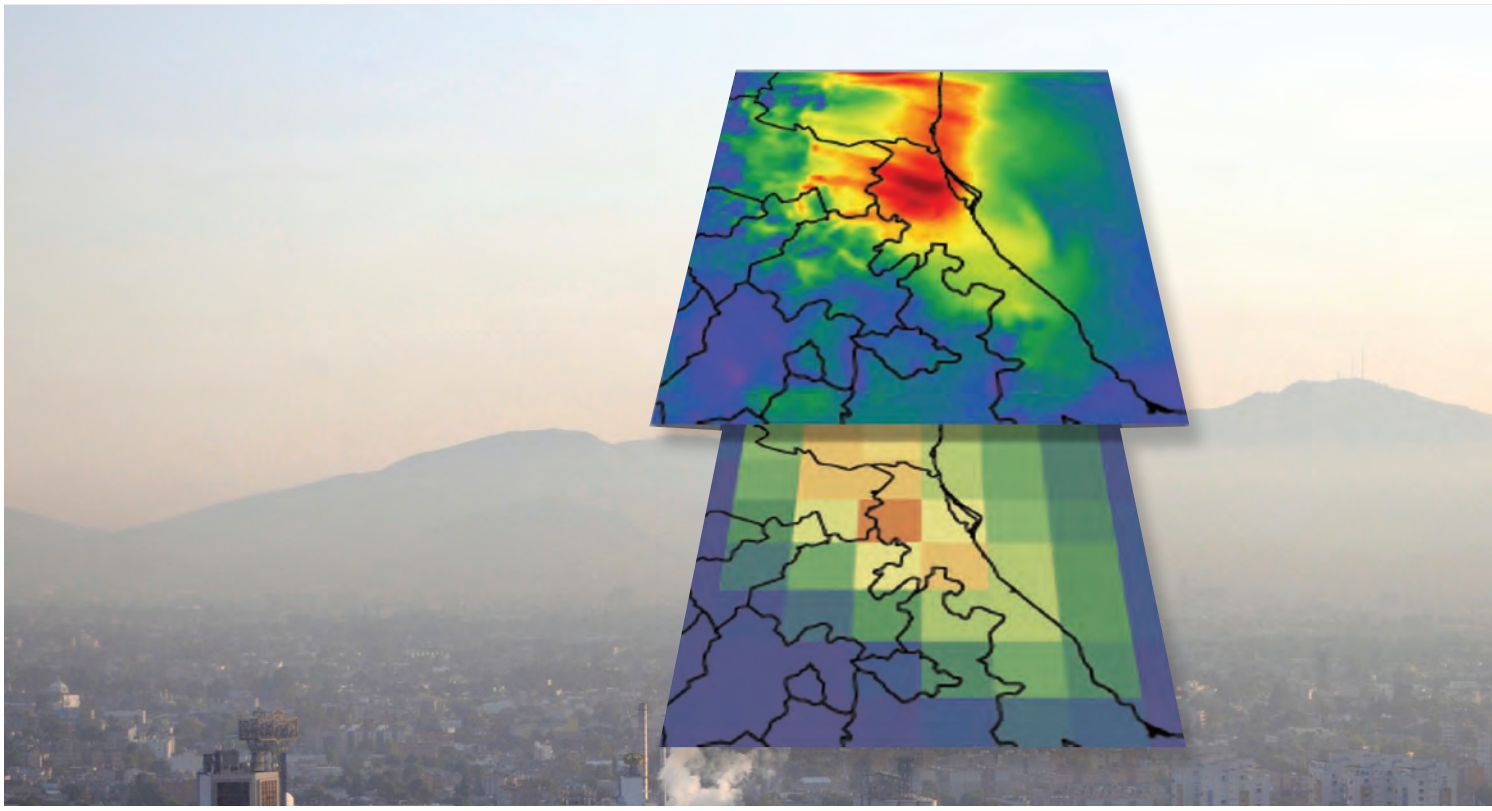
Mohammed AlQuraishi, a postdoctoral researcher at Stanford, realized that determining atomic-level energy potentials in protein–DNA complexes could be treated mathematically as a signal acquisition problem. Compressed sensing allows the recovery of sparse signals from a few carefully constructed but seemingly random measurements. (In mathematical terms, it's a technique for finding sparse solutions to underdetermined linear systems.) It allows a complete signal (such as a visual image) to be communicated and reconstructed with surprisingly little data.

By using the crystal structures of 63 protein–DNA complexes as the sensors, and using the experimentally determined binding affinities of the complexes as the compressive measurements, the researchers determined the "signals"—the attraction between specific pairs of protein and DNA atoms. The results were 90 percent accurate, compared to 60 percent for the best performing alternative computational methods.

Full story: http://www.nersc.gov/news-publications/news/science-news/2012/new-mathematical-method-reveals-where-genes-switch-on-or-off/

Publication: Mohammed AlQuraishi and Harley McAdams, "Direct inference of protein–DNA interactions using compressed sensing methods," PNAS 108, 14819 (2011), doi:10.1073/pnas.1106460108.

Small Particles Have Big Impact



Project Name
Regional Multi-Scale Modeling of Aerosols and Cloud Properties

Project Leader
Jerome Fast, Pacific Northwest National Laboratory

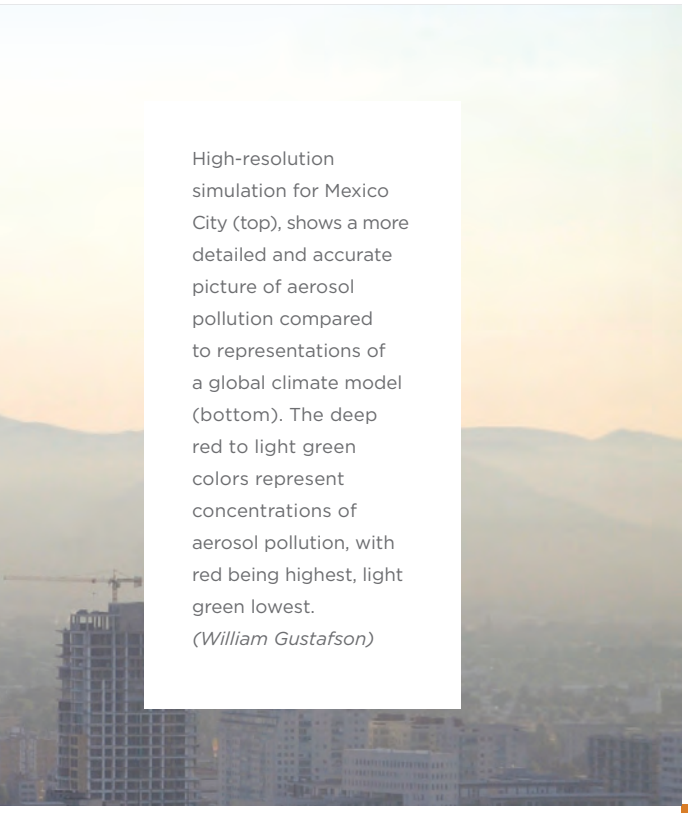
NERSC Resources
Franklin

Other Computing Resources
EMSL

DOE Office
Biological and Environmental Research

Using systems at NERSC, atmospheric scientists at Pacific Northwest National Laboratory (PNNL) have found that small-scale effects of aerosols—tiny particles of dust or pollution in the atmosphere—can add up and over time and lead to large, accumulated errors in climate prediction models.

"Aerosols like ozone, dust, and sea salt in our atmosphere scatter and absorb sunlight," says William Gustafson, an atmospheric scientist at PNNL and principal investigator of the study. "Depending on the type of particle and its elevation above Earth's surface, these particles can tip the energy balance toward heating or cooling." Because global climate models typically calculate atmospheric processes at scales close to 100 by 100 kilometers (62 miles), the characteristics of aerosols are averaged over a large area. This practice distorts the effects of aerosols in climate predictions because, in reality, these particles act on a much smaller scale and can vary according to local atmospheric or geographic features.



High-resolution simulation for Mexico City (top), shows a more detailed and accurate picture of aerosol pollution compared to representations of a global climate model (bottom). The deep red to light green colors represent concentrations of aerosol pollution, with red being highest, light green lowest. *(William Gustafson)*

To quantify this error, Gustafson led a collaboration that looked at changes in the net flux of sunlight—the amount of sunlight that is reflected back into space, absorbed, or allowed to hit the ground by aerosols—using a regional atmospheric model to emulate grids typical of coarse global climate models, as well as detailed grids at scales of 3 by 3 kilometers (1.6 miles). The 30-day-long simulations were based on observations collected in March 2006.

The team's results revealed a 30 percent discrepancy between the coarse and detailed models for aerosol direct radiative forcing—rate of energy change at the top layer of the atmosphere due to the aerosols—over portions of Mexico.

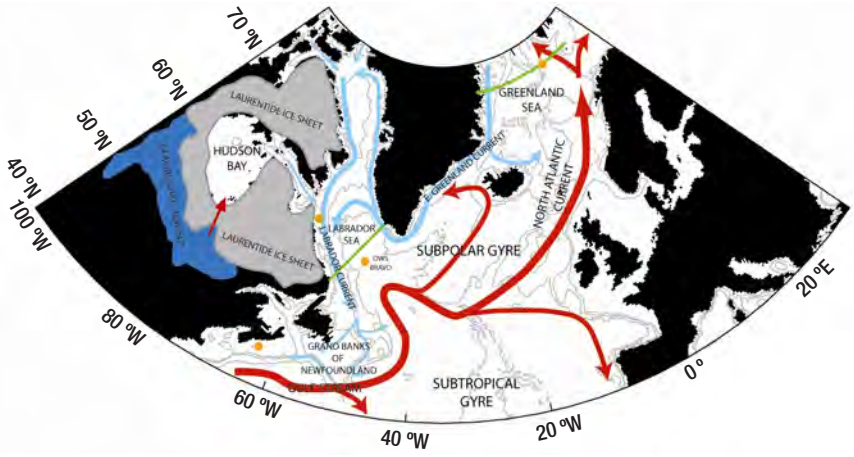
"Until recently, computers weren't fast enough to incorporate detailed effects of aerosols in climate models. These calculations are very computationally expensive," says Gustafson. "On average, if you are trying to predict climate without these particles, there may be 15 variables to consider. With aerosols, there are anywhere between 50 and 500 variables."

With more funding and compute power, Gustafson says he would like to measure these effects on a global level, as well as be able to include the impact on clouds, which was excluded in the present study.

Full story: http://www.nersc.gov/news-publications/news/science-news/2011/small-particles-big-impact/

Publication: Gustafson, W. I., Jr., Y. Qian, and J. D. Fast (2011), Downscaling aerosols and the impact of neglected subgrid processes on direct aerosol radiative forcing for a representative global climate model grid spacing, J. Geophys. Res., 116, D13303, doi:10.1029/2010JD015480.

Detailed Model Changes View of Ancient Climate Change



Project Name
High Resolution Model Development to Quantify the Impact of Icebergs on the Stability of the Atlantic Meridional Overturning Circulation

Project Leader
Alan Condron, University of Massachusetts Amherst

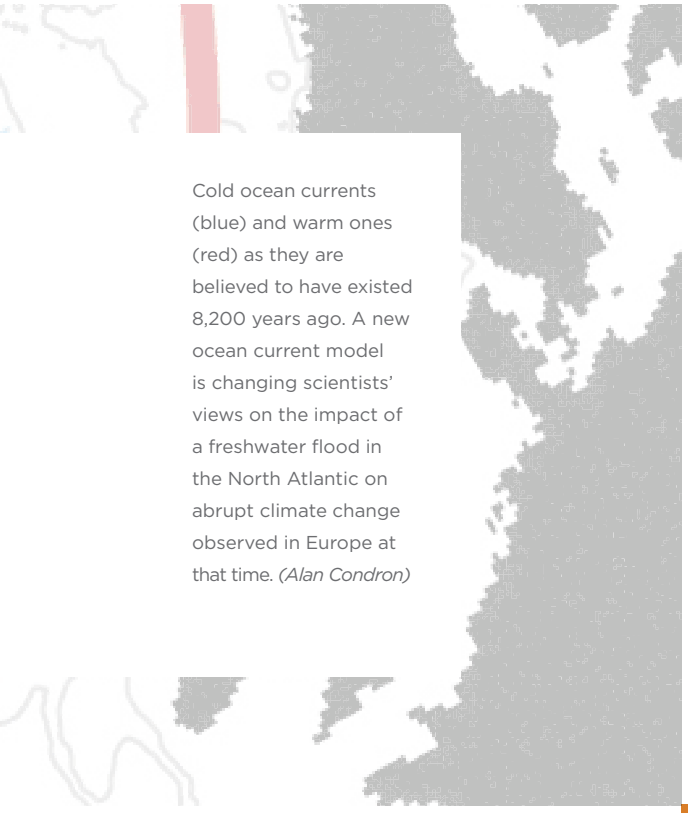
NERSC Resources
Franklin, 1,800 cores for 240 hours

DOE Office
Biological and Environmental Research

Using a new, high-resolution global ocean circulation model and NERSC computers, University of Massachusetts Amherst geoscientist Alan Condron, with Peter Winsor at the University of Alaska, discovered that massive glacial meltwaters that had been assumed to have flooded the entire North Atlantic 8,200 years ago, drastically cooling Europe, instead flowed thousands of miles further south.

The events unfolded when the Laurentide ice sheet, which covered much of Hudson Bay in Arctic Canada, gradually melted during a warm period about 8,200 years ago. The resulting glacial Lake Agassiz catastrophically broke through a kilometers-long ice dam at the bottom of the bay, suddenly dumping thousands of cubic kilometers of fresh water into the Atlantic Ocean.

Because this roughly coincided with the largest abrupt climate change recorded in the last 10,000 years, classic geoscience theory had assumed that the flood covered the surface of the Labrador and Greenland Seas and subpolar gyre with lighter, warmer, and less salty water to trigger the cooling. Scientists have assumed this slowed the Gulf Stream and ocean convection by disrupting the thermohaline circulation, that is, the large-scale temperature- and salt-driven ocean current responsible for our current warm, stable climate.



Cold ocean currents (blue) and warm ones (red) as they are believed to have existed 8,200 years ago. A new ocean current model is changing scientists' views on the impact of a freshwater flood in the North Atlantic on abrupt climate change observed in Europe at that time. *(Alan Condron)*

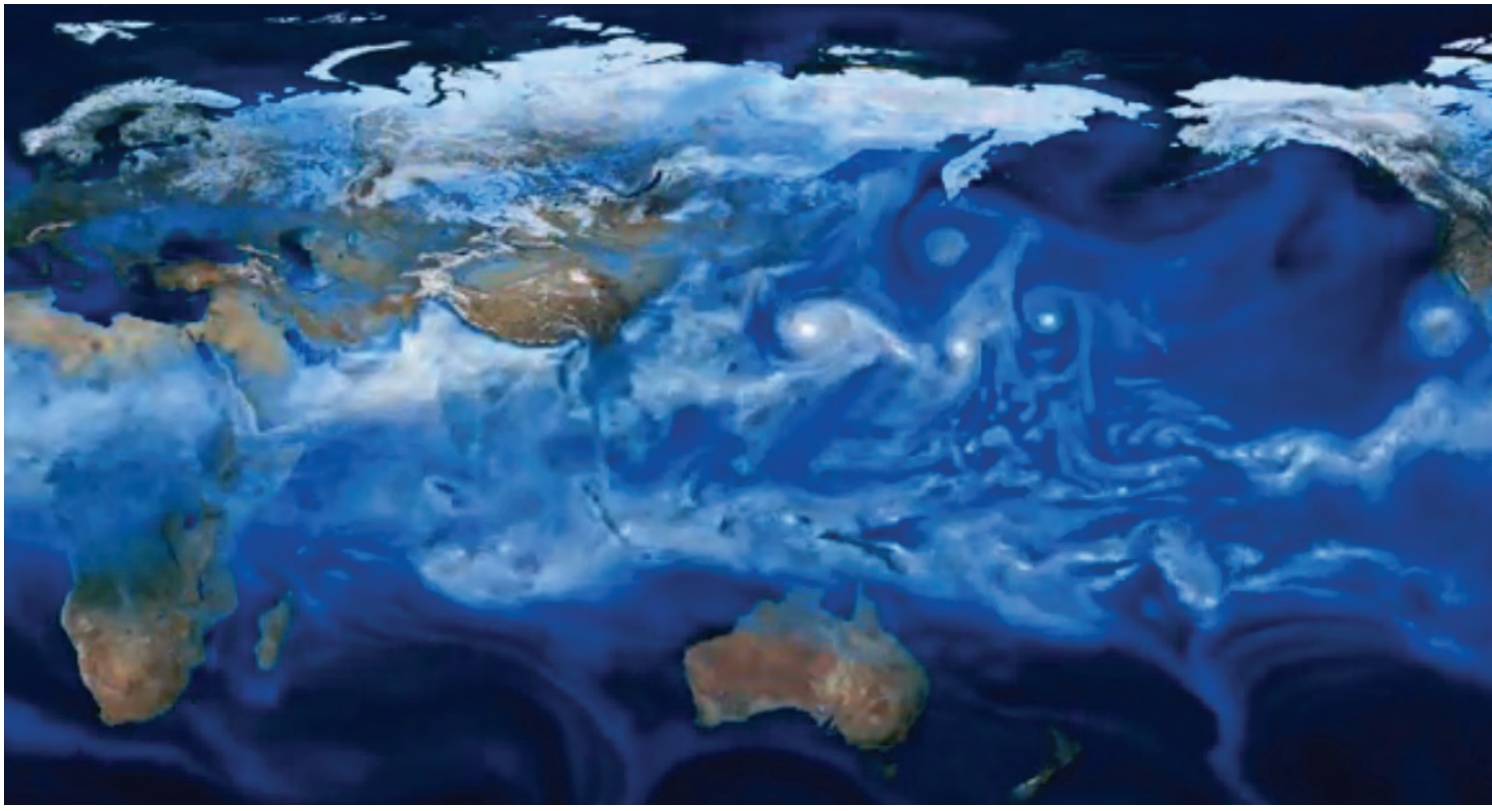
However, the new model, which has 10 to 20 times higher resolution than previous models, suggests the freshwater flood actually skirted past the subpolar gyre, instead flowing much farther south on the continental shelf as a narrow, buoyant, coastal current, eventually streaming into the subtropical North Atlantic. This finding is bolstered by evidence of sedimentary flood deposits along the coastal shelf, as well as the route taken by water flowing out of Hudson Bay today.

"Our results are particularly relevant for how we model the melting of the Greenland and Antarctic ice sheets now and in the future," says Condron. "It's apparent from our results that climate scientists are artificially introducing fresh water into their models much too far north, into parts of the ocean that it never would have reached. This has strong implications for predicting and understanding the stability of our future climate. We're in a similar interglacial period now, in the Holocene. The assumption is that the ocean looked and acted then much the same as it does today. So the coastal currents are assumed to be the same."

Their next step will be to understand the impact caused by this freshwater flooding into the subtropical, rather than the subpolar ocean, to see if it may be related to cooling and climate disruption.

Publication: Condron, A., and P. Winsor (2011), A subtropical fate awaited freshwater discharged from glacial Lake Agassiz, *Geophys. Res. Lett.*, 38, L03705, doi:10.1029/2010GL046011.

A Better Way to Find Extreme Weather Events in Climate Modeling Data



Project Name
High Performance Visualization, Analytics, and I/O

Project Leader
Edward (Wes) Bethel, Lawrence Berkeley National Laboratory

NERSC Resources
Hopper, 7,000 cores for 2 hours

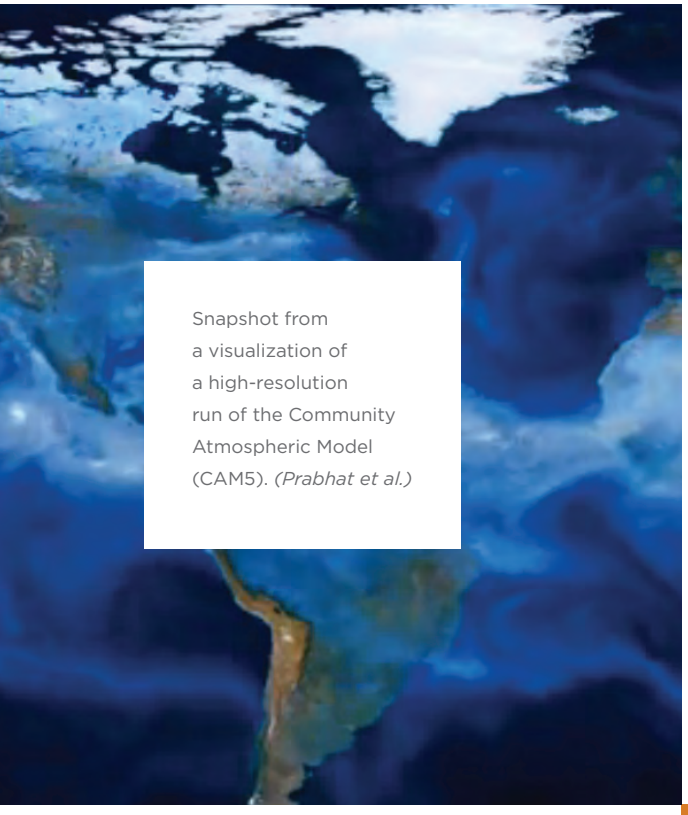
DOE Office
Advanced Scientific Computing Research

You'd think that spotting a category 5 hurricane would never be difficult. But when the hurricane is in a global climate model that spans several decades, it becomes a fleeting wisp among mountains of data.

That's a problem. As scientists develop ever more sophisticated computer models to predict the effects of climate change, one of the things they'll look for are changes in the frequency of extreme weather events such as hurricanes and heavy precipitation. The more data generated by models, however, the more difficult it is to quantify these rare but potent events.

What's needed is an automated way to quickly comb through a climate simulation's huge dataset and tally the weather events that spell big trouble. A team of researchers are developing techniques to do just that.

"We're using state-of-the-art methods in data mining and high performance computing to locate and quantify extreme weather phenomena in the very large datasets generated by today's climate models," says Prabhat, a scientific visualization expert in Berkeley Lab's Computational Research Division. "We want to help answer the question: How will climate change impact the frequency of extreme weather?"



Snapshot from a visualization of a high-resolution run of the Community Atmospheric Model (CAM5). *(Prabhat et al.)*

The research is led by fellow Berkeley Lab scientists Wes Bethel, Bill Collins, and Michael Wehner, and includes scientists from Oak Ridge National Laboratory, Los Alamos National Laboratory, Lawrence Livermore National Laboratory, and the University of California at Berkeley. Prabhat presented examples of their work at the annual meeting of the American Geophysical Union in December 2011.

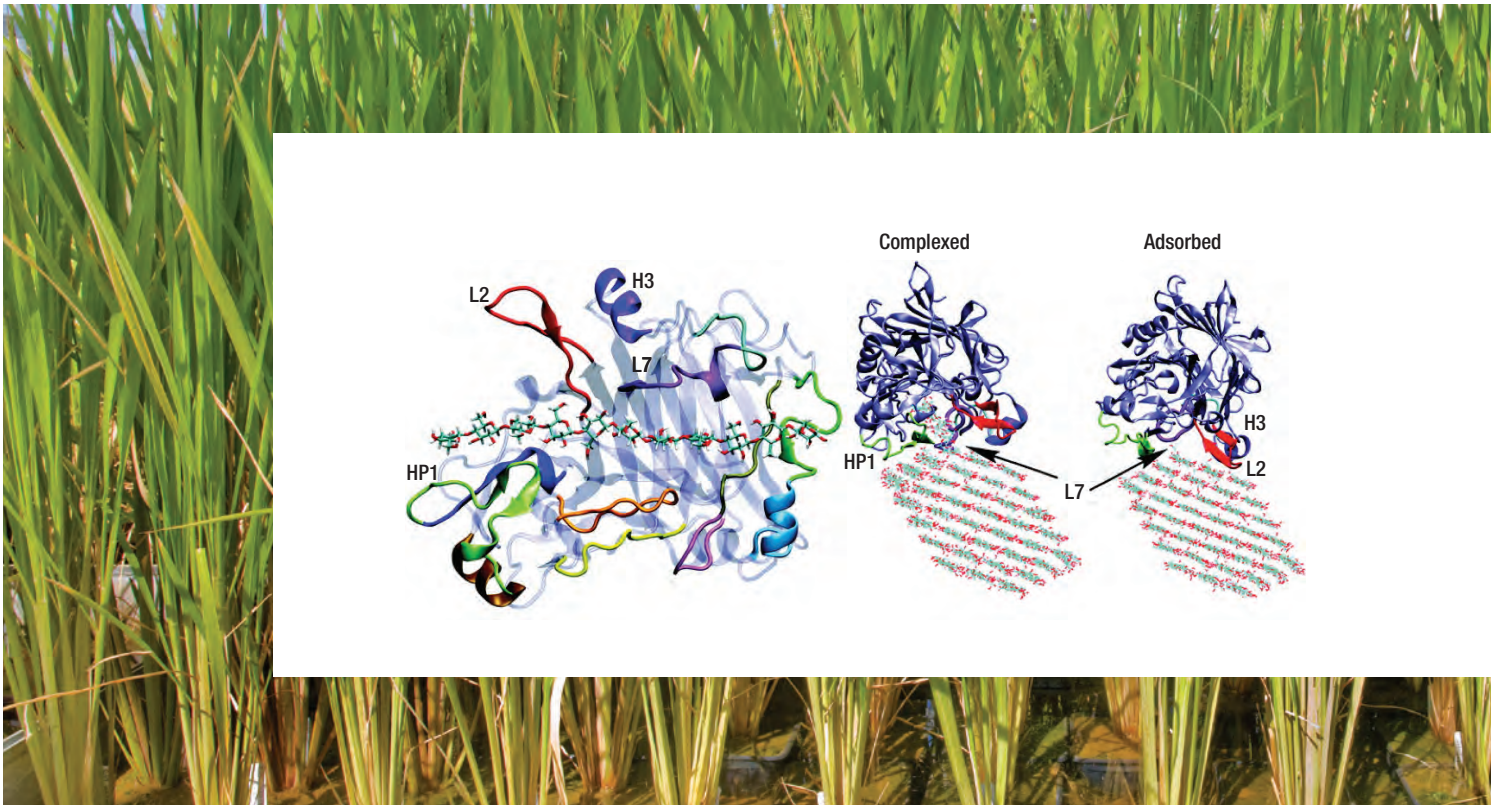
In one example, the scientists began with a recent run of the Community Atmospheric Model (CAM5), which is the latest in a series of global atmosphere models developed primarily at the National Center for Atmospheric Research. The high-resolution simulation spanned from 1979 to 2005 and is being used to demonstrate how well the model reproduces observed tropical cyclones. The 27-year simulation generated 100 terabytes of data.

To detect the hurricanes and cyclones in this simulation, the scientists used code from a model that tracks the intensity of hurricanes and cyclones, and adapted it to run in parallel on large datasets. In this case, the technique churned through the CAM5 simulation in two hours using 7,000 cores on NERSC's Hopper system. It successfully quantified the hurricanes that appear in the simulation.

Full story: http://www.nersc.gov/news-publications/news/science-news/2011/a-better-way-to-id-extreme-weather-events-in-climate-models/

Publication: Prabhat, Oliver Rübel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner, Wes Bethel, "TECA: A Parallel Toolkit for Extreme Climate Analysis," Procedia Computer Science, Volume 9, 2012, Pages 866–876, doi:10.1016/j.procs.2012.04.093.

Turning Grass into Gas for Less



Project Name
Mechanistic Studies of Enzyme-Based Processing of Biomaterials

Project Leader
Jhih-Wei Chu, University of California, Berkeley

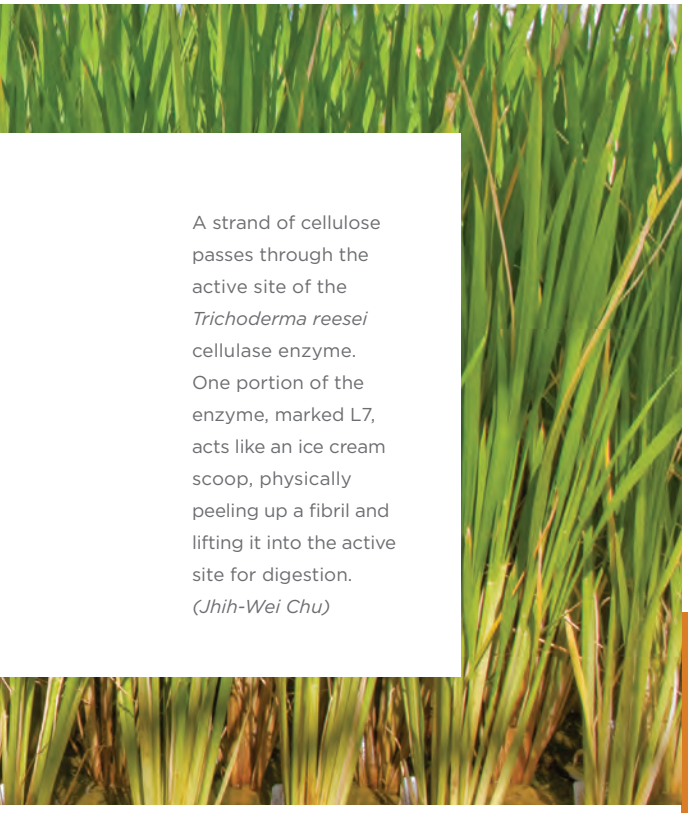
NERSC Resources
Hopper and Carver, 1 million hours

Other Compuring Resources
EMSL, NICS, TACC, UIUC

DOE Office
Basic Energy Sciences

If researchers could unlock the sugars stored in plant cellulose, then liquid biofuels could be made from hardy, high yielding, non-food crops like switchgrass, instead of the corn used to make ethanol. But what makes non-food crops indigestible to humans also makes them challenging raw materials for biofuel production. The sugars needed to make biofuels are locked up tight in cellulose, and researchers have yet to figure out an economical, scalable way to break them loose.

But computer simulations show how two different processes help unzip cellulose's structure, peeling off long strings of sugars. That information should help researchers engineer molecules that do the job better, cheaper, and on a larger scale, says Jhih-Wei Chu, a chemical and biomolecular engineering professor at the University of California, Berkeley, and a principal investigator with the Energy Biosciences Institute (EBI), a collaboration of the university, Berkeley Lab, and the energy company BP. Chu simulated two different approaches to cellulose decomposition: solvent effects and enzyme actions.



A strand of cellulose passes through the active site of the *Trichoderma reesei* cellulase enzyme. One portion of the enzyme, marked L7, acts like an ice cream scoop, physically peeling up a fibril and lifting it into the active site for digestion. *(Jhih-Wei Chu)*

In his first simulation, Chu and collaborators explored how the ionic liquid 1-butyl-3-methylimidazolium chloride (BmimCl) dissolves crystalline cellulose. They found that this solvent is very versatile, attacking a variety of bonds in the crystalline structure. "The Cl anions have strong attraction to hydroxol groups and weak attraction to side chains and sugar rings, while the cations have strong attractions to side chains and linker oxygens," says Chu. "This versatility can potentially be employed as the basis for molecular design and engineering."

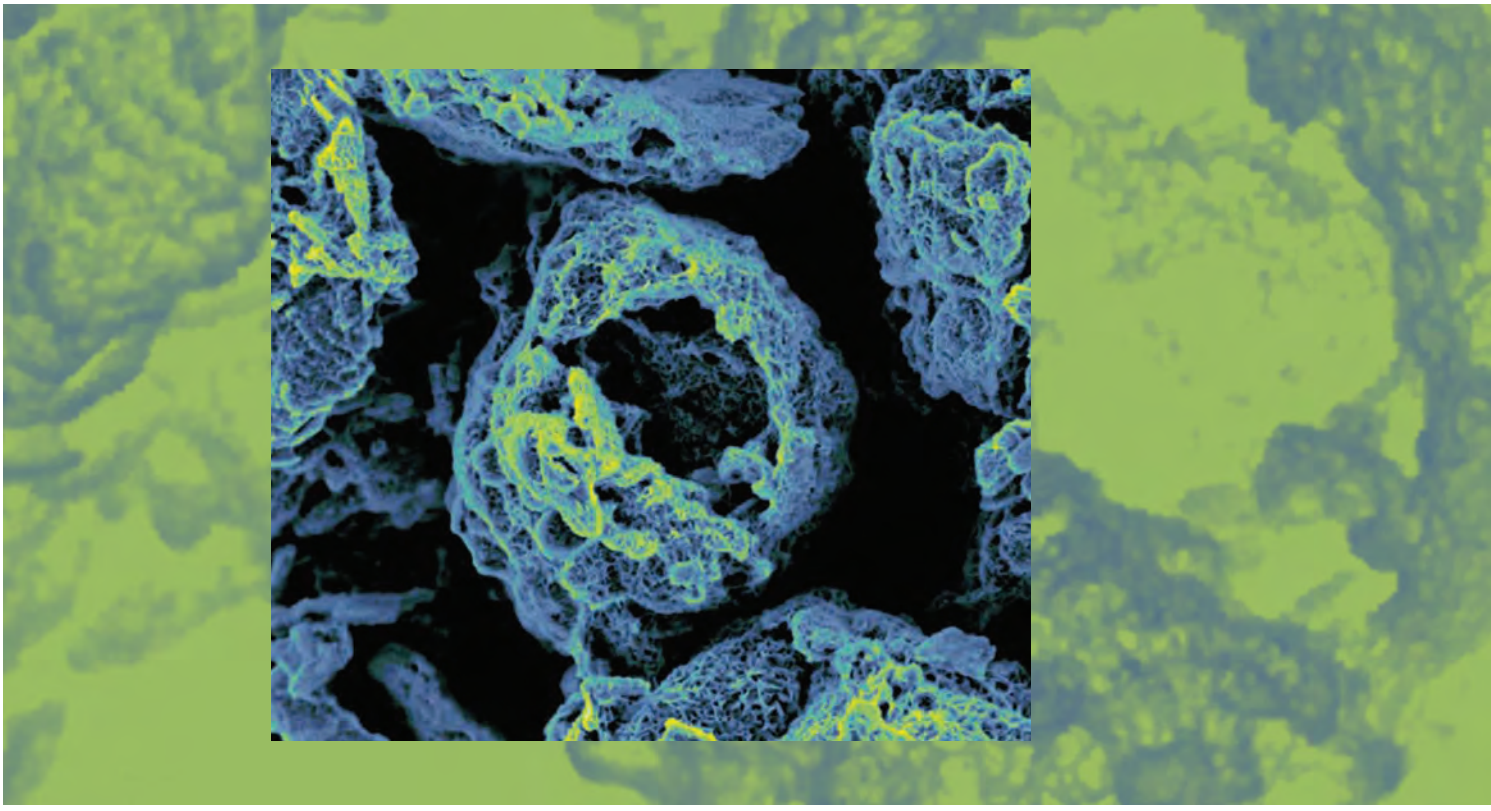
The second simulation involved cellulases, enzymes that can break down cellulose. One of the best sources of cellulases is the fungus *Trichoderma reesei*, discovered in the South Pacific during World War II, where it grew on and digested canvas tents. In the simulation, Chu and collaborators found that a portion of a *T. reesei* cellulase acts something like an ice cream scoop, physically peeling up microfibrils of cellulose and lifting them into the maw of the enzyme for digestion. The simulation hints at ways industrial enzymes already in use might be improved.

Full story: http://www.nersc.gov/news-publications/news/science-news/2011/turning-grass-into-gas-for-less/

Publications: "Dissecting Force Interactions in Cellulose Deconstruction Reveals the Required Solvent Versatility for Overcoming Biomass Recalcitrance," Hyung Min Cho, Adam S. Gross, and Jhih-Wei Chu, Journal of the American Chemical Society **2011** 133 (35), 14033–14041, doi:10.1021/ja2046155.

"Protein Allostery at the Solid–Liquid Interface: Endoglucanase Attachment to Cellulose Affects Glucan Clenching in the Binding Cleft," Yuchun Lin, Jordi Silvestre-Ryan, Michael E. Himmel, Michael F. Crowley, Gregg T. Beckham, and Jhih-Wei Chu, Journal of the American Chemical Society **2011** 133 (41), 16617–16624, doi:10.1021/ja206692g

Bubbles Help Break Energy Storage Record for Lithium–Air Batteries



Project Name
Computational Studies on the Reactivity, Reaction Mechanism, and Kinetics of Catalytic Systems

Project Leader
Donghai Mei, Pacific Northwest National Laboratory

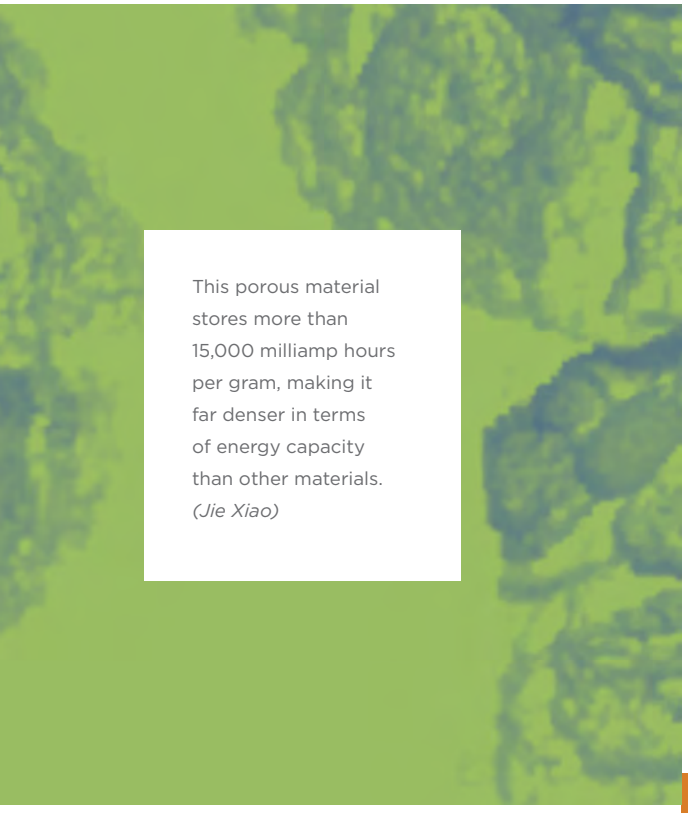
NERSC Resources
Hopper and Carver, up to 256 processors

DOE Office
Basic Energy Sciences

Resembling broken eggshells, graphene structures built around bubbles produced a lithium–air battery with the highest energy capacity to date, according to scientists at Pacific Northwest National Laboratory and Princeton University. This black, porous material could replace the traditional smooth graphene sheets in lithium–air batteries, which become clogged with tiny particles during use. As an added bonus, the team's new material does not rely on platinum or other precious metals, reducing its potential cost and environmental impact.

"This hierarchical structure of self-assembled graphene sheets is an ideal design not only for lithium–air batteries but also for many other potential energy applications," said Jie Xiao, the materials scientist at PNNL who led the study.

Lithium–air batteries could allow for the creation of long-range electric vehicles, able to travel up to 300 miles between charges. Comparatively lightweight, lithium–air batteries still suffer from limited practical capacity and poor cycle life issues. However, this study showed how to maximize the capacity of the batteries.



This porous material stores more than 15,000 milliamp hours per gram, making it far denser in terms of energy capacity than other materials. *(Jie Xiao)*

The team began by combining a binding agent with graphene, a special form of carbon. The binding agent dispersed the graphene in solution, just as soap disperses grease in dishwater. The graphene and binder were then added to water and mixed using a process that created bubbles inside the solution. The graphene and binder formed and hardened around the bubbles. When the bubbles eventually popped, hollow spheres of graphene were left behind. The tiny black particles are only 3 to 4 microns in diameter, 10 times smaller than a human hair.

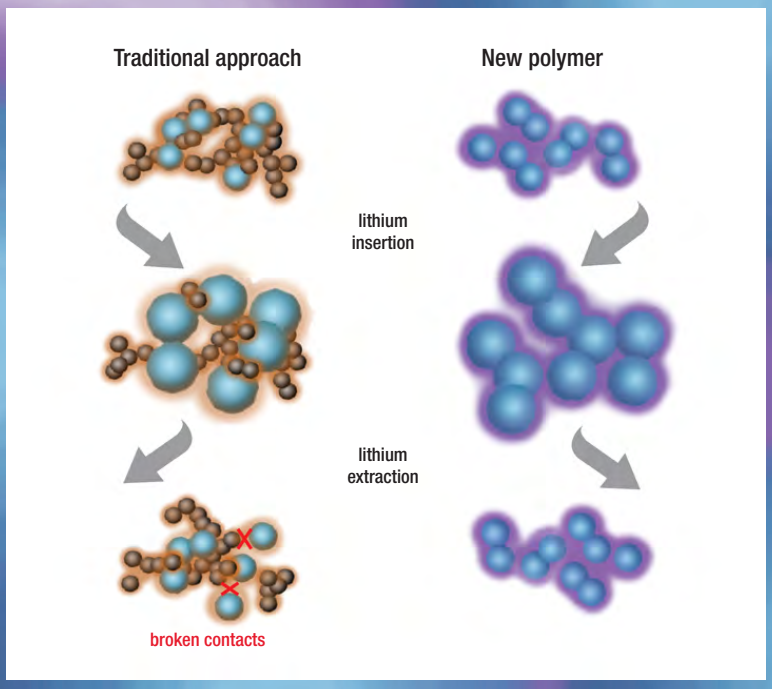
Using both modeling and microscopy, the scientists analyzed the graphene structures and their performance. They performed density functional theory calculations on the Hopper and Carver systems at NERSC, and studied the particles using electron microscopy at the Environmental Molecular Sciences Laboratory. The researchers found that the black porous structures store more than 15,000 milliamp hours per gram of graphene, making it far denser in term of energy capacity than other materials.

The battery is achieving the highest levels of energy capacity in an oxygen-only environment. When operated in ambient air, the capacity drops because the water in the air fouls the lithium metal in the batteries. The PNNL team is working to develop a membrane to block the water and still allow the necessary oxygen to flow. They are also working to make the battery rechargeable.

Full story: http://www.pnnl.gov/science/images/highlights/cmsd/energystorage.pdf

Publication: "Hierarchically Porous Graphene as a Lithium–Air Battery Electrode," Jie Xiao, Donghai Mei, Xiaolin Li, Wu Xu, Deyu Wang, Gordon L. Graff, Wendy D. Bennett, Zimin Nie, Laxmikant V. Saraf, Ilhan A. Aksay, Jun Liu, and Ji-Guang Zhang, Nano Letters **2011** 11 (11), 5071–5078, doi:10.1021/nl203332e.

New Anode Boosts Capacity of Lithium–Ion Batteries



Project Name
Large Scale Nanostructure Electronic Structure Calculations

Project Leader
Lin-Wang Wang, Lawrence Berkeley National Laboratory

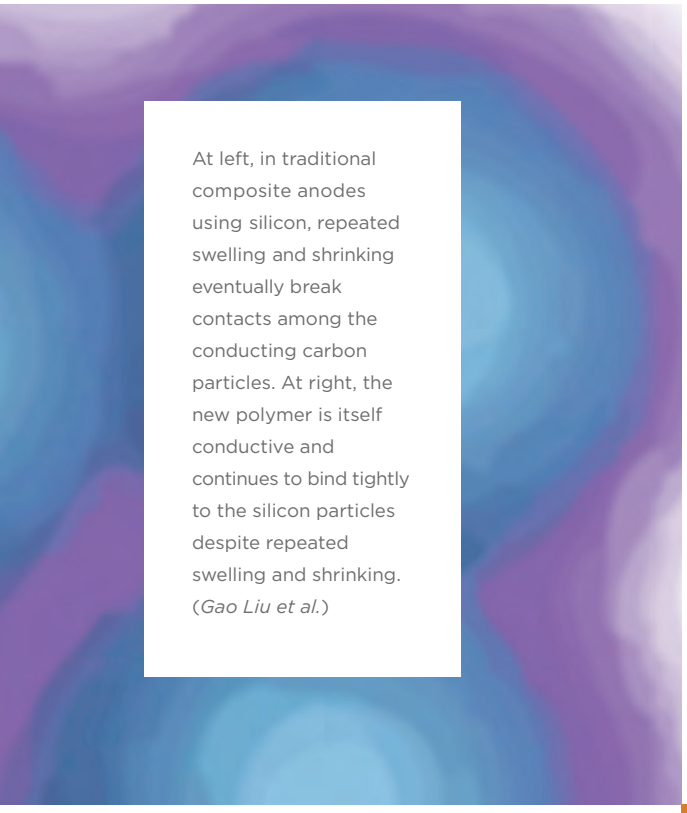
NERSC Resources
Franklin and Hopper

DOE Office
Basic Energy Sciences

Lithium-ion batteries are everywhere, in smart phones, laptops, an array of other consumer electronics, and the newest electric cars. But to lower the cost and extend the range of electric cars, batteries need to store a lot more energy.

A team of scientists at Berkeley Lab has designed a new kind of anode—a critical energy-storing component—capable of absorbing eight times the lithium of current designs. The new type anode has maintained its greatly increased energy capacity after over a year of testing and many hundreds of charge-discharge cycles.

The secret is a tailored polyfluorene-based polymer that conducts electricity and binds closely to lithium-storing silicon particles, even as they expand to more than three times their volume during charging and then shrink again during discharge. The new anodes are made from low-cost materials, compatible with standard lithium-battery manufacturing technologies.



At left, in traditional composite anodes using silicon, repeated swelling and shrinking eventually break contacts among the conducting carbon particles. At right, the new polymer is itself conductive and continues to bind tightly to the silicon particles despite repeated swelling and shrinking. (*Gao Liu et al.*)

Most of today's lithium-ion batteries have anodes made of graphite, which is electrically conducting and expands only modestly when housing the ions between its graphene layers. Silicon can store 10 times more—it has by far the highest capacity among lithium-ion storage materials—but it swells to more than three times its volume when fully charged, which quickly breaks the electrical contacts in the anode.

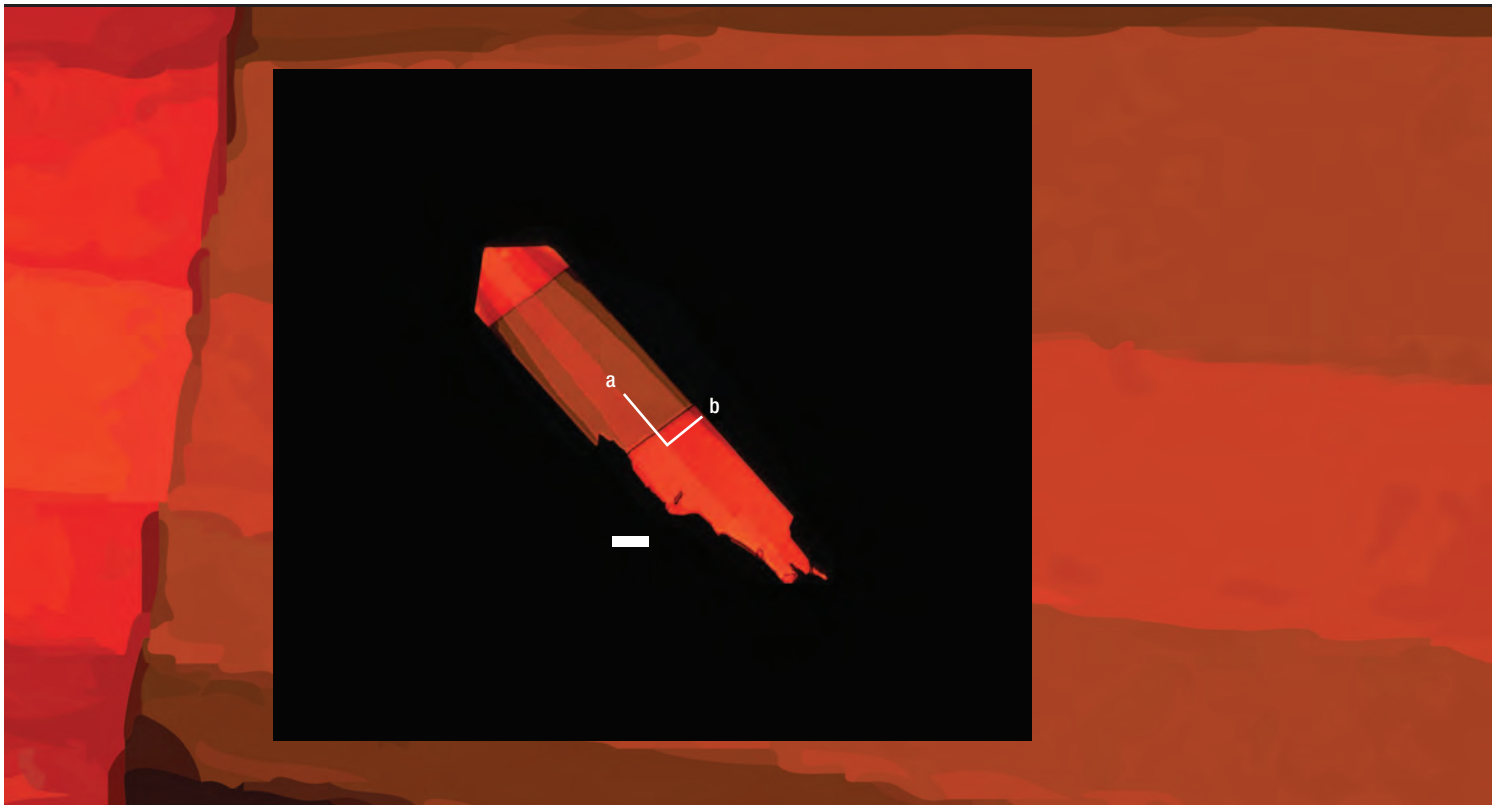
One solution has been to mix silicon particles in a flexible polymer binder, with carbon black added to the mix to conduct electricity. Unfortunately, the repeated swelling and shrinking of the silicon particles as they acquire and release lithium ions eventually push away the added carbon particles. What's needed is a flexible binder that can conduct electricity by itself, without the added carbon—and experiments show that polyfluorene-based polymers can meet those crieteria.

To understand the experimental results and optimize performance, the researchers used supercomputers at NERSC to run first-principles calculations of different versions of the promising polymers until they achieved the desired result. The simulations show that the lithium ions interact with the polymer first, and afterward bind to the silicon particles. When a lithium atom binds to the polymer through a carbonyl group, it gives its electron to the polymer—a doping process that significantly improves the polymer's electrical conductivity, facilitating electron and ion transport to the silicon particles.

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/a-better-lithium-ion-battery-on-the-way/

Publication: Liu, G., Xun, S., Vukmirovic, N., Song, X., Olalde-Velasco, P., Zheng, H., Battaglia, V. S., Wang, L. and Yang, W. (2011), Polymers with Tailored Electronic Structure for High Capacity Lithium Battery Electrodes. Adv. Mater., 23: 4679–4683, doi:10.1002/adma.201102421.

Speeding Up Design of Organic Semiconductors



Project Name
Computational Design of Novel Organic Photovoltaic Materials with Improved Hole and Electron-Transport Properties

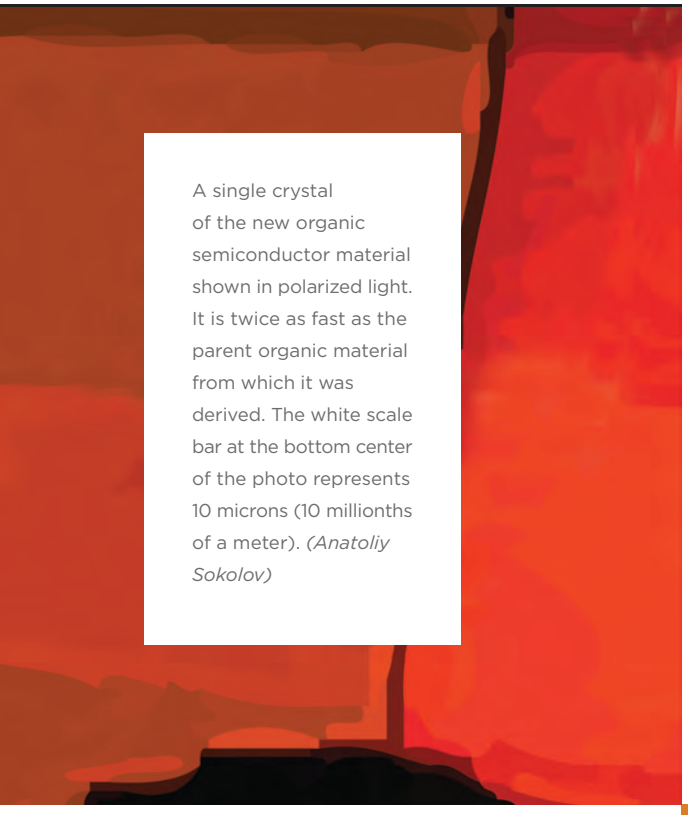
Project Leader
Alán Aspuru-Guzik, Harvard University

NERSC Resources
Carver, Franklin, Hopper

Other Computing Resources
Harvard University

If, one day, your iPad rolls up like a newspaper or solar panels are painted on your house, you can be grateful for progress in the design of organic semiconductors—carbon-based materials that move electrical charges from one place to another. Inorganic semiconductors, such as silicon, which are more widely used today, allow electrons to move quickly, but are rigid and expensive to grow in high purity. In contrast, today's organic semiconductors tend to have low electron mobility—they are not yet fast enough to drive high definition displays—but they can potentially be used on flexible substrates, processed cheaply, and even sprayed like ink or paint, opening a broad range of applications.

Now a team led by researchers at Stanford and Harvard universities has developed a new organic semiconductor material that is among the speediest yet. The scientists also accelerated the development process by using a predictive approach that lopped many months—and could lop years—off the typical timeline. Developing a new organic electronic material in the past has been a long, hit-or-miss process, requiring researchers to synthesize large numbers of candidate materials and then test them. The Stanford and Harvard-led group used computational screening to narrow the field of candidates before synthesizing the most promising material.



A single crystal of the new organic semiconductor material shown in polarized light. It is twice as fast as the parent organic material from which it was derived. The white scale bar at the bottom center of the photo represents 10 microns (10 millionths of a meter). *(Anatoliy Sokolov)*

The new organic semiconductor—predicted using a computer modeling program developed by Alán Aspuru-Guzik of Harvard and Joshua Schrier of Haverford College, and then synthesized by Zhenan Bao's research group at Stanford University—conducts charges much faster than the silicon used in most of today's display electronics. That means it could be used to make brighter displays that provide crisper video. And the new material is sufficiently speedy to make electronics for organic light-emitting diode (OLED) displays used in cell phones and televisions.

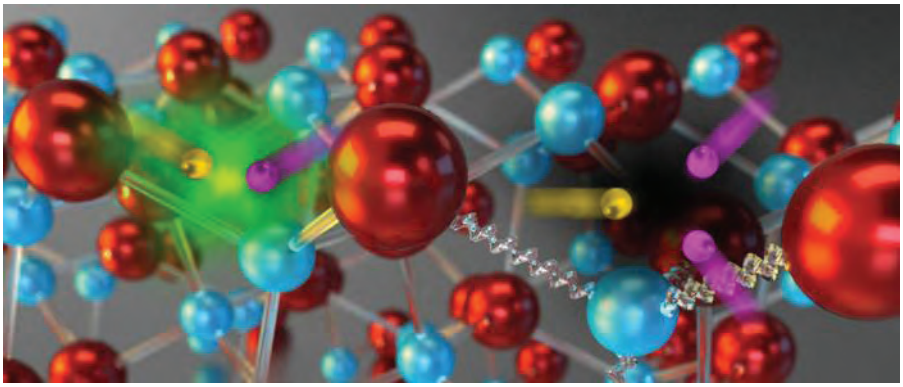
The researchers started with an organic semiconductor called DNTT, which has already demonstrated promise in the lab. First, the computer program generated several possible variations on this large carbon-based molecule by adding and subtracting components. The program then predicted how these variations would behave and screened for the most promising—those that seemed likeliest to conduct charges very fast.

When researchers in the Bao laboratory synthesized and tested the molecule predicted to have the best properties, it behaved as expected. Transistors made from the material operate 10 times faster than transistors made from amorphous silicon, the material used in today's display electronics; the new material is the second-fastest organic semiconductor yet made. The experiment is a proof of principle for using computers to develop new organic materials.

Publication: A.N. Sokolov et al., "From computational discovery to experimental characterization of a high hole mobility organic crystal," Nature Communications 2:437, doi:10.1038/ncomms1451 (2011).

DOE Office
Basic Energy Sciences

Solving the Mystery of LED "Droop"



Project Name
First-Principles Modeling of Nitride and Oxide Materials for Applications in Electronic and Optoelectronic Devices

Project Leader
Chris Van de Walle, University of California, Santa Barbara

NERSC Resources
Franklin and Hopper, up to 4,000 processors

Other Computing Resources
California NanoSystems Institute, NSF TeraGrid

Despite being cool, ultra-efficient, and long lasting, the light-emitting diode (LED) has yet to conquer the general lighting market due to a problem called "efficiency droop." New findings from simulations at NERSC have unearthed droop's elusive cause, researchers say, paving the way for wider LED use.

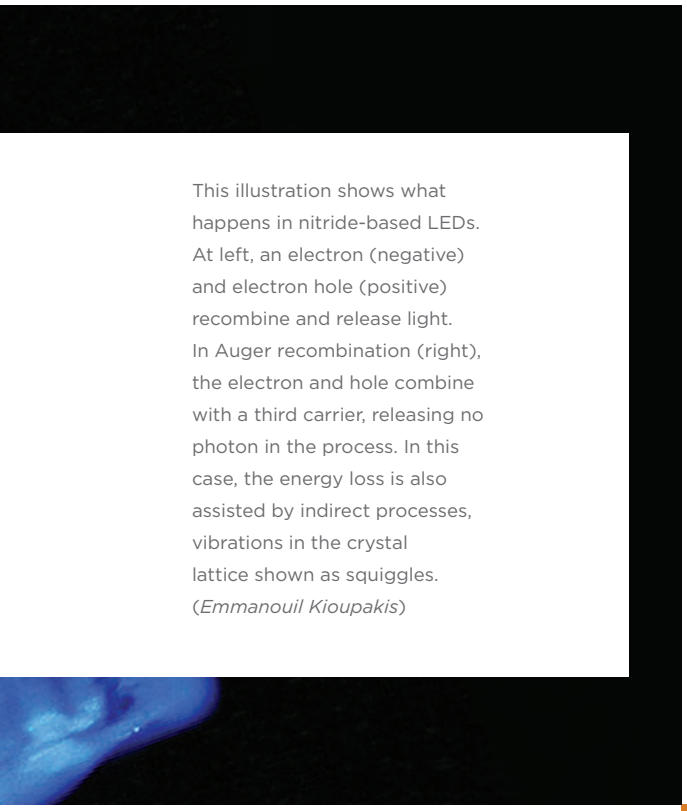
At low power, the nitride-based LEDs used to generate white light are astoundingly efficient; but pump up the current to room-illuminating levels and efficiency takes a nosedive. "If you use more electricity, you do get more light, but the efficiency isn't as good, and that adds to the higher cost of LED-based lighting," says Emmanouil Kioupakis, a researcher at the University of California, Santa Barbara (UCSB) and lead author of the study.

Theoretically, LED-based lights can produce up to 300 lumens per watt, about three times better than today's most energy-efficient alternative, compact fluorescent lights (CFLs). Because of droop, however, today's high-power LED lights perform about the same or slightly better than CFLs.

The new study does not solve the problem of droop, but it does set researchers and engineers on the right path to do so, says Chris Van de Walle, head of the UCSB research group that carried out the work: "Identifying the root cause of the problem is an indispensable first step toward devising solutions." The research relies on detailed, quantum-mechanical simulations of the inner workings of nitride-based LEDs to finger the culprit.

The exact cause of drooping photon production in nitride-based LEDs has been hotly debated between two main camps: One thinks the electrons become so energized that they overflow and escape the "quantum well" where trapped electrons emit photons. The other camp, to which Van de Walle's group belongs, thinks that the energy created by an electron/hole combination—physicists call it "recombination"—is being quietly swallowed by another charge carrier (a hole or an electron) inside the quantum well. This phenomenon is called Auger recombination.

Their results not only confirmed the importance of Auger recombination, but they also uncovered an unsuspected accomplice: carrier scattering. "This carrier scattering opens up additional recombination channels where the energy is eventually lost to heat," Kioupakis explains. In fact, the effects of this indirect Auger recombination were found to be about 1,000 times stronger than the direct process, the team found. This research suggests ways engineers might work around the problem.



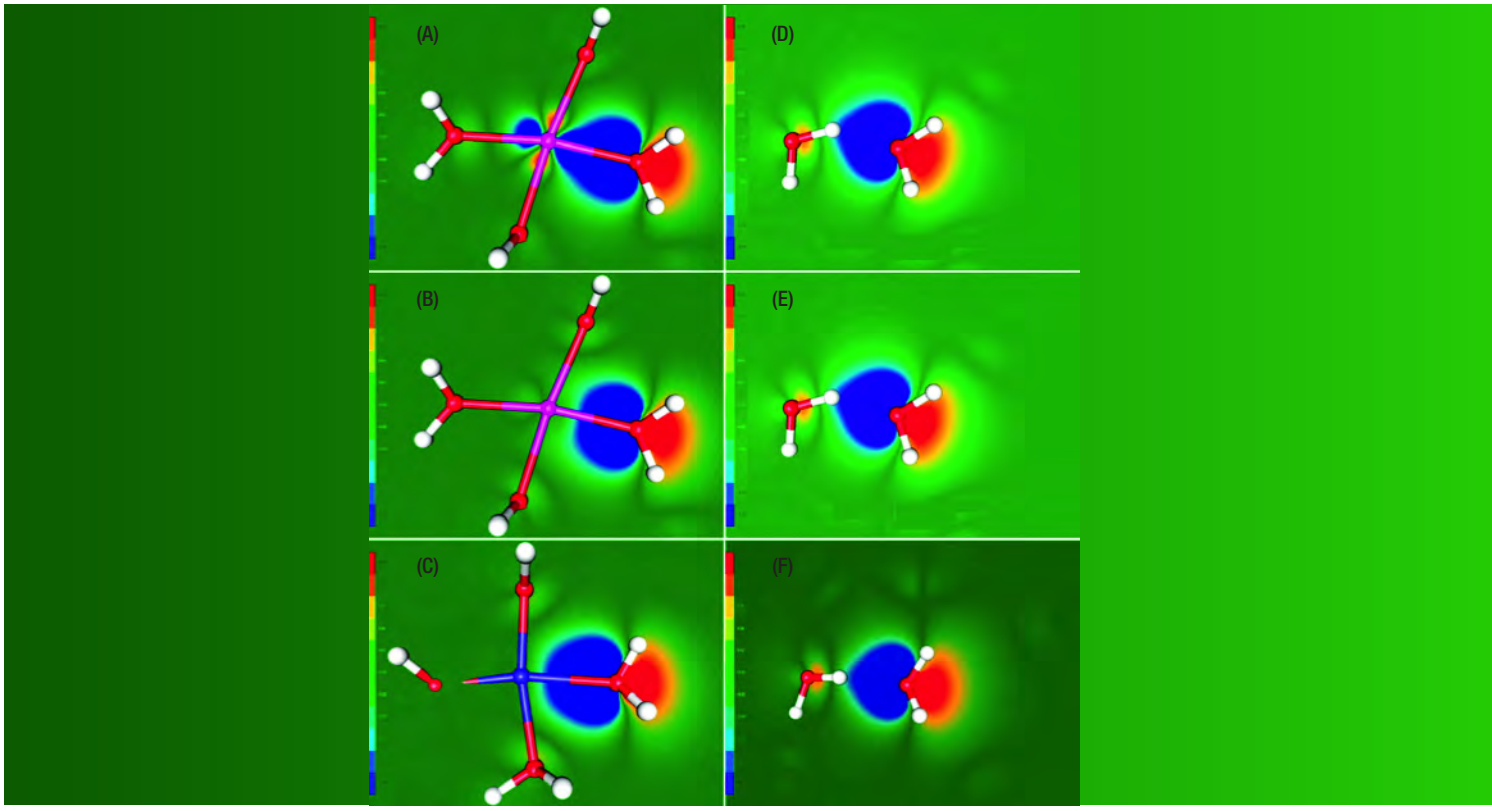
This illustration shows what happens in nitride-based LEDs. At left, an electron (negative) and electron hole (positive) recombine and release light. In Auger recombination (right), the electron and hole combine with a third carrier, releasing no photon in the process. In this case, the energy loss is also assisted by indirect processes, vibrations in the crystal lattice shown as squiggles. (*Emmanouil Kioupakis*)

DOE Office
Basic Energy Sciences

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/led-lighting-comes-out-of-the-dark/

Publication: Emmanouil Kioupakis, Patrick Rinke, Kris T. Delaney, and Chris G. Van de Walle, "Indirect Auger recombination as a cause of efficiency droop in nitride light-emitting diodes," Appl. Phys. Lett. **98**, 161107 (2011), doi:10.1063/1.3570656.

Modeling the Bonds of Iron and Water



Project Name
Computational Studies in Molecular Geochemistry

Project Leader
Andrew Felmy, Pacific Northwest National Laboratory

NERSC Resources
Franklin, Bassi

Other Computing Resources
EMSL

DOE Office
Basic Energy Sciences

Using supercomputers at NERSC, scientists have successfully modeled the atomic interactions between a high spin ferric iron ion and 64 water molecules for the first time. The model is based entirely on the fundamental equation of quantum mechanics, known as the Schrödinger equation, and uses no approximations. The simulation ran five times longer than any previous model and contained twice as many water molecules as preceding studies.

Understanding the structure and dynamics of hydration shells surrounding highly charged metal ions provides critical insights into many chemical processes. Researchers say this unprecedentedly accurate model will offer valuable insights into the process of underground mineral formation and extraction, as well as the transport of toxic materials in groundwater. By understanding these key reactions, scientists hope to develop effective strategies for carbon sequestration and other environmental remediation projects.

The study was carried out by the group of John Weare (Chemistry and Biochemistry Department, University of California, San Diego) in collaboration with researchers at Pacific Northwest National Laboratory. The project was led by Stuart Bogatko as part of his doctoral research.

"We can use quantum mechanical equations to gain valuable insights about Earth's chemistry by accurately modeling how metal ions and other materials react in water at different temperatures and pressures," says Eric Bylaska of the Pacific Northwest National Laboratory, co-author of the paper. "But until this study, accurate on-the-fly representations of these reactions were difficult to obtain."

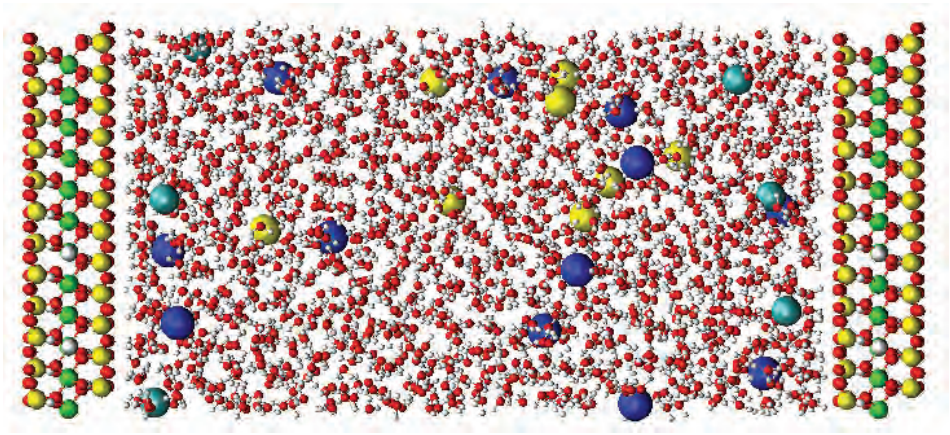
He notes that in geochemistry, conditions like pressure, temperature, and pH change quickly and constantly. The problem is further complicated by the fact that atoms in the molecular system are moving. The accuracy of Bogatko's work comes from solving the Schrödinger equation for all 517 electrons in the iron ion–water solution system. This ensures a parameter free model, meaning that scientists can successfully predict how the molecular system will react in a wide range of conditions on the fly. When implemented in the NWChem chemistry code, this technique accurately predicted the structure of iron ions in a water solution—including the motion of water molecules in the hydration shells, the interactions of electrons of the solvating molecules with the metal ion, the formation of ion pairs, and water exchange mechanisms.

Water-ion pair orbitals in the first (left) and second (right) hydration shells of high spin iron and aluminum. (*Bogatko et al.*)

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/modeling-the-bonds-of-iron-and-water/

Publication: "First Principles Simulation of the Bonding, Vibrational, and Electronic Properties of the Hydration Shells of the High-Spin Fe^{3+} Ion in Aqueous Solutions," Stuart A. Bogatko, Eric J. Bylaska, and John H. Weare, *The Journal of Physical Chemistry A* **2010** 114 (5), 2189–2200, doi:10.1021/jp904967n

Discovering the Molecular Details of CO_2 Reservoirs



Project Name
Clay Mineral Surface Geochemistry

Project Leader
Ian Bourg, Lawrence Berkeley National Laboratory

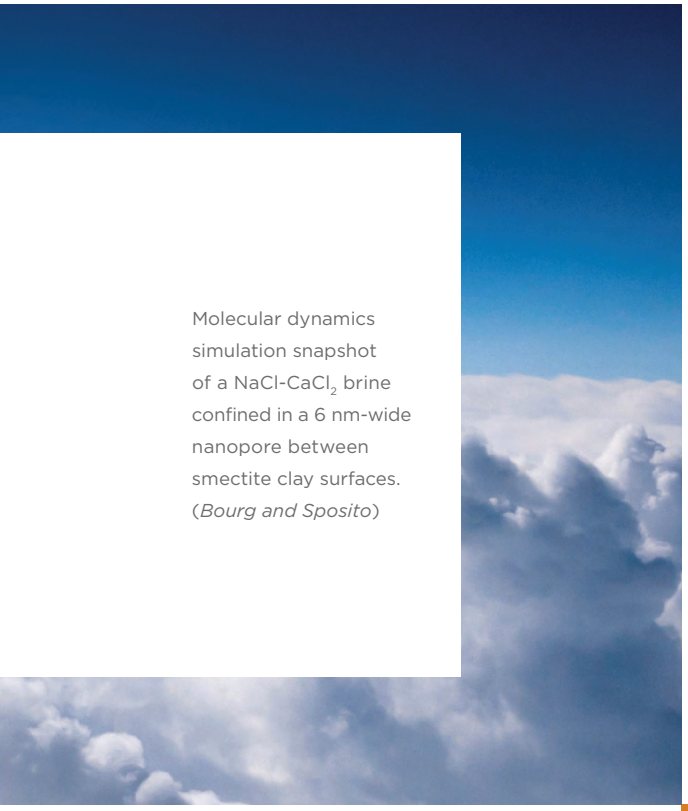
NERSC Resources
Franklin, Hopper, Carver

DOE Office
Basic Energy Sciences

Geologic carbon sequestration—trapping carbon dioxide (CO_2) from fossil fuel-burning power plants underground—is a key strategy for mitigating global warming; and saline aquifers are prime candidates for CO_2 storage. The brine that permeates porous rock formations deep underground is unusable for crops or drinking; but given enough time, it can lock away some CO_2 in the form of a weak acid. The rock lid that prevents brine from welling up to the surface (the caprock) traps injected CO_2.

Some carbon from CO_2 may also be transformed into carbonate minerals over hundreds or thousands of years. In the meantime, how can we be sure that CO_2 won't leak out?

The Center for Nanoscale Control of Geologic CO_2 (NCGC)—a collaborative effort led by Berkeley Lab under the auspices of DOE's Energy Frontiers Research Center (EFRC) program—is working to predict the performance and stability of carbon reservoirs by understanding of the behavior of aqueous solutions confined at high temperature and pressure in nanoscale rock pores. One NCGC team, led by Ian Bourg, is studying potential CO_2 leakage through clay caprocks by studying the behavior of water and ions near clay mineral surfaces.



Molecular dynamics simulation snapshot of a NaCl-$CaCl_2$ brine confined in a 6 nm-wide nanopore between smectite clay surfaces. (*Bourg and Sposito*)

The electrical double layer (EDL)—two parallel layers of charge surrounding an object in a liquid—plays a key role in ion adsorption at clay-water interfaces. However, despite decades of research, molecular-scale structure and dynamics in the EDL are not fully understood, especially at the high electrolyte concentrations of brine aquifers considered for CO_2 sequestration.

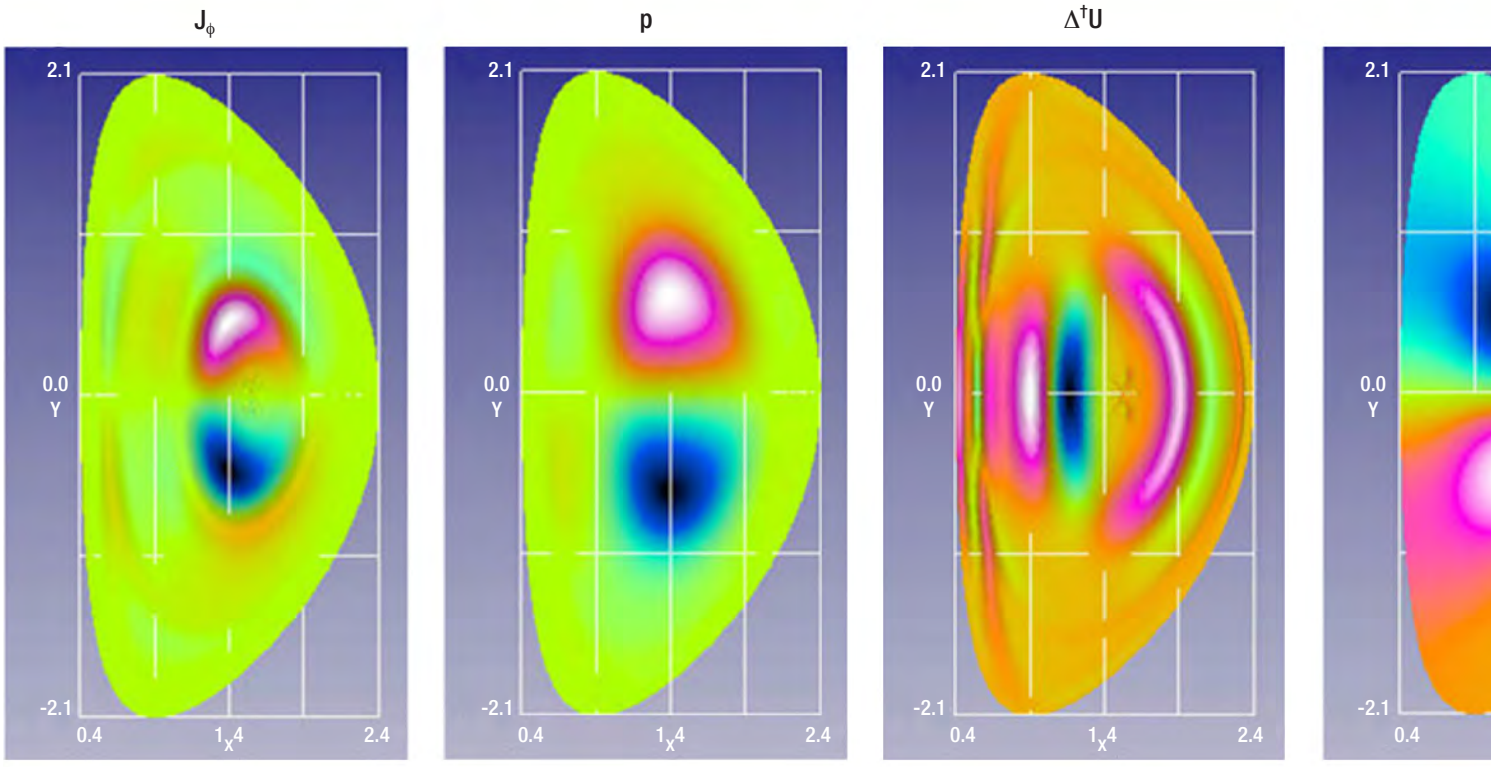
Bourg's team wanted to know if salinity influences the molecular-scale structure of the EDL, if charged surfaces modulate ion-ion association (and, eventually, CO_2 precipitation as carbonate minerals), and how fast water and solutes diffuse near clay-brine interfaces. They conducted molecular dynamics simulations of 6 nm-wide nanopores between smectite clay surfaces for a range of brine concentrations to determine the properties of the EDL as a function of salinity in a sodium chloride–calcium chloride brine.

The simulations found that anion exclusion balances a significant fraction of the clay structural charge at high ionic strengths. This may shed light on the ionic-strength-dependence of the clay–water interfacial energy and of clay–water–CO_2 wetting angles.

This study—unprecedented in detail—will help set the parameters for larger-scale reactive transport models of clay caprocks, which will in turn be used to predict the stability of CO_2 reservoirs in saline aquifers.

Publication: Ian C. Bourg and Garrison Sposito, "Molecular dynamics simulations of the electrical double layer on smectite surfaces contacting concentrated mixed electrolyte (NaCl–$CaCl_2$) solutions," Journal of Colloid and Interface Science 360 (2011) 701–715, doi:10.1016/j.jcis.2011.04.063.

Simulating Magnetic Instabilities in NSTX and ITER

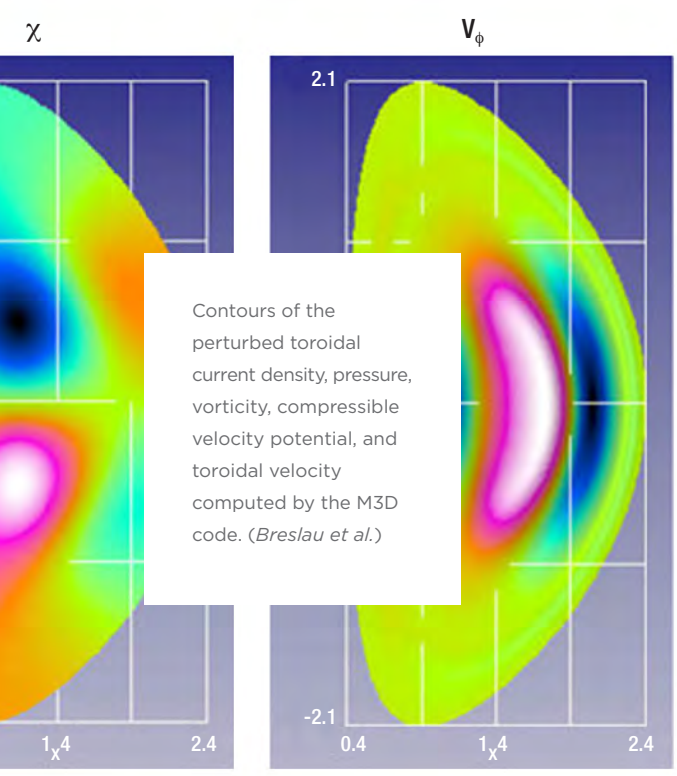


Project Name
3D Extended MHD Simulation of Fusion Plasmas

Project Leader
Stephen Jardin, Princeton Plasma Physics Laboratory

NERSC Resources
Franklin, up to 1,024 cores

DOE Office
Fusion Energy Sciences



Contours of the perturbed toroidal current density, pressure, vorticity, compressible velocity potential, and toroidal velocity computed by the M3D code. (*Breslau et al.*)

Controlled nuclear fusion has long held out the promise of clean, essentially limitless energy. Ultra-hot ionized gases are confined within a magnetic field at high pressures until the atoms fuse, releasing energy, much like the sun does. Learning how to control the stability of fusion plasma is a crucial step toward making fusion energy a reality.

Magnetohydrodynamics (MHD) is a physical model that can be used to analyze plasma instabilities. MHD stability at high beta—a measure of plasma pressure—is essential for a cost-effective fusion reactor. Small changes can trigger large effects if the plasma is in an unstable equilibrium.

One of the more elusive types of instability is the neoclassical tearing mode (NTM). NTMs tear the magnetic field lines, producing magnetic islands with lower energy levels and breaking up the magnetic surfaces that confine the plasma.

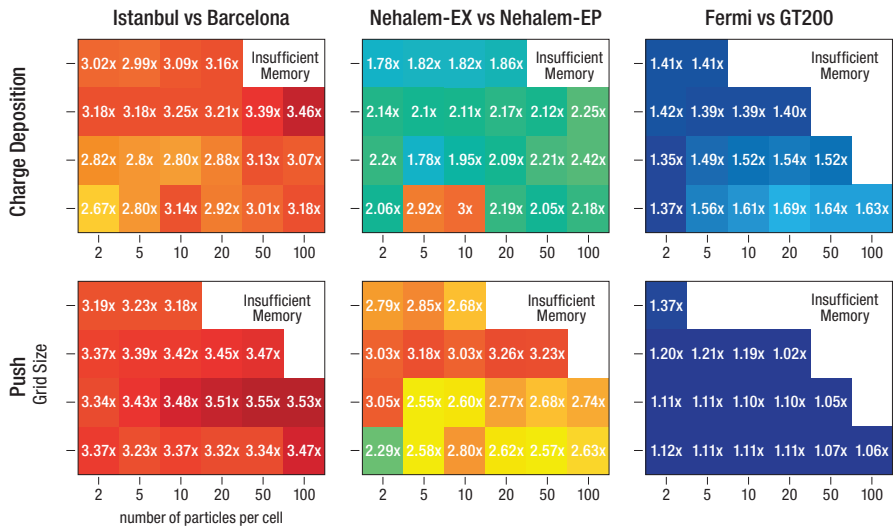
In the National Spherical Torus Experiment (NSTX) at Princeton Plasma Physics Laboratory (PPPL), NTMs have been observed to exert a braking effect on plasma rotation, leading to plasma disruption. In a disruption, the plasma can lose all its energy within milliseconds. As higher-energy experiments like ITER, the international fusion energy experiment, go online, disruptions could cause structural damage to the reactor.

Understanding what causes NTMs will help in avoiding or mitigating these instabilities. NTM triggering events can often be found in diagnostic data. But in a particular subset of NSTX runs, the NTMs seemed to appear with no apparent trigger. A team of PPPL researchers led by Joshua Breslau did a detailed computational analysis of one of these runs using several MHD codes. Their simulations not only revealed the conditions that created the NTM, but they even reproduced the onset time to within milliseconds.

It is not yet clear whether such a "triggerless" NTM could happen at ITER, but the PPPL team concluded, "the predictive capability demonstrated for NSTX gives us some confidence that our model is capable of predicting the onset of these modes in ITER, should they exist."

Publication: J.A. Breslau et al., "Onset and saturation of a non-resonant internal mode in NSTX and implications for AT modes in ITER," 2011 Nucl. Fusion **51** 063027, doi:10.1088/0029-5515/51/6/063027.

Fusion PIC Code Optimization on Emerging Architectures



Project Name
Scientific Application Performance on Candidate Petascale Platforms

Project Leader
Leonid Oliker,
Lawrence Berkeley National Laboratory

NERSC Resources
Dirac

Other Computing Resources
UC Berkeley

The next decade of high performance computing (HPC) systems will see a rapid evolution of node architectures as power and cooling constraints limit increases in microprocessor clock speeds. In response to this trend, computer architects are dramatically increasing on-chip parallelism to improve performance. During the next decade, the level of parallelism on a single microprocessor will rival the number of nodes in the most massively parallel supercomputers of the 1980s. By 2020, extreme-scale HPC systems are anticipated to have on the order of 100,000–1,000,000 sockets, with each socket containing between 100 and 1,000 potentially heterogeneous cores.

Assessing the tradeoffs of differing supercomputing designs using real large-scale simulations is a key step towards making effective use of vastly complex and expensive HPC systems. The project "Scientific Application Performance on Candidate Petascale Platforms," led by Leonid Oliker of Berkeley Lab, meets this challenge by running full applications with real input data, at the scale desired by application scientists, on today's leading-edge architectures.

In one important study, Kamesh Madduri and collaborators at Berkeley Lab, Princeton Plasma Physics Lab, and Kookmin University in South Korea optimized a fusion energy application—GTC, a petascale gyrokinetic toroidal fusion code for studying plasma microturbulence—to take advantage of four modern multi-core architectures.

For GTC's key computational components (charge deposition and particle push), the team explored efficient parallelization strategies across a broad range of emerging multicore designs, including the Intel Nehalem-EX, the AMD Opteron Istanbul, and the highly multithreaded Sun UltraSparc T2+. They also performed the first study on tuning gyrokinetic particle-in-cell (PIC) algorithms for graphics processors, using the NVIDIA C2050 (Fermi). Their paper discusses several novel optimization approaches for gyrokinetic PIC, including mixed-precision computation, particle binning and decomposition strategies, grid replication, SIMDized atomic floating-point operations, and effective GPU texture memory utilization.

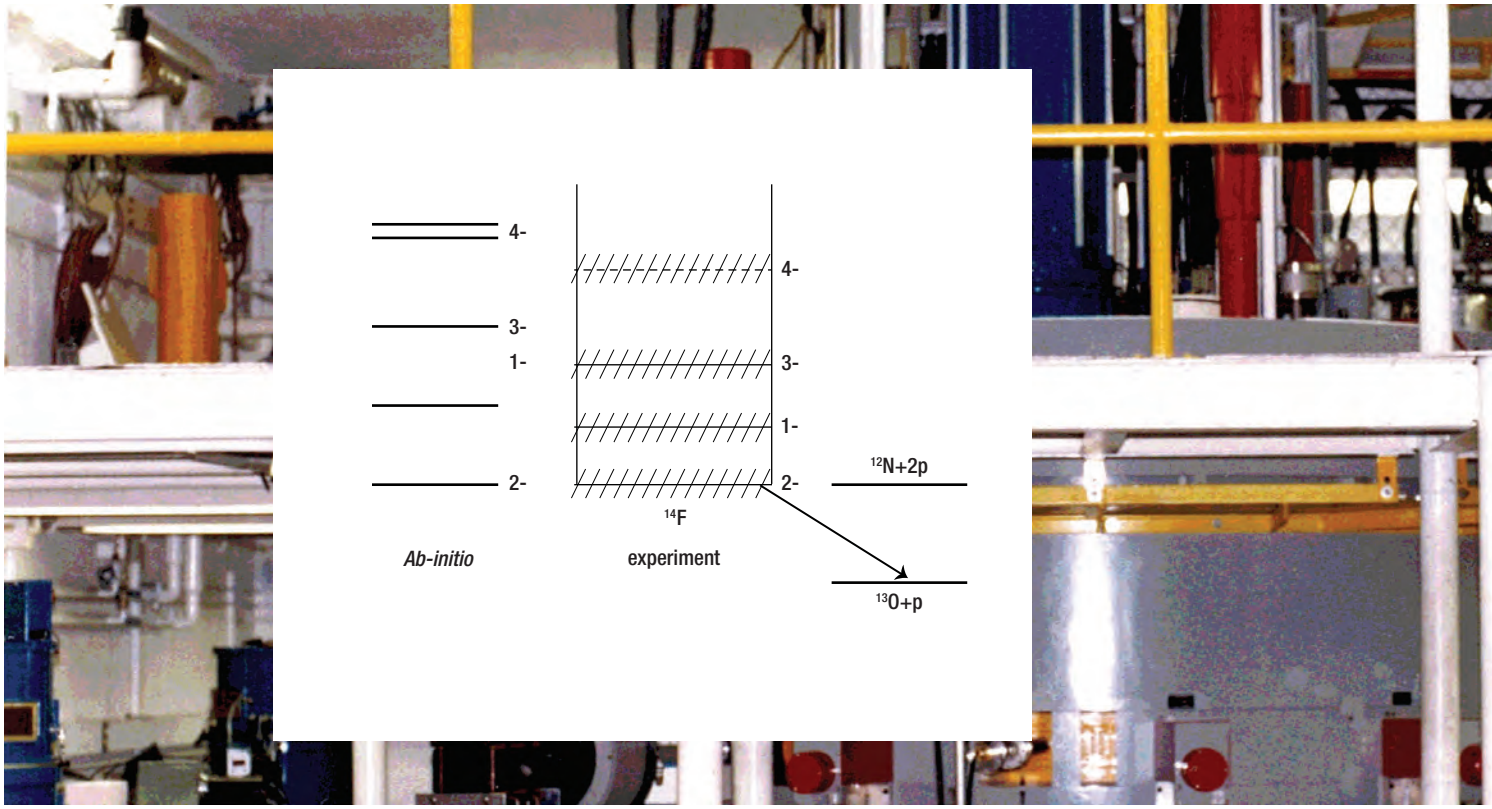
Overall, the researchers achieved significant performance improvements of 1.3–4.7x on these complex PIC kernels, despite the inherent challenges of data dependency and locality. Their work points to several architectural and programming features that could significantly enhance PIC performance and productivity on next-generation architectures.

Performance ratio of Istanbul vs. Barcelona SMPs (left column), Nehalem-EX vs. Nehalem-EP SMPs (middle column), and Fermi vs. GT200 GPUs (right column) for the best implementations of charge deposition (top row) and push (bottom row). (*Madduri et al.*)

DOE Office
Advanced Scientific Computing Research

Publication: Kamesh Madduri, Eun-Jin Im, Khaled Z. Ibrahim, Samuel Williams, Stéphane Ethier, Leonid Oliker, "Gyrokinetic particle-in-cell optimization on emerging multi- and manycore platforms," Parallel Computing, Vol. 37, Issue 9, Sept. 2011, pp. 501–520, doi:10.1016/j.parco.2011.02.001.

Proton Dripping Tests a Basic Force of Nature



Project Name
Structure and Reactions of Hadrons and Nuclei; Non-Perturbative Solutions of Strongly Interacting Particles and Fields

Project Leader
James Vary, Iowa State University

NERSC Resources
Franklin, up to 23,980 cores

Other Computing Resources
OLCF

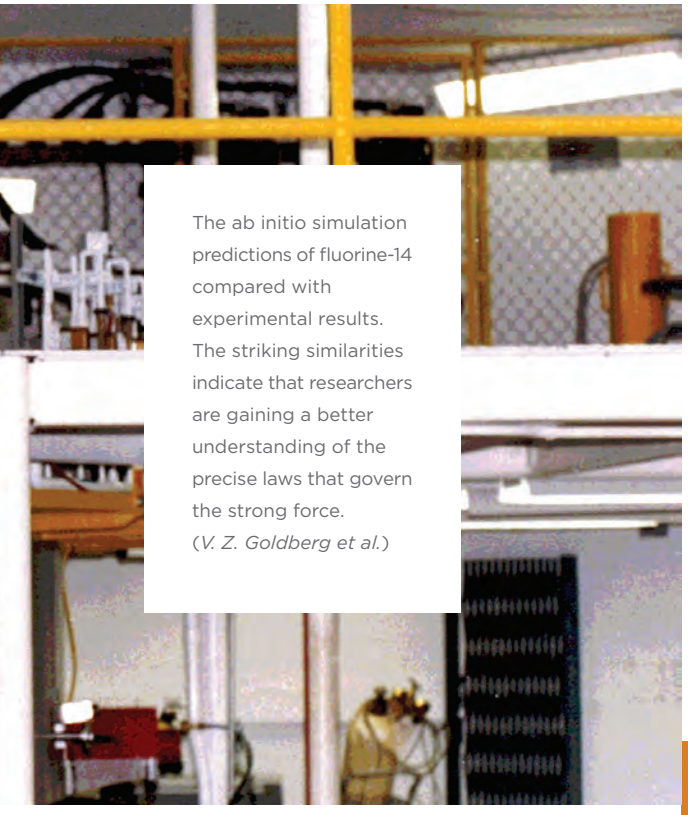
Like gravity, the strong interaction is a fundamental force of nature. It is the essential "glue" that holds together atomic nuclei (composed of protons and neutrons) to form atoms, the building blocks of nearly all the visible matter in the universe. Despite its prevalence in nature, researchers are still searching for the precise laws that govern the strong force. However, the recent discovery of an extremely exotic, short-lived nucleus called fluorine-14 in laboratory experiments may indicate that scientists are gaining a better grasp of these rules.

Fluorine-14 is composed of nine protons and five neutrons. It exists for a tiny fraction of a second before a proton "drips" off, leaving an oxygen-13 nucleus behind. A team of researchers led by James Vary, a professor of physics at Iowa State University, first predicted the properties of fluorine-14 with the help of scientists in Berkeley Lab's Computational Research Division, as well as supercomputers at NERSC and the Oak Ridge Leadership Computing Facility. These fundamental predictions served as motivations for experiments conducted by Vladilen Goldberg's team at Texas A&M's Cyclotron Institute, which achieved the first observation of fluorine-14.

Researchers have so far discovered about 3,000 nuclei in laboratory experiments and suspect that 6,000 more could still be created and studied. Understanding the properties of these nuclei will give researchers insights into the strong force, which could in turn be applied to develop and improve future energy sources.

With these goals in mind, the DOE SciDAC program brought together teams of theoretical physicists, applied mathematicians, computer scientists, and students from universities and national laboratories to create a computational project called the Universal Nuclear Energy Density Functional (UNEDF), which uses supercomputers to predict and understand the behavior of a wide range of nuclei, including their reactions, and to quantify uncertainties.

Fluorine-14 was simulated with a code called Many Fermion Dynamics–nuclear (MFDn) that was developed on NERSC systems over the past two decades and optimized by the UNEDF team. The UNEDF contributions improved code performance by a factor of three and increased the mathematical precision. Without these improvements, the prediction of fluorine-14 would not have been possible.



The ab initio simulation predictions of fluorine-14 compared with experimental results. The striking similarities indicate that researchers are gaining a better understanding of the precise laws that govern the strong force. (*V. Z. Goldberg et al.*)

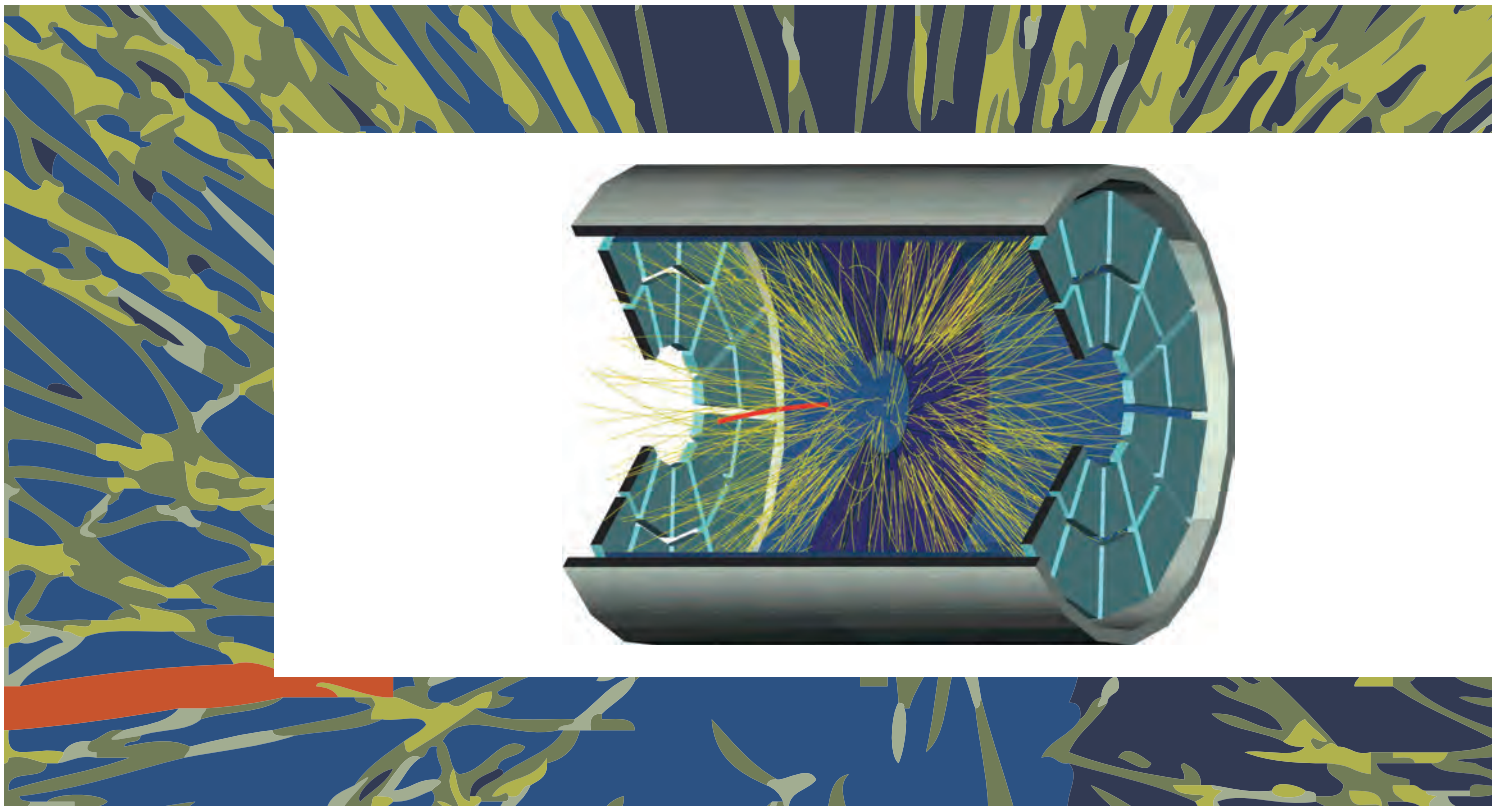
DOE Office
Nuclear Physics

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/proton-dripping-tests-a-basic-force-of-nature/

Publication: P. Maris, A. M. Shirokov, and J. P. Vary, "*Ab initio* nuclear structure simulations: The speculative ^{14}F nucleus," Physical Review C **81**, 021301(R) (2010), doi:10.1103/PhysRevC.81.021301.

V. Z. Goldberg et al., "First observation of ^{14}F," Physics Letters B **692**, 307–311 (2010), doi:10.1016/j.physletb.2010.07.054.

Heaviest Antimatter Particle Detected with NERSC Help



Project Name
STAR Detector Simulations and Data Analysis

Project Leader
Grazyna Odyniec, Lawrence Berkeley National Laboratory

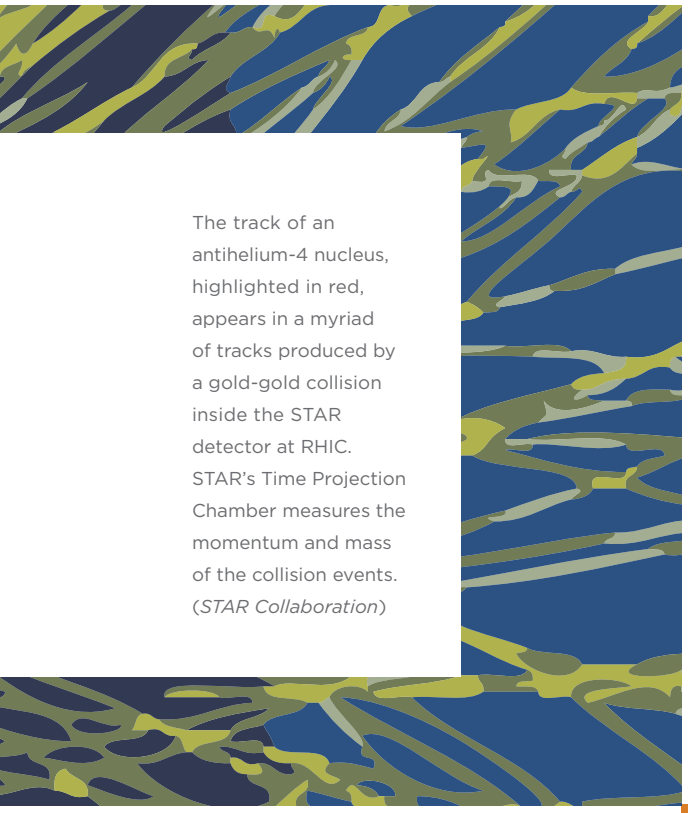
NERSC Resources
PDSF

DOE Office
High Energy Physics

Eighteen examples of the heaviest antiparticle ever found, the nucleus of antihelium-4, have been made in the STAR experiment at RHIC, the Relativistic Heavy Ion Collider at Brookhaven National Laboratory. The finding was not unexpected, but it is a milestone for scientists exploring a fundamental puzzle of physics: Why is there any matter at all? The next possible heavyweight antimatter particle would be produced under such rare circumstances that this record is likely to stand for years to come.

Collisions of energetic gold nuclei inside STAR briefly recreate conditions in the hot, dense early universe only millionths of a second after the Big Bang. Since equal amounts of matter and antimatter were created in the Big Bang, they should have completely annihilated one another; but for reasons still not understood, only ordinary matter seems to have survived. Today this excess matter forms the visible universe we know.

Roughly equal amounts of matter and antimatter are also produced in heavy-ion (gold nuclei) collisions at RHIC. The resulting fireballs expand and cool quickly, so the antimatter can avoid annihilation long enough to be detected in the Time Projection Chamber (TPC) at the heart of STAR.



The track of an antihelium-4 nucleus, highlighted in red, appears in a myriad of tracks produced by a gold-gold collision inside the STAR detector at RHIC. STAR's Time Projection Chamber measures the momentum and mass of the collision events. (*STAR Collaboration*)

"The STAR experiment is uniquely capable of finding antihelium 4," says the STAR experiment's spokesperson, Nu Xu, of the Nuclear Science Division (NSD) at Berkeley Lab. "STAR already holds the record for massive antiparticles, last year [2010] having identified the anti-hypertriton, which contains three constituent antiparticles. With four antinucleons, antihelium-4 is produced at a rate a thousand times lower yet. To identify the 18 examples required sifting through the debris of a billion gold-gold collisions."

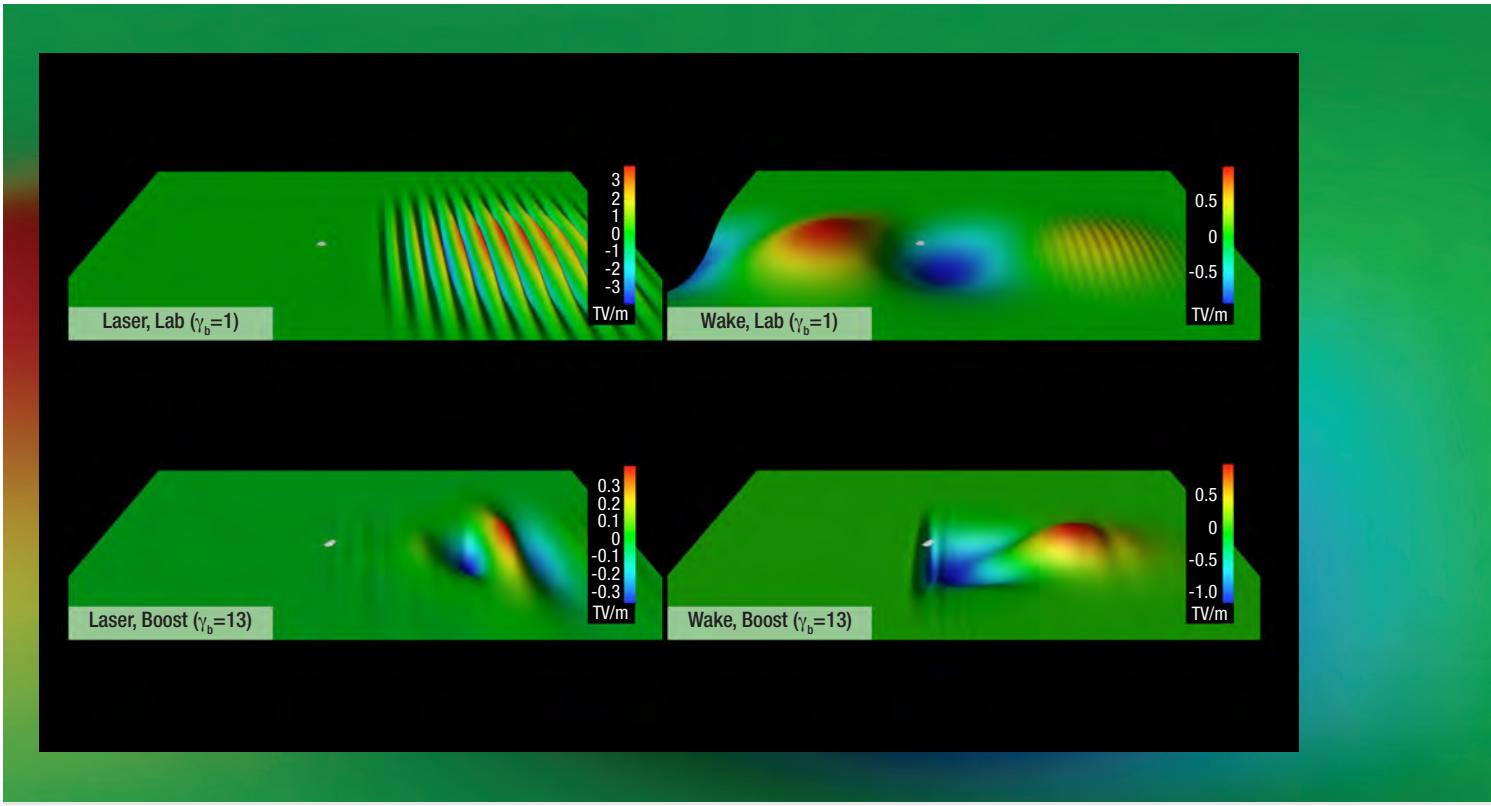
That sifting was done on the PDSF high energy physics computing cluster hosted at NERSC, using 1 petabyte of data stored on NERSC's HPSS system. "STAR uses PDSF extensively for analysis and simulation essential to discoveries such as the anti-alpha [antihelium-4] that require good understanding of the detector response," says Hank Crawford, a STAR collaborator at Berkeley Lab's NSD and the University of California Berkeley's Space Sciences Laboratory. "Our signature for the antihelium-4 is based on finding 'fat' tracks in the time-projection chamber at STAR's center. Understanding these tracks and the TPC response is dependent on PDSF resources."

The scientists measured the antihelium-4 production rate in nuclear interactions, and found that it is consistent with expectations based on a statistical coalescence of antiquarks from the soup of quarks and antiquarks generated in RHIC collisions. Knowing the production rate of these antinuclei is important to a wide range of scientific disciplines, including searches for new phenomena in the cosmos.

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/heaviest-antimatter-particle-detected-with-nersc-help/

Publication: The STAR Collaboration, "Observation of the antimatter helium-4 nucleus," Nature **473**, 353–356 (19 May 2011), doi:10.1038/nature10079.

Boosting the Next Wave of Accelerators



Project Name
Simulation of Laser-Plasma Particle Accelerators

Project Leader
Cameron Geddes, Lawrence Berkeley National Laboratory

NERSC Resources
Franklin, 2,016 cores

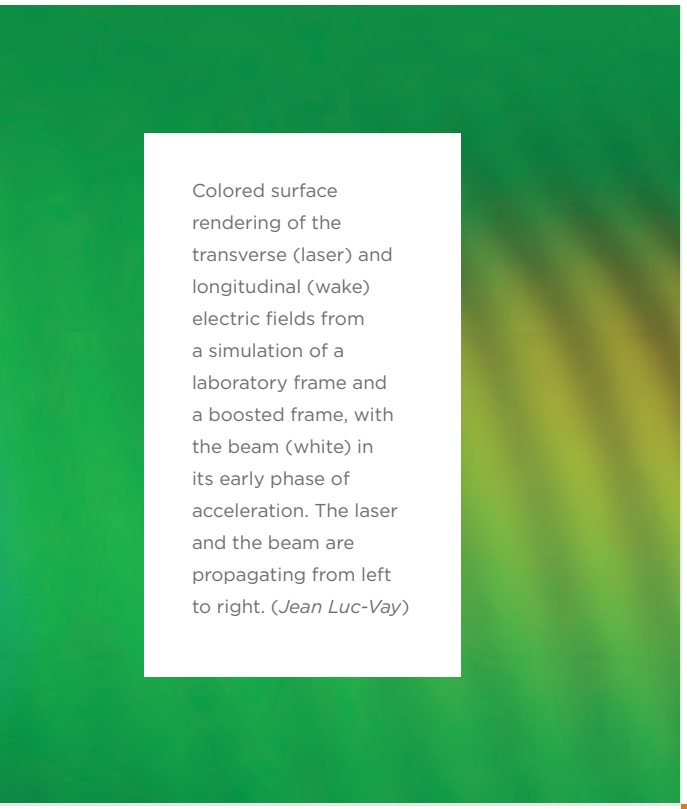
DOE Office
High Energy Physics

Albert Einstein's most famous thought experiment is proving its worth once again as researchers use it to help speed up the modeling (and thus design) of so-called "tabletop" accelerators.

Particle accelerators help physicists unlock the fundamental secrets of matter and the beginnings of our universe. But conventional accelerators are large and expensive. An emerging new class of compact laser-plasma accelerators is being designed to cost less and pack more power into much smaller spaces than today's machines.

However, generating computer models of laser-plasma accelerators has proven difficult, slow, and costly. To solve this problem, a team of scientists computing at NERSC, led by Jean-Luc Vay of Berkeley Lab, has perfected a new method that generates models anywhere from 10,000 to a million times faster than before.

This "boosted frame" approach uses a feature of special relativity to calculate in a just a few hours models that once took days, or even weeks, to complete. The method even allows researchers to execute models that weren't previously within the realm of possibility.



Colored surface rendering of the transverse (laser) and longitudinal (wake) electric fields from a simulation of a laboratory frame and a boosted frame, with the beam (white) in its early phase of acceleration. The laser and the beam are propagating from left to right. (*Jean Luc-Vay*)

Early attempts to apply the boosted-frame method to laser-plasma wakefield simulations were limited by numerical instabilities, small errors that are magnified the longer a calculation runs. Calculations could still be speeded up tens or even hundreds of times, but that was only a tiny fraction of the full promise of the method.

Vay's team showed that using a particular boosted frame, that of the wakefield itself, gave them near-optimal calculation speeds. Also, calculating fewer oscillations made it easier to deal with numerical instabilities. Combining these advantages with special techniques for interpreting the data between frames, Vay and his team were able to exploit the full potential of the boosted-frame principle.

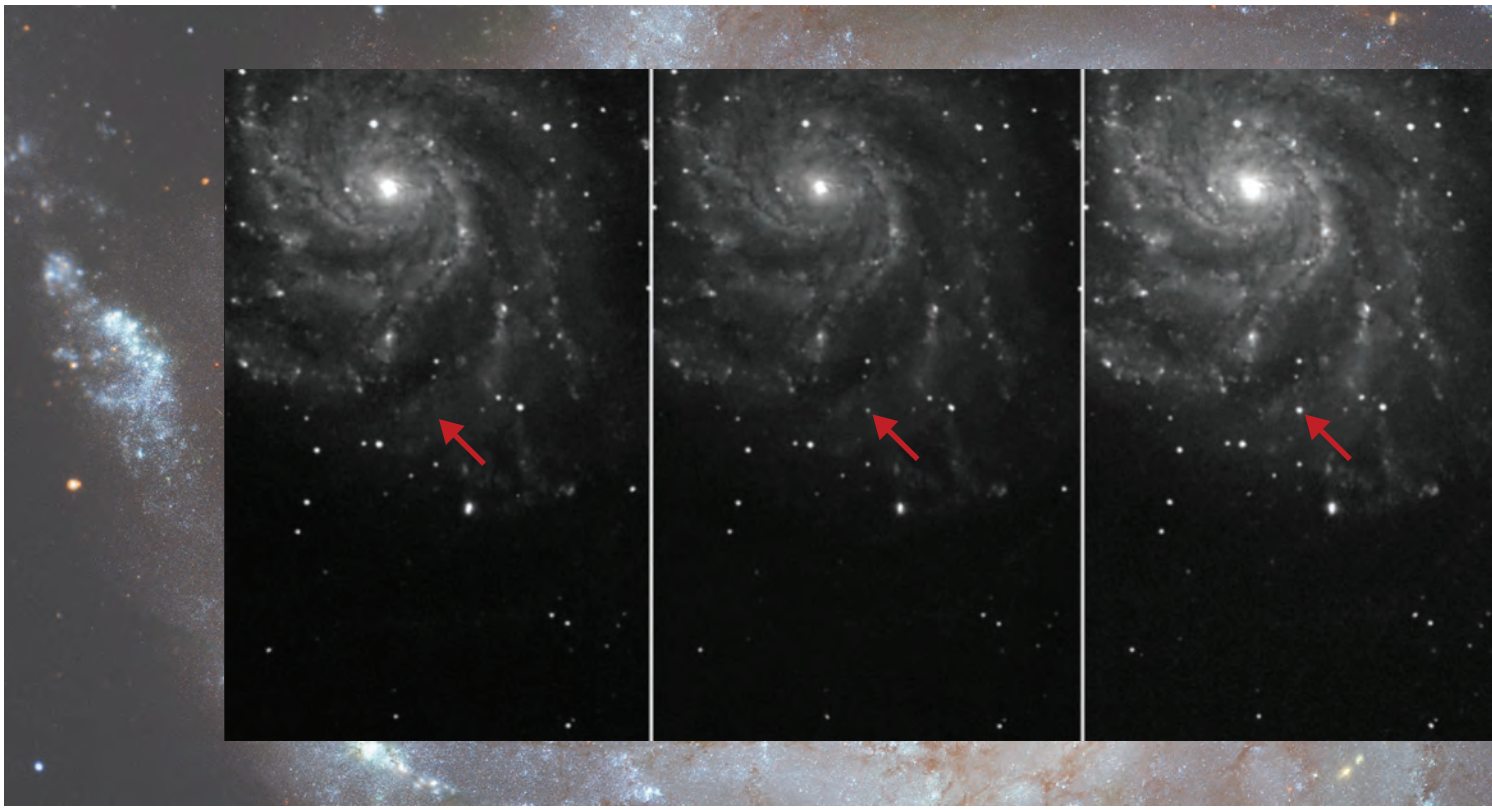
"We produced the first full multidimensional simulation of the 10 billion-electron-volt design for BELLA [the Berkeley Lab Laser Accelerator]," says Vay. "We even ran simulations all the way up to a trillion electron volts, which establishes our ability to model the behavior of laser-plasma accelerator stages at varying energies."

Speedier calculations will allow scientists to simulate the detailed physics they need in order to develop laser plasma accelerators not only for future particle colliders, but for other applications, as well.

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/boosting-the-next-wave-of-accelerators/

Publication: J.-L. Vay, C. G. R. Geddes, E. Cormier-Michel, and D. P. Grote, "Effects of hyperbolic rotation in Minkowski space on the modeling of plasma accelerators in a Lorentz boosted frame," Phys. Plasmas **18**, 030701 (2011), doi:10.1063/1.3559483.

Supernova Caught in the Act Reveals First Direct Evidence of Its Progenitor System



Project Name
Palomar Transient Factory

Project Leader
Peter Nugent, Lawrence Berkeley National Laboratory

NERSC Resources

- Data transfer nodes accept 100 GB of telescope data each night.
- Data analysis generates 300 GB on 2–4 Carver nodes for 12 hours each night.
- 300 GB database on Science Gateway nodes.
- Archived data on over 800 million candidate detections.

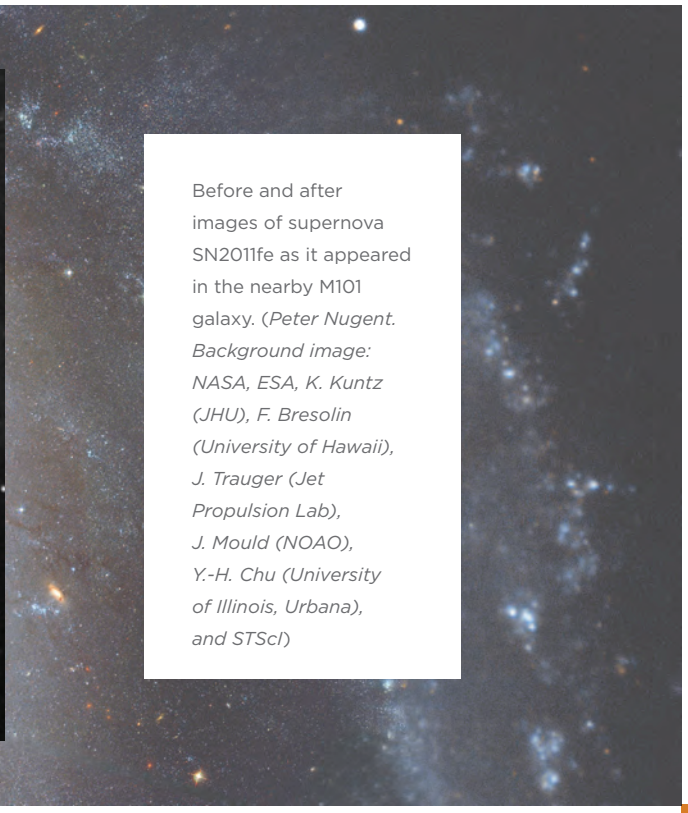
Type Ia supernovae (SN Ia's) are the extraordinarily bright and remarkably similar "standard candles" astronomers use to measure cosmic growth, a technique that in 1998 led to the discovery of dark energy—and 13 years later to a Nobel Prize, "for the discovery of the accelerating expansion of the universe." The light from thousands of SN Ia's has been studied, but until now their physics—how they detonate and what the star systems that produce them actually look like before they explode—has been educated guesswork.

Scientists long ago developed models of Type Ia supernovae based on their evolving brightness and spectra. The models assume the progenitor is a binary system—about half of all stars are in binary systems—in which a very dense, very small white-dwarf star orbits a companion, from which it sweeps up additional matter. When the white dwarf grows to a size about 1.4 times the mass of our sun, its own gravity causes it to collapse and explode as a supernova.

Peter Nugent co-leads Berkeley Lab's Computational Cosmology Center and also leads the Lab's collaboration in the multi-institutional Palomar Transient Factory. On August 24, 2011, searching data as it poured into NERSC from an automated telescope on Mount Palomar in California, Nugent spotted a remarkable object. It was shortly confirmed as a Type Ia supernova in the Pinwheel Galaxy, some 21 million light-years away—unusually close by cosmic standards.

Spotting the supernova just 11 hours after it exploded, the researchers were able to calculate the actual moment of the explosion to within 20 minutes, and gathered unprecedented data about the early stages of the explosion, including the first direct evidence that a supernova started as a white dwarf, evidence that the companion star was not much bigger than our sun, and surprising details about the non-uniform, "clumpy" nature of the explosion. Understanding how these giant explosions create and mix materials is important because supernovae are where most of the elements in the universe—and even our own bodies—are created.

The Palomar Transient Factory has discovered 1,500 supernovae to date, including 1,000 Type Ia supernovae.



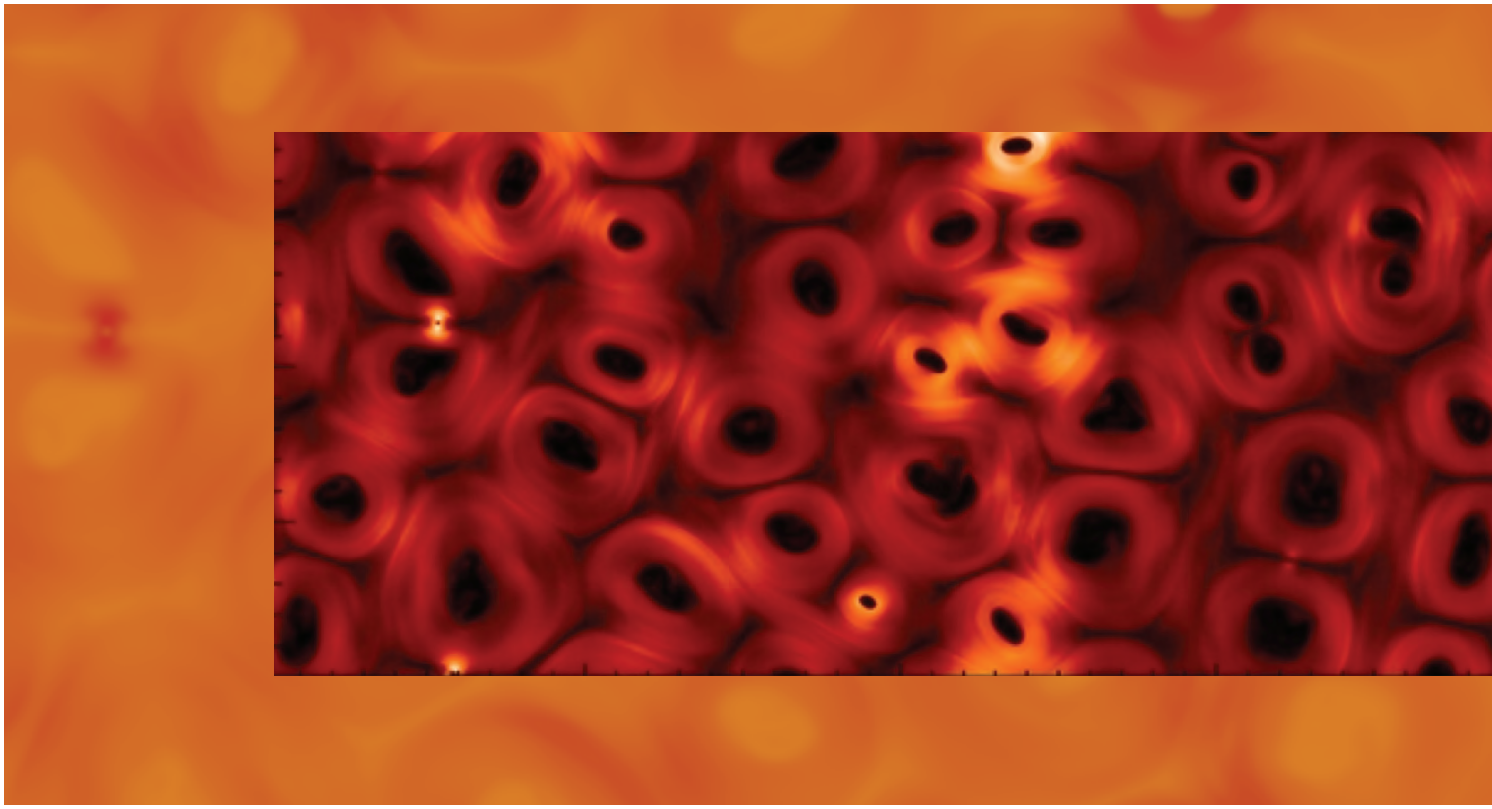
Before and after images of supernova SN2011fe as it appeared in the nearby M101 galaxy. (*Peter Nugent. Background image: NASA, ESA, K. Kuntz (JHU), F. Bresolin (University of Hawaii), J. Trauger (Jet Propulsion Lab), J. Mould (NOAO), Y.-H. Chu (University of Illinois, Urbana), and STScI*)

DOE Office
High Energy Physics

Full Story: www.nersc.gov/news-publications/news/science-news/2011/closest-type-ia-supernova-in-decades-solves-a-cosmic-mystery/

Publication: P.E. Nugent et al., "Supernova SN 2011fe from an exploding carbon–oxygen white dwarf star," Nature 480, 344 (2011), doi:10.1038/nature10644.

Solar System's Magnetic Edge More Turbulent Than Expected



Project Name
Turbulence, Transport and Magnetic Reconnection in High Temperature Plasma

Project Leader
William Dorland, University of Maryland

NERSC Resources
Franklin, 8,192 cores running simultaneously for 20 hours

DOE Office
Fusion Energy Sciences

NASA's Voyager probes have reached the end of our solar system, where they've found neither giants nor dragons, but something nearly as surprising—a turbulent froth of magnetic bubbles.

Using new computer models to analyze Voyager data, scientists computing at NERSC have found that the sun's distant magnetic field is made up of bubbles about 100 million miles wide. The bubbles are created when magnetic field lines reorganize, a process known as magnetic reconnection. This new model suggests the bubbles are self-contained structures disconnected from the solar magnetic field and may help scientists explain how some very strong cosmic rays make it to Earth.

A frame from a visualization showing how magnetic bubbles about 100 million miles across form at the edge of our solar system. It was created on NERSC's Franklin system using 8,192 processor cores running simultaneously for 20 hours. (*James F. Drake*)

When a magnetic field gets severely folded like this, interesting things can happen. Lines of magnetic force crisscross and reconnect. The crowded folds of the skirt reorganize themselves, sometimes explosively, into foamy magnetic bubbles. Theories dating back to the 1950s had predicted a very different scenario: The distant magnetic field of the sun was supposed to curve around in relatively graceful arcs, eventually folding back to rejoin the sun.

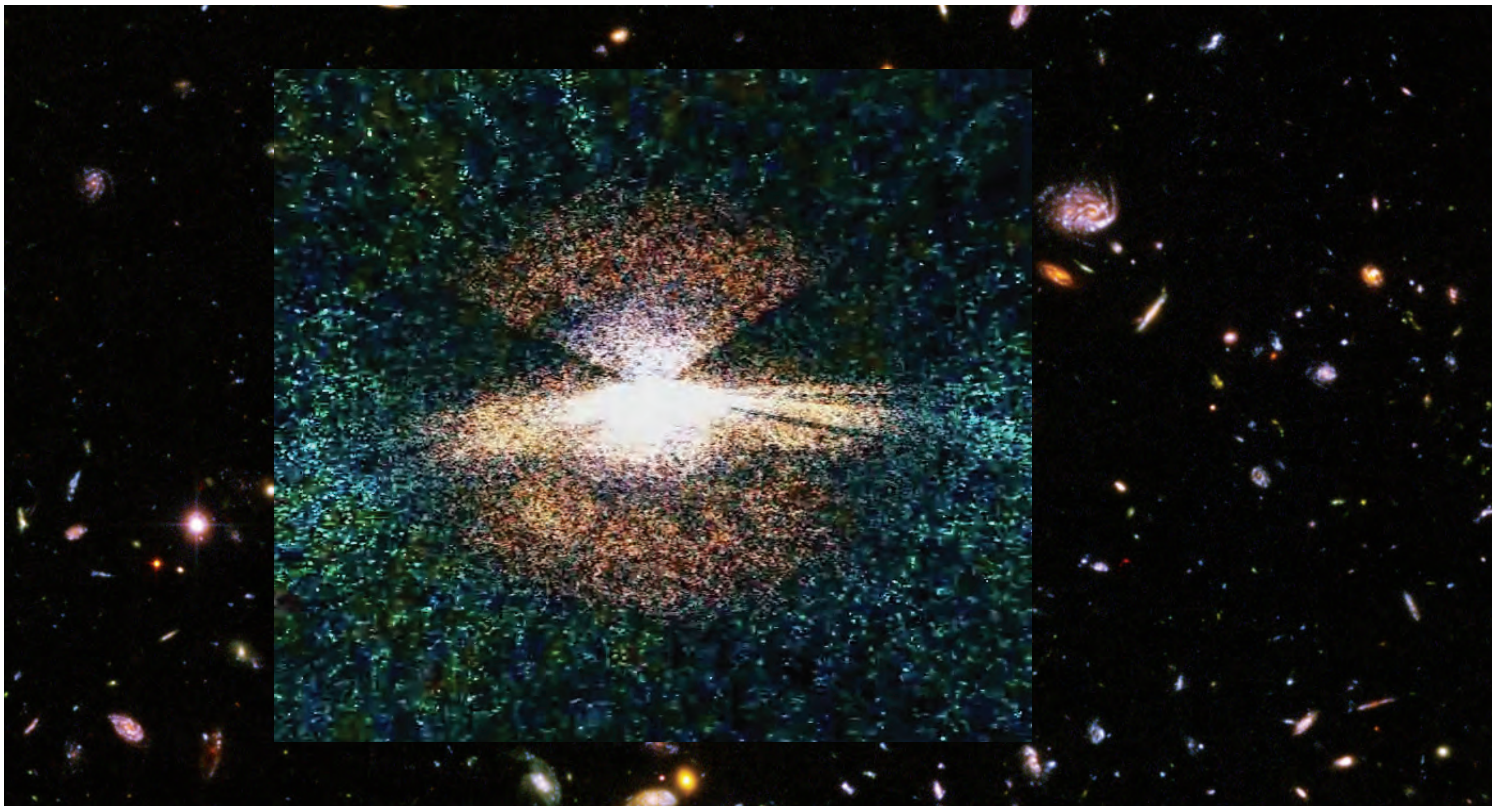
"We came up with a theory about how reconnection might be behind the data the Voyagers were returning," says James F. Drake, a University of Maryland physicist. The team proposed that magnetic reconnection was the source not only of the odd Voyager data, but of strong cosmic rays once thought to originate from the "bow shock," the point where the solar wind slows abruptly. "With observation, and then a theoretical idea backed by simulations carried out at NERSC, and then reconfirmation by observation, I think we've really convinced people that we may have been right about this," says Drake.

Understanding the structure of the sun's magnetic field will allow scientists to understand how our star interacts with the rest of the galaxy.

Full Story: http://www.nersc.gov/news-publications/news/science-news/2011/magnetic-bubbles/

Publication: M. Opher et al., "Is the magnetic field in the heliosheath laminar or a turbulent sea of bubbles?" Astrophysical Journal, **734**, 71 (2011), doi:10.1088/0004-637X/734/1/71.

Measuring the Distant Universe in 3D



Project Name
Baryon Oscillation Spectroscopic Survey

Project Leader
Peter Nugent, Lawrence Berkeley National Laboratory

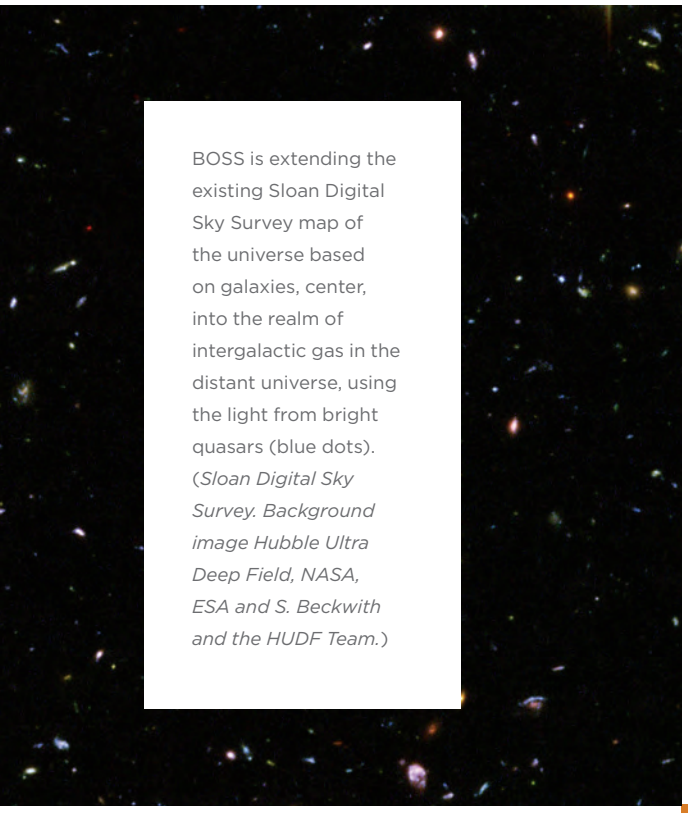
NERSC Resources
Franklin, 1,500 cores for ~30 hours

Other Computing Resources
LBNL

DOE Office
High Energy Physics

The biggest 3D map of the distant universe ever made, using light from 14,000 quasars—supermassive black holes at the centers of galaxies many billions of light years away—has been constructed by scientists with the third Sloan Digital Sky Survey (SDSS-III). The map is the first major result from the Baryon Oscillation Spectroscopic Survey (BOSS), SDSS-III's largest survey, whose principal investigator is David Schlegel of Berkeley Lab.

BOSS is the first attempt to use baryon acoustic oscillation (BAO) as a precision tool to measure dark energy. Baryon oscillation refers to how matter clumps in a regular way throughout the universe, a physical manifestation of the expansion of the universe. Until now, 3D maps showing this oscillation have been based on the distribution of visible galaxies. BOSS is the first survey to map intergalactic hydrogen gas as well, using light from distant quasars.



BOSS is extending the existing Sloan Digital Sky Survey map of the universe based on galaxies, center, into the realm of intergalactic gas in the distant universe, using the light from bright quasars (blue dots). (*Sloan Digital Sky Survey. Background image Hubble Ultra Deep Field, NASA, ESA and S. Beckwith and the HUDF Team.*)

"Quasars are the brightest objects in the universe, which we use as convenient backlights to illuminate the intervening hydrogen gas that fills the universe between us and them," says Anže Slosar of Brookhaven National Laboratory. "We can see their shadows, and the details in their shadows"—specifically, the absorption features in their spectra known as the Lyman-alpha forest—"allow us to see how the gas is clumped along our line of sight."

The new map demonstrates that indeed it is possible to determine variations in the density of intergalactic hydrogen gas at cosmological distances and thus to measure the effects of dark energy at those distances.

Shirley Ho and Martin White of Berkeley Lab worked with Slosar to develop simulated Lyman-alpha data that could be used to test their analytic methods and codes. Comparing the real data to the simulation confirmed that the search is working as hoped.

The distance scale of the new map corresponds to an early time in the history of the universe, when the distribution of matter was nearly uniform. Any effects of dark energy detected so early would settle basic questions about its nature. By the end of the BOSS survey, researchers will be able to measure how fast the universe was expanding 11 billion years ago with an accuracy of a few percent.

Full Story: http://newscenter.lbl.gov/news-releases/2011/05/01/boss-quasars/

Publication: Anže Slosar et al., "The Lyman-α forest in three dimensions: measurements of large scale flux correlations from BOSS 1st-year data," Journal of Cosmology and Astroparticle Physics, JCAP09(2011)001, doi:10.1088/1475-7516/2011/09/001.

NISE Program Encourages Innovative Research

NERSC allocates 10 percent of the total hours on its computational systems through the NERSC Initiative for Scientific Exploration (NISE) program, which provides computer time for exploring new areas of research. Selection criteria used in 2011 were:

- **A new research area not covered by the existing ERCAP proposal: this could be a tangential research project or a tightly coupled supplemental research initiative.**
- **New programming techniques that take advantage of multicore compute nodes by using OpenMP, Threads, UPC or CAF: this could include modifying existing codes, creating new applications, or testing the performance and scalability of multicore programming techniques.**
- **Developing new algorithms that increase researchers' scientific productivity, for example by running simulations at a higher scale or by incorporating new physics.**

Fifty-four projects received NISE awards in 2011, as listed below:

Project: New Potential Energy Surface for Ozone Molecule
PI: Dmitri Babikov, Marquette University
NISE award: 1,450,000 hours
Research areas/applications: Atmospheric chemistry, environmental science

Project: Hybrid OpenMP/MPI Approach to R-Matrix Scattering
PI: Connor Balance, Auburn University
NISE award: 600,000 hours
Research areas/applications: Atomic physics, plasma physics, fusion energy

Project: Global Effects on the Dynamics of Plasmoids and Flux Ropes During Magnetic Reconnection
PI: Amitava Bhattacharjee, University of New Hampshire
NISE award: 1,000,000 hours
Research areas/applications: Plasma physics, astrophysics, fusion energy

Project: MD Simulations of Liquid Water Films at the Boundary Between Gas Hydrate and Mineral Surfaces
PI: Ian C. Bourg, Lawrence Berkeley National Laboratory
NISE award: 300,000 hours
Research areas/applications: Geochemistry, carbon sequestration

Project: Systems Biology Knowledge Base
PI: Shane Canon, Lawrence Berkeley National Laboratory
NISE award: 100,000 hours
Research areas/applications: Biological systems, computer science

Project: Sparse Lanczos and GMRES Solvers in Co-Array Fortran (CAF)
PI: Pierre Carrier, University of Minnesota, Twin Cities
NISE award: 80,000 hours
Research areas/applications: Computer science, materials science

Project: Numerical Study of Mode Conversion and Ion Heating in Laboratory-Relevant Plasmas
PI: Paul Cassak, West Virginia University
NISE award: 200,000 hours
Research areas/applications: Plasma physics, astrophysics, fusion energy

Project: Large Scale Numerical Study of Bootstrap Current in Edge Pedestal Plasma
PI: Choong-Seock Chang, Courant Institute of Mathematical Sciences, New York University
NISE award: 4,000,000 hours
Research areas/applications: Plasma physics, fusion energy

Project: Ab Initio Simulation of Mechanical Properties of Bulk Metallic Glass and Mo-Based Alloys
PI: Wai-Yim Ching, University of Missouri, Kansas City
NISE award: 7,650,000 hours
Research areas/applications: Materials science, nanoscience, fossil energy technologies

Project: Improving Algorithms and Resolution for the Ocean-Atmosphere Reanalysis for Climate Applications (OARCA) 1850–2012
PI: Gil Compo, University of Colorado at Boulder
NISE award: 7,000,000 hours
Research areas/applications: Climate science

Project: High Resolution Modeling to Identify the Location of the Meltwater Flood Responsible for the Younger Dryas Cold Episode
PI: Alan Condron, Pennsylvania State University
NISE award: 1,000,000 hours
Research areas/applications: Climate science

Project: Numerical Investigations of Symmetry Breaking, Magnetism and the Pseudogap in the Three-Orbital Hubbard Model
PI: Thomas Devereaux, SLAC National Accelerator Laboratory
NISE award: 1,000,000 hours
Research areas/applications: Materials science

Project: Breaking Magnetic Field Lines with Turbulence
PI: James Drake, University of Maryland
NISE award: 500,000 hours
Research areas/applications: Plasma physics, astrophysics, fusion energy

Project: Development of CALPHAD Type Approach to Thermal Conductivity Aided by First-Principles Phonon Approach
PI: Huazhi Fang, Pennsylvania State University
NISE award: 500,000 hours
Research areas/applications: Materials science

Project: An Efficient Real-Space Approach to Calculating Auger Spectra
PI: Keith Gilmore, Lawrence Berkeley National Laboratory
NISE award: 800,000 hours
Research areas/applications: Materials science, nanoscience, energy technologies

Project: Surface Instabilities and Subgrid Scale Models in Computational Fluid Dynamics
PI: James Glimm, Stony Brook University
NISE award: 500,000 hours
Research areas/applications: Applied mathematics, fusion energy

Project: Self-Healing of Polymer Films
PI: Gary Grest, Sandia National Laboratories
NISE award: 3,500,000 hours
Research areas/applications: Materials science, nanoscience

Project: Computational Studies of Methanol Steam Reforming
PI: Hua Guo, University of New Mexico
NISE award: 300,000 hours
Research areas/applications: Chemistry, fuel cells

Project: Laser Acceleration of Ions Using Few Times Critical Density Gas Targets
PI: Michael Helle, Naval Research Laboratory
NISE award: 1,000,000 hours
Research areas/applications: Accelerator physics

Project: An Integrative Strategy to Model Complex Biological Assemblies
PI: Ivaylo Ivanov, Georgia State University
NISE award: 960,000 hours
Research areas/applications: Biological systems

Project: Unveiling Microbial Carbon Cycling Processes in Key U.S. Soils using “omics”
PI: Janet Jansson, Lawrence Berkeley National Laboratory
NISE award: 1,000,000 hours
Research areas/applications: Biological systems

Project: Semiclassical Approaches for Clean Energy Resources
PI: Puru Jena, University of California Berkeley
NISE award: 100,000 hours
Research areas/applications: Materials science, energy technologies

Project: Tall Tower Wind Energy Monitoring and Numerical Model Validation in Southern Nevada
PI: Michael Kaplan, Desert Research Institute
NISE award: 750,000 hours
Research areas/applications: Environmental science

Project: Threading EnergyPlus Using OpenMP
PI: Noel Keen, Lawrence Berkeley National Laboratory
NISE award: 100,000 hours
Research areas/applications: Environmental science, computer science

Project: First-Principles Calculation of Phonon-Assisted Optical Absorption in Indirect-Band-Gap Semiconductors
PI: Emmanouil Kioupakis, University of California, Santa Barbara
NISE award: 1,000,000 hours
Research areas/applications: Materials science, energy technologies

Project: Deep Defect States and Valence Skipping in Narrow Band Gap Semiconductors
PI: Mal Soon Lee, Michigan State University
NISE award: 340,000 hours
Research areas/applications: Materials science, energy technologies

Project: Material Simulations in Joint Center for Artificial Photosynthesis (JCAP)
PI: Nathan Lewis, California Institute of Technology
NISE award: 5,000,000 hours
Research areas/applications: Materials science, energy technologies

Project: High Fidelity Study of Multiscale Multiphase Lubricant Flow in High-Speed Gear Systems
PI: Xiaoyi Li, United Technologies Corporation Research Center
NISE award: 550,000 hours
Research areas/applications: Combustion technologies, computer science

Project: Semiclassical Approaches for Clean Energy Resources
PI: Jian Liu, University of California Berkeley
NISE award: 1,350,000 hours
Research areas/applications: Chemistry, energy technologies

Project: First Principles Characterization of van der Waals Dispersion Interactions
PI: Deyu Lu, Brookhaven National Laboratory
NISE award: 350,000 hours
Research areas/applications: Materials science, energy technologies

Project: Simulating ETG Plasma Turbulence with Impurities
PI: David Mikkelsen, Princeton Plasma Physics Laboratory
NISE award: 3,000,000 hours
Research areas/applications: Plasma physics, fusion energy

Project: Coupled Electronic and Nuclear Dynamics in Enzymes and Photocatalysts
PI: Thomas Miller, California Institute of Technology
NISE award: 750,000 hours
Research areas/applications: Chemistry, energy technologies

Project: Proton-Coupled Electron Transfer in Photocatalysis and Protein Translocation in Biosynthesis: Bridging Lengthscales and Timescales in Molecular Simulation
PI: Thomas Miller, California Institute of Technology
NISE award: 10,000,000 hours
Research areas/applications: Chemistry, biological systems, energy technologies

Project: Three Dimensional Models for Martensitic Phase Transformations
PI: Benson Muite, University of Michigan
NISE award: 200,000 hours
Research areas/applications: Environmental science, energy technologies

Project: Molecular Simulations of the Release of Phosphate and ADP from F1-ATPase to Aid in the Understanding of the Function of F1-ATPase
PI: Victor Ovchinnikov, Massachusetts Institute of Technology
NISE award: 1,000,000 hours
Research areas/applications: Biosciences, biofuels

Project: Turbulence over Complex Terrain: A Wind Energy Perspective
PI: Edward Patton, National Center for Atmospheric Research
NISE award: 5,500,000 hours
Research areas/applications: Environmental science, energy technologies

Project: Performance Enhancements for Three-Dimensional Fast Fourier Transforms Library P3DFFT
PI: Dmitry Pekurovsky, University of California, San Diego
NISE award: 350,000 hours
Research areas/applications: Computer science, fluid dynamics, astrophysics, materials science, oceanography, tomography

Project: Accelerated Materials Design Towards the Materials Genome
PI: Kristin Persson, Lawrence Berkeley National Laboratory
NISE award: 2,500,000 hours
Research areas/applications: Materials science, energy technologies

Project: Triple Photoionization of the Li Atom
PI: Mitch Pindzola, Auburn University
NISE award: 3,000,000 hours
Research areas/applications: Atomic physics, plasma physics, fusion energy

Project: Implementation of C-MAD Parallel Code to Include New Physics and Capabilities for Future Particle Colliders
PI: Mauro Pivi, SLAC National Accelerator Laboratory
NISE award: 300,000 hours
Research areas/applications: Accelerator physics

Project: Morphology of Young Core-Collapse Supernova Remnants from Multi-Physics Three-Dimensional Simulations
PI: Tomasz Plewa, Florida State University
NISE award: 1,000,000 hours
Research areas/applications: Astrophysics

Project: Large High-Resolution Galaxy Simulation Program
PI: Joel Primack, University of California, Santa Cruz
NISE award: 1,500,000 hours
Research areas/applications: Astrophysics

Project: Towards High Performance Urban Canopy Flows
PI: Joe Prusa, Teraflux Corporation
NISE award: 400,000 hours
Research areas/applications: Climate science

Project: Evaluating PGAS scientific graph analysis codes on the Gemini interconnect
PI: Jason Riedy, Georgia Institute of Technology
NISE award: 250,000 hours
Research areas/applications: Computer science

Project: Representing Convective Clouds in Global Circulation Models
PI: David Romps, Lawrence Berkeley National Laboratory
NISE award: 1,000,000 hours
Research areas/applications: Climate science

Project: Simulation of Inelastic Decoherence in Electronic Transport Through Nanoscale Structures
PI: Sayeef Salahuddin, University of California, Berkeley
NISE award: 750,000 hours
Research areas/applications: Materials science, nanoscience, energy technologies

Project: Improving Scalability of an Eigensolver for the Kohn-Sham Electronic Structure Problem
PI: Grady Schofield, University of Texas at Austin
NISE award: 1,000,000 hours
Research areas/applications: Materials science, computer science

Project: Phonon, Thermodynamic and Diffusion Properties of Lithium-Ion Batteries from First-Principles Calculations
PI: ShunLi Shang, Pennsylvania State University
NISE award: 500,000 hours
Research areas/applications: Materials science, energy technologies

Project: CACHE Novel Architecture Research
PI: Brian Van Straalen, Lawrence Berkeley National Laboratory
NISE award: 470,000 hours
Research areas/applications: Computer science

Project: High Performance Simulation of Reactive Transport Processes at the Pore Scale
PI: Brian Van Straalen, Lawrence Berkeley National Laboratory
NISE award: 1,200,000 hours
Research areas/applications: Applied mathematics, carbon sequestration

Project: First-Principles Calculations of Multiferroic Oxides
PI: Yi Wang, Pennsylvania State University
NISE award: 1,2500,000 hours
Research areas/applications: Materials science, electronics, spintronics

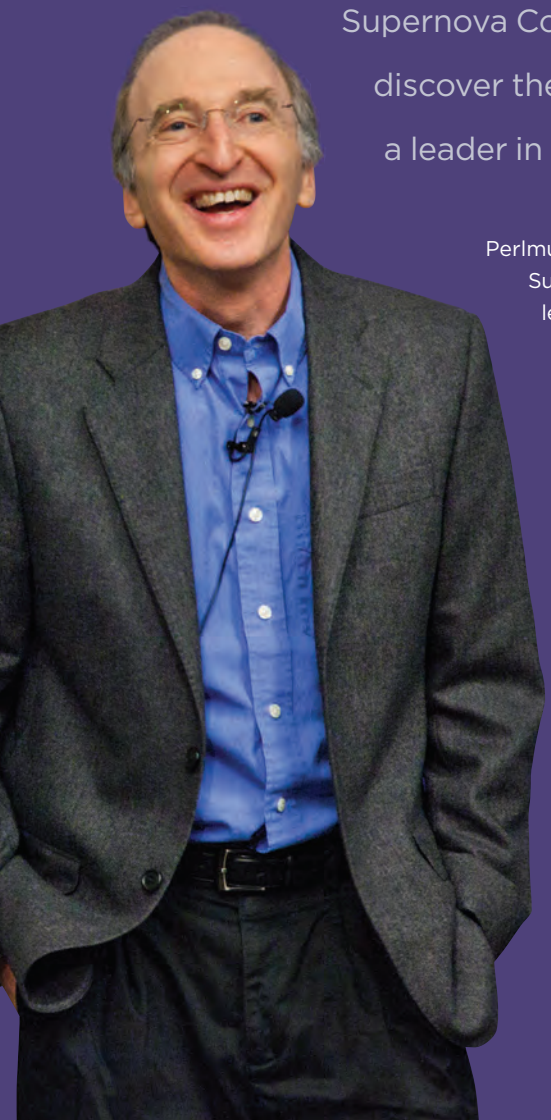
Project: High Resolution Integration of CAM5 for Extreme Weather Research
PI: Michael Wehner, Lawrence Berkeley National Laboratory
NISE award: 5,600,000 hours
Research areas/applications: Climate science

Project: Modeling and Simulation of High Dimensional Stochastic Multiscale PDE Systems at the Exascale
PI: Nicholas Zabaras, Cornell University
NISE award: 1,500,000 hours
Research areas/applications: Applied mathematics, energy technologies

Project: Theoretical Investigation on the Properties of Energy Materials
PI: Qianfan Zhang, Stanford University
NISE award: 400,000 hours
Research areas/applications: Materials science, nanoscience, energy technologies

Nobel Prize for Physics Leads List of Honors for NERSC Users in 2011

Saul Perlmutter, an astrophysicist at Lawrence Berkeley National Laboratory and a professor of physics at the University of California at Berkeley, won the 2011 Nobel Prize in Physics "for the discovery of the accelerating expansion of the universe through observations of distant supernovae." Perlmutter heads the international Supernova Cosmology Project, which pioneered the methods used to discover the accelerating expansion of the universe, and he has been a leader in studies to determine the nature of dark energy.



Perlmutter shared the prize with Brian Schmidt and Adam Riess, leader of the High-z Supernova Search Team and first author of that team's analysis, respectively, which led to their almost simultaneous announcement of accelerating expansion.

The Supernova Cosmology Project used a robotic telescope equipped with a digital detector instead of photographic plates. Its digital images were compared with earlier images using "subtraction" software. By 1994 the SCP team could discover supernovae "on demand," and Perlmutter realized he would soon need more computing power to analyze the growing flow of data. NERSC's move to Berkeley Lab in 1996 provided the perfect opportunity for his team.

With a Laboratory Directed Research and Development (LDRD) grant, the NERSC and Physics divisions jointly hired a postdoc. Peter Nugent—now leader of the NERSC Analytics Team and co-leader of Berkeley Lab's Computational Cosmology Center (C3)—helped the group develop parallel algorithms that could run on 128 cores at once, taking great advantage of NERSC's "Mcurie" system, a 512-processor Cray T3E-900 supercomputer.

To analyze the data from 40 supernovae for errors or biases, Nugent simulated 10,000 exploding supernovae at varying distances under varying circumstances. These were then plotted and compared with the observed data to detect any biases affecting observation or interpretation. The Cray T3E supercomputer was also used to check and recheck their work by resampling the data and running calculations that helped determine the reliability of their measurements thousands of times.

These rigorous, supercomputer-powered analyses of potential biases reduced the uncertainties in the data and helped Perlmutter's team win widespread acceptance of their conclusions in the scientific community. His research team is believed to have been the first to use supercomputers to analyze and validate observational data in cosmology. This melding of computational science and cosmology sowed the seeds for more projects, establishing Berkeley Lab and NERSC as centers for the emerging field.

Other NERSC users who received awards and honors for their achievements include:

Fellow of the American Academy of Arts and Sciences
Martin Head-Gordon, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

Member of the National Academy of Sciences
James W. Demmel, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

Members of the National Academy of Engineering
Karsten Pruess, *Lawrence Berkeley National Laboratory*
Ramamoorthy Ramesh, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

Fellows of the American Association for the Advancement of Science
Alexander V. Balatsky, *Los Alamos National Laboratory*
Alan L. Balch, *University of California, Davis*
Shiv N. Khanna, *Virginia Commonwealth University*
Chung-Pei Ma, *University of California, Berkeley*
Alfred Z. Msezane, *Clark Atlanta University*
Robert L. Sugar, *University of California, Santa Barbara*
Martin White, *University of California, Berkeley*

Member of the American Philosophical Society
Graham R. Fleming, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

Fellows of the American Physical Society (APS)
Edward A. Baron, *University of Oklahoma*
Robert Edwards, *Thomas Jefferson National Accelerator Facility*
Chueng-Ryong Ji, *North Carolina State University*
Feng Liu, *University of Utah*
Serdar Ogut, *University of Illinois, Chicago*
Dvora Perahia, *Clemson University*
Gary Staebler, *General Atomics*

APS J.J. Sakurai Prize for Theoretical Particle Physics
Ian Hinchliffe, *Lawrence Berkeley National Laboratory*

American Chemical Society (ACS) F. Albert Cotton Award in Synthetic Inorganic Chemistry
Alan L. Balch, *University of California, Davis*

Journal of Physical Chemistry A Festschrift
Graham R. Fleming, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

Fellow of the Materials Research Society
James Chelikowsky, *University of Texas at Austin*

Glenn T. Seaborg Actinide Separations Award
Darleane Hoffman, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

John Simon Guggenheim Fellowhip
Martin White, *University of California, Berkeley, and Lawrence Berkeley National Laboratory*

Innovations

NERSC is known as one of the best-run scientific computing facilities in the world. It provides some of the largest computing and storage systems available anywhere; but what distinguishes the center is its success in creating an environment that makes these resources effective for scientific research.

NERSC systems are reliable and secure, and provide a state-of-the-art scientific development environment with the tools needed by the diverse community of NERSC users. NERSC offers scientists intellectual services that empower them to be more effective researchers. For example, many of our consultants are themselves domain scientists in areas such as material sciences, physics, chemistry, and astronomy, well-equipped to help researchers apply computational resources to specialized science problems.

This section provides examples of innovations NERSC implemented in 2011 to increase scientific productivity.

Petaflops Power to NERSC

NERSC marked a major milestone in May 2011, putting its first petascale supercomputer into the hands of its 4,000 scientific users. The flagship Cray XE6 system, named "Hopper" in honor of American computer scientist Grace Murray Hopper, is capable of more than one quadrillion floating point operations per second, or one petaflops. At the time of its acceptance into full production status, Hopper was ranked by the TOP500 list as the second most powerful supercomputer in the United States.

Before this milestone, several pioneering users put the machine through its paces while making scientific discoveries in a broad set of areas, from fundamental science to clean energy alternatives and severe storm modeling.

"We are very excited to make this unique petascale capability available to our users, who are working on some of the most important problems facing the scientific community and the world," said Kathy Yelick, NERSC Director. "With its 12-core AMD processor chips, the system reflects an aggressive step forward in the industry-wide trend toward increasing the core counts, combined with the latest innovations in high-speed networking from Cray. The result is a powerful instrument for science. Our goal at NERSC is to maximize performance across a broad set of applications, and by our metric, the addition of Hopper represents an impressive five-fold increase in the application capability of NERSC."

Peter Ungaro, president and CEO of Cray, emphasized the collaboration between the two organizations in tailoring the machine for NERSC. "Our partnership with NERSC has been important in further increasing these capabilities in our Cray XE6 supercomputer, not only for NERSC but for all our customers around the world. We worked together to improve the functionality and performance of our external services offerings from our Custom Engineering organization, as well as put our new Gemini system interconnect to use across an amazingly broad set of scalable applications that needed all the performance they could get to achieve their scientific goals. We're proud that Hopper is the first petascale system at NERSC, and I'm convinced it will be a great tool for their users as they strive toward the next scientific breakthroughs."

"Uptake of the Hopper system by early science users was quite impressive," said Jonathan Carter, Computer Sciences Deputy and lead on the Hopper procurement project. "We started with a select set of users and gradually opened the system up to all interested users as the testing period progressed. From the earliest access, the system was heavily utilized and very popular." Here is a selection of their stories:

From left to right: Horst Simon (Berkeley Lab deputy director), Kathy Yelick (NERSC division director), Dan Hitchcock (DOE ASCR), and Paul Alivisatos (Berkeley Lab director) celebrating Hopper's transition from testing phase into full production mode.

Understanding DNA Replication

Our DNA is damaged every day from exposure to the sun's ultraviolet light, secondhand smoke, toxins released by mold, and various other factors. Fortunately, nature has equipped us all with genes that repair and replicate DNA. But when these repair systems go awry, the result may be fatal cancerous tumors or degenerative diseases. A team of researchers from Georgia State University ran simulations on Hopper to understand the basic mechanisms of repair, which could lead to new preventions and treatments.

"My group has been very pleased with our experience running on Hopper," said Ivaylo Ivanov, an assistant professor of chemistry at Georgia State University, who is part of the team. "Our parallel jobs seem to scale much better on Hopper compared to similar systems. I suspect it has to do with the better processor interconnect network. This makes it possible to run on many more processor cores and make quicker progress on our simulations."

Improving the Accuracy of Storm Surge Forecasts

In the event of a hurricane, storm surges are the greatest threat to life and property along the coast. In fact, experts estimate that most of the 1,500 deaths caused by Hurricane Katrina were the result of a surge that occurred when winds that moved cyclonically around the storm in the Gulf of Mexico pushed water toward the shore. The ability to simulate these events on a computer is a powerful tool for evaluating risk, designing hurricane protection systems, planning evacuations, and analyzing the physics of a storm.

Today, supercomputers built on parallel architectures with fast networks allow researchers to get these simulations with a fast turnaround. But their codes need to run at peak efficiency on tens of thousands of processors to get detailed results as fast as possible. After all, emergency planners typically need to get anywhere from two to four days of storm simulations in an hour of computer time to plan for human safety.

Researchers Seizo Tanaka and Patrick Kerr of the University of Notre Dame's Department of Civil Engineering and Geological Sciences, and Jay Ratcliff of the U.S. Army Corps of Engineers tested the scalability of their high-resolution storm surge code using Hopper pre-acceptance time.

"Advance time on Hopper has helped greatly in the scaling benchmarks as well in the simulations I've been performing related to hurricane modeling and sea level rise impacts," said Ratcliff.

Harnessing Wind Power with Hopper

Although wind power technology is close to being cost-competitive with fossil fuel plants for generating electricity, wind turbine installations still provide less than one percent of all U.S. electricity. Because scientists don't

Hurricane Katrina on August 26, 2005.

have detailed knowledge about how unsteady flows interact with wind turbines, many turbines underperform, suffer permanent failures or break down sooner than expected.

Since standard meteorological datasets and weather forecasting models do not provide detailed information on the variability of conditions needed for the optimal design and operation of wind turbines, researchers at the National Center for Atmospheric Research (NCAR) developed a massively parallel large-eddy simulation (LES) code for modeling turbulent flows in the planetary boundary layer—the lowest part of the atmosphere, which interacts with the shape and ground cover of the land.

With approximately 16,000 processor cores on Hopper, the NCAR team simulated the turbulent wind flows over hills in unprecedented resolution and increased the scalability of their code to ensure that it will be able to take advantage of peta- and exascale computer systems.

"The best part of Hopper is the ability to put previously unavailable computing resources toward investigations that would otherwise be unapproachable," said Ned Patton of NCAR, who heads the investigation. "We seriously couldn't make the progress we have been without NERSC's support. We find NERSC's services to be fantastic and truly appreciate being able to compute there."

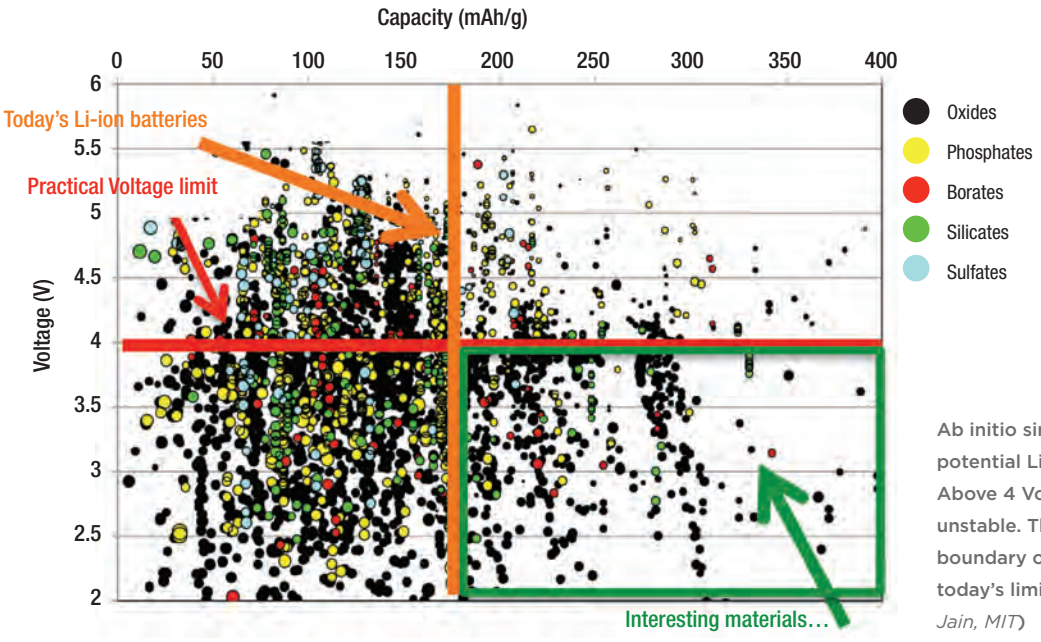
Accelerating Advanced Materials Development

New materials are crucial to building a clean energy economy—for everything from batteries to photovoltaics to lighter weight vehicles—but today the development cycle is too slow: around 18 years from conception to commercialization. To speed up this process, a team of researchers from Berkeley Lab and the Massachusetts Institute of Technology (MIT) teamed up to develop a new computational tool. Called the Materials Project, it was launched in October 2011.

California's San Gorgonio pass wind farm.

The tool, which sits on NERSC's science gateway infrastructure, was developed with support from the Department of Energy and a Laboratory Directed Research and Development grant from Berkeley Lab.

"Our vision is for this tool to become a dynamic 'Google' of material properties, which continually grows and changes as more users come on board to analyze the results, verify against experiments and increase their knowledge," says



Ab initio simulation of over 20,000 potential Li-ion cathode compounds. Above 4 Volts, solvents become unstable. This research is pushing the boundary of capacity upward beyond today's limit of 170 mAh/g. (*Anubhav Jain, MIT*)

Kristin Persson, a Berkeley Lab chemist and one of the founding scientists behind the Materials Project. "So many scientists can benefit from this type of screening. Considering the demand for innovative clean energy technology, we needed most of these materials yesterday."

Taking a Genome-Like Approach

The Materials Project employs an approach to materials science inspired by genomics. But rather than sequencing genomes, researchers are using supercomputers to characterize the properties of inorganic compounds, such as their stability, voltage, capacity, and oxidation state. The results are then organized into a database with a user-friendly web interface that gives all researchers free and easy access and searching.

"First-principles calculations have reached the point of accuracy where many materials properties, relevant for photovoltaics, batteries and thermoelectrics, can be reliably predicted," says Gerbrand Ceder, an MIT professor of materials science and engineering and founder of the Materials Project.

A better battery—one that is cheaper and has more power and energy while being safe—could finally make possible the dream of an electric vehicle reaching performance and cost parity with a gasoline-powered car. But beyond batteries, novel materials could transform a host of other industries, from food packaging to buildings. For example, the Materials Project is working with several entities interested in making stronger, corrosion-resistant, lightweight aluminum alloys, which could make possible lighter vehicles and airplanes.

"Materials innovation today is largely done by intuition, which is based on the experience of single investigators," says Persson, who works in Berkeley Lab's Environmental Energy Technologies Division. "The lack of comprehensive knowledge of materials, organized for easy analysis and rational design, is one of the foremost reasons for the long process time in materials discovery."

President Obama has recognized the importance of advanced materials with his announcement in June 2011 of the Materials Genome Initiative "to double the speed with which we discover, develop, and manufacture new materials." Many of the concepts of that initiative were inspired by the Materials Project, Persson said.

With the help of supercomputers at the Department of Energy's NERSC, the Berkeley Lab Lawrencium cluster, and systems at the University of Kentucky, the Materials Project database currently contains the structural and energetic properties of more than 28,000 inorganic compounds, and up to hundreds more are added every day. Researchers are continuously adding new properties to enable true rational design of new materials for a wide variety of applications.

A Gateway for Science

To build the Materials Project web tool, the team approached computer systems engineers at NERSC who have extensive experience building web-based interfaces and technologies. These science gateways make it easier for researchers to access computational resources and share data with the rest of their community.

"The Materials Project represents the next generation of the original Materials Genome Project, developed by Ceder's team at MIT," says Shreyas Cholia, a NERSC computer engineer who helped develop the Materials Project tool. "The core science team worked with developers from NERSC and Berkeley Lab's Computational Research Division to expand this tool into a more permanent, flexible, and scalable data service built on top of rich modern web interfaces and state-of-the-art NoSQL database technology."

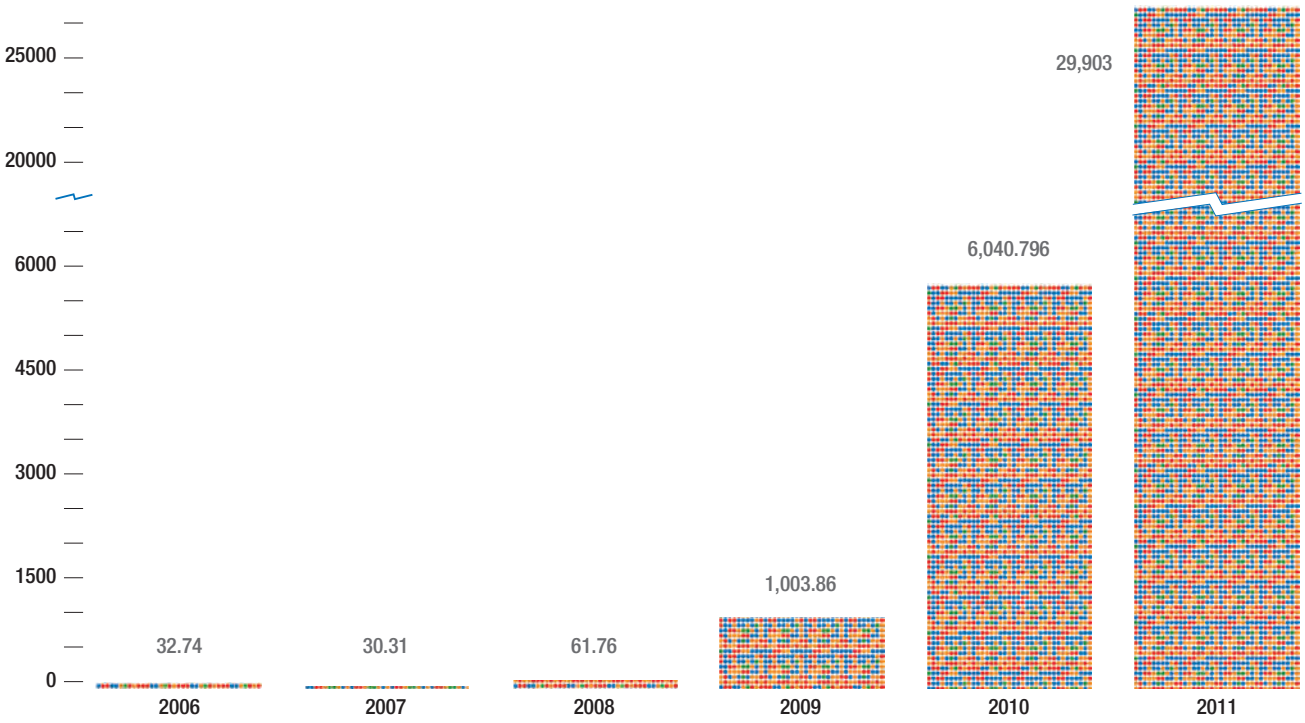
In addition to Persson and Cholia, other Berkeley Lab contributors to this project include Michael Kocher, Daniel Gunter, Annette Greiner, David Skinner and David Bailey. MIT collaborators include Gerbrand Ceder, Shyue Ping Ong, Anubhav Jain, Geoffroy Hautier and Evgueni Chtykov.

"At NERSC we have a long history of engaging with science teams to create web-based tools that allow scientists to share and access data, perform computations, and interact with NERSC systems using web-based technologies, so it was a perfect match," adds Cholia.

Supporting Diverse Workloads on HPC Platforms

The high performance computing (HPC) platforms at centers like NERSC provide amazing computing capability; however, many users with growing computing needs have encountered barriers when trying to take advantage of these systems. Genomics is a clear example of this. Improvements in sequencing technology have led to dramatic year-by-year improvements (see chart). This has led to a commensurate increase in the computing requirements for this domain.

This rapid growth in computing requirements was one reason that the DOE Joint Genome Institute (JGI) turned to NERSC to help meet its computing needs. Since HPC systems have not been specifically architected to match this workload, genomics researchers have typically avoided HPC systems in the past. However, the scale of systems like Hopper coupled with the rapidly growing computing needs of genomics have forced a reevaluation. NERSC has worked with users in the genomics community to understand their



DOE JGI sequence productivity: total bases (in billions of bases or Gb).

workloads and developed tools to help lower the barrier to entry for these users. A few examples of the innovations we have developed include a task farmer, tools to create a private virtual cluster, and file caching utilities for data intensive applications.

Task Farmer

Many of the workloads in the genomics community, as well as other data intensive communities, are high-throughput oriented. Unfortunately, the typical scheduling policies as well as the run-time systems on typical HPC systems are not well suited for these workloads. To address this shortcoming, NERSC has developed a task farmer that simplifies running high-throughput workloads.

The task farmer provides several critical capabilities, including load balancing, fault tolerance, and aggregation of output. The goal of the design is to make it trivial to run serial applications at scale. The task farmer uses a client-server architecture. The clients are launched in parallel as part of a parallel job submission. This helps address many of the scheduling policies constraints, since the throughput workload is scheduled as a single parallel job. The task farmer tracks the execution of tasks and automatically reschedules tasks that fail to complete. The framework also maintains a checkpoint of job progress and can be easily restarted after a failure or running out of wall time. The framework can also automatically collect output from the tasks and aggregate it into a single set of output files. This avoids the need to create output files for each task execution, which can lead to a large number of small files that are inefficient for the file system and difficult to manage.

The task farmer was initially developed to simplify performing alignment of sequence data using BLAST (Basic Local Alignment Search Tool—software that compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches). The task farmer has been used to perform critical BLAST analysis for a variety of metagenomic and microbial data sets. The framework has since been generalized and used for other genomic analysis including hidden Markov model search (HMMER), assembly, and data reorganization.

Virtual Private Clusters

While the task farmer is useful for running many workloads, there is still a need for a more flexible tool for complex workflows. Workloads that have complex dependencies or need to dynamically generate new job steps based on the results from previous steps are more aligned with the capabilities of traditional schedulers and resource managers like Sun Grid Engine (SGE) or Maui Torque. However, making policy changes to an HPC scheduler typically requires careful review and may result in a negative impact for other workloads.

To help address this situation, NERSC has developed tools to create a virtual private cluster (VPC) within a parallel allocation. In 2011 NERSC implemented and beta tested

a prototype with Sun Grid Engine; the prototype tool is called MySGE. Using MySGE, a user requests a large parallel allocation from the system scheduler (Moab in the case of Hopper). When the parallel job is started, the VPC system starts a personal SGE scheduler for the user and starts up the execution daemons on the allocated compute nodes. The user can then submit the workload to the personal scheduler, and these jobs will be executed on the allocated compute nodes.

The jobs can include large job arrays, large numbers of serial steps, and job dependencies which are difficult to efficiently run on a typical HPC system. Furthermore, all of the services for the VPC run as the users, so MySGE doesn't require any special privileges to run. The user even has the ability to change the scheduling policies and configuration within the VPC. This means the user can create special queues or change job priorities for jobs runs in the VPC.

MySGE is ideal for users who have already implemented tools to submit and monitor jobs for throughput environments and would like to easily execute these workflows on an HPC system. MySGE also allows users to dynamically allocate additional nodes to the VPC, allowing the system to be adjusted based on the compute needs. NERSC is also working on an implementation that will use the Torque resource manager which is called MyTorque.

File Cacher

One challenge to scaling up some data intensive workloads can come from reading-in large input or reference data sets. For example, when BLAST runs, the user specifies a reference database that is used to compare and align against. This reference database can be several gigabytes in size and is accessed via a memory-mapped file descriptor. At scale, this can place a large load on the file system, since the data is typically read in page size chunks (4 kilobytes) by all the compute cores in parallel.

To help address this, NERSC developed a file caching mechanism. The cacher starts by reading-in the input data prior to starting the application. Since the cacher can read the input data in large blocks, it can achieve very good I/O efficiency. The data is stored in a shared memory segment. A shared library is preloaded when executing the application. The shared library intercepts I/O calls to the cached files and reads the data directly from the shared memory space. Using this method, read-only data can be read-in at close to peak bandwidth of the system. This reduces the startup costs by over an order of magnitude at scale and significantly reduces the stress on the file system.

Ongoing Efforts

NERSC has developed tools like the task farmer, MySGE, and the file cacher to lower the barrier of entry and enable new user communities to easily make use of the incredible computing capabilities at NERSC. NERSC is investigating other methods to support these types of workloads on HPC platforms. For example, NERSC is evaluating Cray's Cluster Compute Mode (CCM) to see what role it can play for high-throughput, data intensive workloads. Given the growing need for computational cycles from genomics and other data-intensive workloads, extending the capabilities of HPC platforms to these communities will be critical to supporting scientific discovery for all of DOE.

Increasing Scientific Productivity by Tracking Data

New experimental instruments, supercomputers, and networks are contributing to record levels of scientific productivity. To effectively meet the increasing scientific demand for data storage systems and services, NERSC's staff must first understand how data moves within the facility. Until recently, the process of obtaining these insights was extremely tedious because the statistics came from multiple sources, including network router statistics, client and server transfer logs, storage and accounting reports—all saved as very large, independently formatted text files.

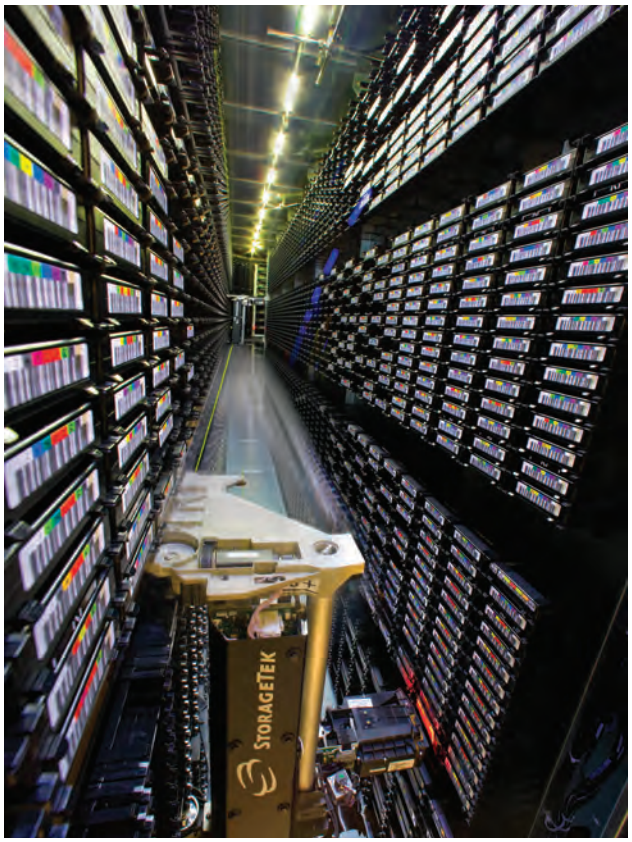
Now a dynamic database created by the NERSC Storage Systems Group continually collects statistics from all of these sources and compiles them into a single, searchable repository. The system also automatically generates daily email reports and graphs that illustrate how data moves in and out of the facility's HPSS archival storage system, which is the largest repository of scientific data at the center.

The daily reports help the Storage Systems Group understand the frequency, amount, and method of data movement between the archive and other systems in the center. Because 50 percent of all data movement activities within the center involve HPSS, this capability allows them to identify user bottlenecks and gives them an indication of how we should invest in storage solutions.

Strategy for Planning Hardware and Software

When all the Storage Systems Group had were text files of various daily logs and reports, identifying specific events was like looking for a needle in a haystack, and quantifying trends proved to be extremely time-consuming. But thanks to the new 50 gigabyte database, which contains several years of historical information with all the current information, a quick query will allow any analyst to instantly find a historical event—like when a new piece of hardware was installed—and quantify its usefulness within seconds.

For instance, one query showed that NERSC's Cray XT4 (Franklin) and XE6 (Hopper) systems contributed to



NERSC's HPSS storage system.

the largest amount of data movement within the storage archive in 2010, which was not surprising. But the tool also revealed that NERSC's Data Transfer Nodes (DTNs)—which are dedicated servers for performing transfers between local storage resources like HPSS and the NERSC Global Filesystem, and wide area resources like Leadership Computing Facilities at Argonne and Oak Ridge—are major data movement systems for HPSS. As a result, the Storage Systems Group determined that the center should invest in more DTNs to aid users.

Another aspect of the database that has been especially useful for storage analysts is the client software statistics, which look at the software used to transfer data to HPSS over time. These statistics have been critical in directing NERSC's efforts to improve software solutions in support of user data movement over time.

A query of the HPSS software used between May 1, 2010 and 2011 shows that the vast majority of data transfers to or from HPSS used the HSI client, which provides a UNIX-like interface into the system, or the HTAR client, which is similar to UNIX tar and is recommended for archiving small files. Due to the high usage of HSI on the Franklin and pre-production Hopper systems, the center made specific improvements in the way this client was deployed on Hopper. These improvements allowed users to achieve twice the bandwidth to HPSS for multi-gigabyte sized files.

Speeding Up Data Access for Users

For many data centers, tape libraries represent a cost, energy, and space efficient solution for storing ever-increasing amounts of scientific data. NERSC houses four SL8500 libraries, each composed of four library storage modules (LSMs), and each module contains about 2,500 tape cartridges and a collection of drives to read them. In total the center has about 40,000 tapes for user data archives and HPSS backups.

When a scientist makes a request to retrieve data from the tape library, a robot locates the tape where the requested information is stored, grabs it, and drops it into a tape drive that will read it to the user. This entire process occurs within a few seconds, unless all of the drives in that particular LSM are full. In that case, the robot will look for an available drive in another LSM.

Movements between LSMs are relatively slow because the cartridge either has to be deposited in an elevator and moved to another LSM within the same library, or go through a pass-through port into another library. The user will observe this as slower access time to their data. This really becomes a problem when the user is requesting data spread across numerous cartridges in LSMs with unavailable tape drives.

The new database allows staff to easily identify and monitor such events by including a section in the daily report that shows cartridge movements between LSMs. These reports allow storage analysts to track cartridge movements across LSMs daily, and determine if re-arranging cartridges and drive locations will speed up data access for the user.

Now that the usefulness of this database is becoming increasingly apparent, the Storage Systems Group would like to see the database grow to include statistics from the center's largest disk storage repository, the NERSC Global File system (NGF). The team would also like to design a graphical user interface (GUI) and analytics framework on top of the database so that all NERSC staff can use this information on demand for troubleshooting and decision-making.

Safeguarding Data with Parallel Incremental Backups

To ensure that users' data is always available when they need it, NERSC is extremely vigilant about regularly backing up persistent user data (that is, data that does not change from session to session) to tape. For many years, these backups were done with fairly basic

mechanisms like *dump*, *tar*, and *cpio*. But as the center transitioned to a shared *project* file system—which is mounted on all major NERSC systems to facilitate data sharing—it became increasingly clear that these methods could not keep up. That's when Matthew Andrews of the NERSC Storage Systems Group started developing the Parallel Incremental Backup System (PIBS).

The group realized that they needed to have daily backup to ensure that users have current and relevant backup data, and that any system used to do this needed to be fast and scalable to meet the demands of future growth.

The center-wide *project* file system at NERSC aims to provide well-performing, persistent, no-purge storage to projects where scientists are sharing both code and data. In 2006 when the system was initially installed, it had 80 TB capacity with only 20 TB of data stored. At this stage the traditional backups were still feasible.

However, just one year later the system grew to 180 TB capacity with 60 TB of data stored. At this point, Andrews and his colleagues in the Storage Systems Group had developed a few scripts to split up jobs by project, and then run them in parallel to back up files onto tape cartridges. This method was quite effective for small- to mid-range projects, but larger projects still took weeks to back up.

This is when they saw a need for a custom backup system for NERSC, and started developing the parallel backup system called PIBS. A single installation of the PIBS software can back up multiple file systems. Within PIBS, each file system is subdivided into logical subtrees or "backup sets." The purpose of these sets is to determine its priority in the daily backup schedule, and whether the system will do a full or incremental backup.

Multiple instances of a schedule type allow full backups of different directories at different frequencies, or on different days, thus ensuring that the backup is current and minimizes disruption to users. After all, backing up files should not take up half of the system's bandwidth all day.

Every day the file system metadata is scanned to determine what files have changed and need to be backed up. A set of daemons, or background programs, will then write *.tar* files using the lists generated in the metadata scan, and transfer daemons store the *.tar* files to HPSS.

Although the system was just fully implemented at the end of 2010, the Storage Systems Group has already achieved some major successes. In February 2011, when the Storage Systems Group upgraded to a new *project* file system, they were able to leverage the scalability of the new software by using it to restore 500 TB of data in just seven days with no disruption to users. While staff members set up the new file system, users continued to save to the existing *project* file system while PIBS continued to back up the new data to HPSS. When the new system was up and running, the NERSC team used the file system backup in HPSS to restore the data.

Recent improvements to the PIBS software should enable data to be restored at a rate of over 100 TB per day using 10 Oracle T10KB tape drives.

This system is truly an innovation from the NERSC Storage Systems Group. It used to take weeks to copy from an old to new *project* system, and the group was competing with users for bandwidth to do it. With PIBS they got the new system up and running in three days with near zero impact to users.

Data Transfer Nodes Facilitate Scientific Collaborations

The ability to reliably move and share data around the globe is essential to scientific collaboration. That's why three DOE scientific computing centers—the Argonne and Oak Ridge Leadership Computing Facilities and NERSC—have teamed up to focus on optimizing wide area network (WAN) transfers.

This ongoing effort began several years ago when each site deployed dedicated transfer nodes (DTNs), optimized for carrying data between the DOE facilities. Today, engineers from each site continue to meet regularly with DOE's Energy Sciences Network (ESnet) staff to develop strategies for optimizing bandwidth performance. This collaboration is called the Data Transfer Working Group, and their effort is having a huge impact on scientific research.

For example, Chesley McColl, Associate Scientist at the National Oceanic and Atmospheric Administration's (NOAA's) Earth System Research Laboratory, typically computes at one facility and analyzes data at another. He recently moved an 80 terabyte dataset from the high performance storage system at Oak Ridge to NERSC in just two months. Without the DTNs this would have taken as long as six months.

Because the DTNs mounted on NERSC's filesystems proved so useful to other researchers, the PDSF user community decided to mount one on their cluster too. PDSF is a networked distributed computing cluster designed primarily to meet the detector simulation and data analysis requirements of physics, astrophysics and nuclear science collaborations. Although PDSF is housed at NERSC, it is different from the center's other systems

because it does not rely on allocations. Instead, each collaboration or group receives shares in the batch system proportional to their contributions to shared resources like compute nodes or staffing. NERSC's DTNs are important benchmarks for the PDSF transfer nodes. Running WAN transfer tests on both sets of nodes and comparing them ensures that the PDSF hardware is working well.

Because the perfSONAR network monitoring applications are deployed on all of the DTNs, users can now identify "choke points" in their data movements and work with NERSC staff to fix them. In the past, there have been issues with WAN transfers to PDSF, and finding the problem required extensive effort. But with the DTNs' well defined and maintained configuration, if there is a problem, staff can look at the transfer and system logs, identify the problem, and fix it.

NERSC has also enabled Globus Online on their DTNs. This hosted service automates tasks associated with moving files between sites, or "endpoints," including retrying failed transfers, recovering from faults automatically whenever possible, and reporting status. So users simply log into the system, choose a start and end point, hit transfer, and the system does the rest.

Many scientists are familiar with tools like scp and GridFTP, but with Globus Online NERSC can offer a much simpler and faster method for moving data. Globus Online actually makes web-based data syncing an easy, nearly trivial process, so users don't have to be IT or middleware experts to move their files.

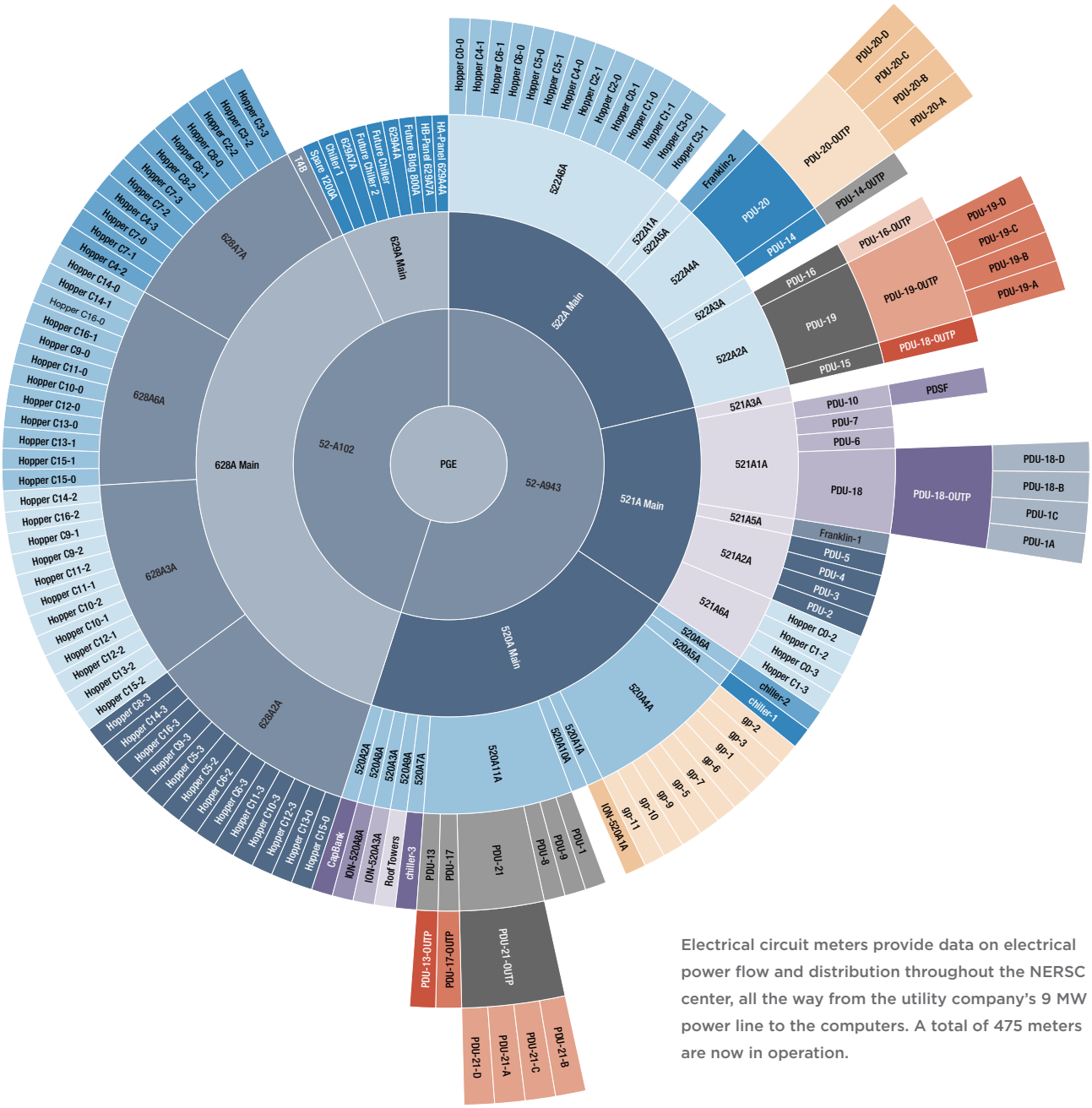
Increasing NERSC's Energy Efficiency

In FY2010, NERSC started instrumenting its machine room with state-of-the-art wireless monitoring technology to gather information on variables important to machine room operation, including air temperature, pressure and humidity, and chilled water flow and temperature. With the addition of new sensors for Hopper and Carver/Magellan, a total of 1,036 sensors are currently installed.

In 2011, NERSC added 475 electrical circuit meters to collect power usage data at four different levels of the Oakland Scientific Facility's electrical infrastructure (see diagram). Cray supercomputers also have their own built-in sensors to monitor power, temperature, and air pressure at various points in the cabinets.

NERSC's Outreach, Software, and Programming Group created database systems that collect and archive all of this environmental, power, and Cray data every five minutes; and they developed a web-based interface that provides a single view of the data.

This integrated monitoring infrastructure provides both real-time information to Operations staff and trend information that can be used for optimizing performance—for example, optimizing airflow management. This data is also being used by NERSC staff working with data center efficiency experts from Berkeley Lab's Environmental Energy and Technologies Division to analyze and optimize power usage. Future machine room design modifications will also take advantage of this data.



Electrical circuit meters provide data on electrical power flow and distribution throughout the NERSC center, all the way from the utility company's 9 MW power line to the computers. A total of 475 meters are now in operation.

User Support & Outreach

With over 550 projects using more than 500 different codes, NERSC is constantly working to improve the productivity of all 4,000 scientists and engineers who use the center. NERSC closely collaborates with users to improve their productivity and the performance of their applications. Software installed or optimized for one user or group is installed for the benefit of the entire NERSC community of users. Tools that are found to be useful for one group become supported by NERSC for all users.

This section presents a few examples of how NERSC's service-oriented efforts have benefitted its users, as indicated in the 2011 NERSC user survey. The 2011 survey garnered responses from 410 users, representing about 70 percent of all computational hours used at the center in 2011 through June 30. NERSC users' Overall Satisfaction score for 2011 was the highest ever recorded in the 13 years the survey has been in its current form: 6.54 out of a possible 7 points for "very satisfied."

Providing Rapid, Quality Front-Line User Support

The heart of NERSC's support services is its Consulting and Account Support staff. NERSC provides rapid responses and advice to hundreds of users each month. Relatively simple questions are answered immediately, while more complex and challenging problems develop into longer-term collaborations. In both cases—and everything in between—NERSC users benefit from NERSC technical staff's deep knowledge of high performance computing to overcome obstacles that otherwise might hold up research for months.

NERSC's consultants are on duty 8 a.m. to 5 p.m. Pacific Time during standard work days, immediately fielding phone calls, emails, and questions submitted through the web. During off hours, NERSC's Computer Operations and Networking staff is available 24x7 for password resets and system status inquiries.

The consultants work at the help desk for 4½-hour shifts on a rotating basis. Users appreciate instant access to HPC professionals because it allows them to overcome problems—sometimes simple ones—that could otherwise immediately block their work. Typical issues involve network connection and authentication problems, errors in batch scripts, quota issues, and allocation issues.

In addition to the outstanding satisfaction scores, users offered many comments on NERSC's services in response to the question "What does NERSC do best? How does NERSC distinguish itself from other computing centers you have used?" Some of the comments:

"NERSC's people are very talented and are on par with the staff at the best supercomputing centers in the world. I currently run on 8 petaflop/s worth of supercomputers on two continents, and I count NERSC among the top 3 places I use."

"NERSC consulting has been helpful with environment variables for my memory-intensive large runs."

"Great user support, software support is exhaustive on mature systems"

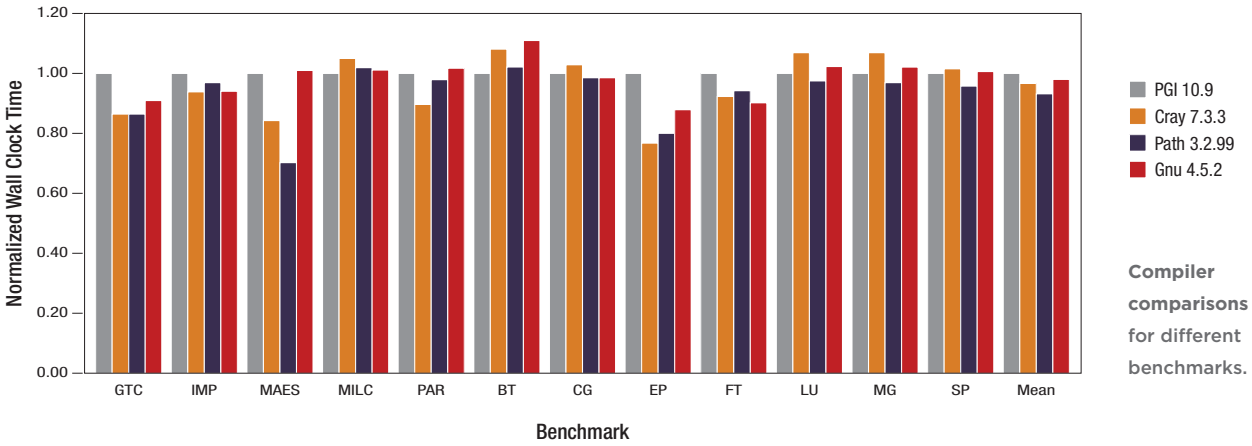
"NERSC has superb technical support staff. I'll never forget the experience of calling at 3 a.m. on a weekend for a password reset and getting a staff member right away who was much more awake than I was."

"The consulting is good and they always follow through. With webinars the training has become easily available."

"User Services and Viz/Analysis support are outstanding."

"Web pages are very useful, help line & support is friendly & reliable."

"Beyond simply furnishing computer time and resources, I have found that NERSC's consulting services are extremely helpful. Whenever I have run into a problem, the help desk has jumped in to investigate and almost always solved the issue within a day or two. For a large computing facility, this is a fabulous service. In addition, we recently had a couple of folks from NERSC come down to interview our group about our current and future computing needs to use in planning the next big cluster. This sort of user-centric planning is no doubt part of what makes NERSC great."



Compiler comparisons for different benchmarks.

"The staff is the most consistently helpful of any HPC systems I have ever used."

"When I eventually start up my own research group, getting resources at NERSC will be a top priority as the environment is extremely conductive to scientific research."

NERSC Users Welcome Expanded Training Efforts

As a direct result of a 2010 Operational Assessment recommendation, NERSC renewed its training efforts in 2011. In addition to its traditional training during the annual NERSC Users Group (NUG) Meeting, NERSC conducted a two-day workshop for Cray XE6 users at its facility in Oakland, joining with members of the Cielo team from Los Alamos National Laboratory and staff from Cray, Inc. Both the NUG training and the XE6 training were concurrently broadcast over the web.

In addition, NERSC held a number of web-based training events (webinars) through 2010–2011. In all, NERSC put on eight events for its users from July 1, 2010 to June 30, 2011, with an aggregate attendance of about 375. In addition, NERSC staff contributed to lectures and tutorials at UC Berkeley, at SC10, and at the Astrosim Summer School on Computational Astrophysics, Nicolaus Copernicus University, Torun, Poland.

NERSC's users responded positively to the training classes, as indicated by the satisfaction score increase of 0.51 points on the 2011 User Survey. Additional surveys were conducted after each class, with 97.8 percent of respondents indicating that the training was "useful to me."

All the presentation materials from the classes are available on the NERSC website, and video recordings of most sessions are also online.

Comparing Compilers to Improve Code Performance

Compilers can be said to be the most common software used by scientists on HPC systems. Every application that runs on the NERSC systems is making a use of a compiler. Despite the obvious importance of compilers, users often don't get enough guidance on how to choose the best one for their application or what compiler flags to use. The Hopper system currently has four compilers installed: Portland Group, Cray, Pathscale, and GNU. (Additionally, due to large demand, NERSC has recently purchased the Intel compilers for the Hopper system as well.) This past year NERSC did a study comparing compiler performance and compiler flags for different benchmarks. The intention of the study is to educate users on the available compilers and encourage them to try a new compiler for their application.

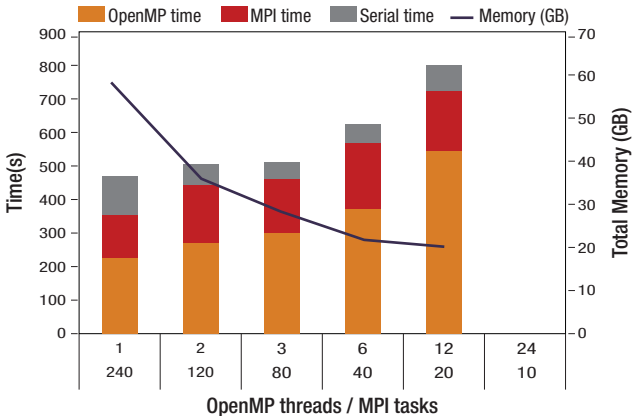
The chart "Compiler comparisons" shows the performance of micro-benchmarks and five full applications using the four compilers available on the Hopper system. The Portland Group compilers are the default on the Hopper system and are stable, reliable compilers that show solid performance for many applications. As the graph shows, however, a number of applications perform 20 and even 30 percent faster using alternative compilers. The MAESTRO code, for example, shows roughly an 18 percent performance improvement with the Cray compiler and over 30 percent improvement with the Pathscale compiler. The complete study is available to users on the NERSC website[1].

[1] http://www.nersc.gov/users/computational-systems/hopper/performance-and-optimization/compiler-comparisons/

Changing compilers is often the simplest step a user can take to improve an application's performance. Additionally, the large number of NERSC users makes it impossible to work with each scientist one-on-one for code optimization, and so having accessible tutorials and performance tips on the website is key.

Learning How to Use Hopper's Multi-Core Nodes Effectively

A combined team of NERSC and Cray staff has formed a Cray Center of Excellence to examine programming models beyond pure MPI on the 24-core Hopper nodes. One of the key findings with hybrid MPI-OpenMP codes is that although OpenMP may not alter the performance of an application very much, using OpenMP can dramatically decrease memory usage, allowing larger problems to be addressed. The core change in architecture of Hopper from that of earlier systems is the increasing number of cores per node and the decreasing amount of memory available per core. The majority of NERSC users run pure MPI applications. To ease the transition to systems with less memory available per task, NERSC has studied a number of benchmark codes and put together a series of tutorials, talks, and presentations to educate users on how to use the 24-core Hopper nodes effectively.



Performance of the fvCAM benchmark on Hopper as a function of MPI tasks and OpenMP threads—in other words, the performance of the code using a fixed number of nodes and differing amounts of MPI tasks and OpenMP threads. As one decreases the number of MPI tasks and increases the number of OpenMP threads, the overall memory usage decreases significantly. For example, using three threads and 80 MPI tasks, the performance decreases by only 6 percent compared to using single-threaded MPI tasks, whereas the memory requirement is reduced by almost 50 percent. Using six threads incurs a substantial performance penalty, in principle because of a growth in the time taken in OpenMP parallelized parts of the code. For 12 threads, there are also NUMA effects which add to the performance decrease.

As with the compiler comparison study, to reach a large number of users, NERSC creates online tutorials and presents optimization tips at workshops and conferences. The OpenMP study of CAM, GTC, Paratec, and PMAMR is available on the NERSC website.[2]

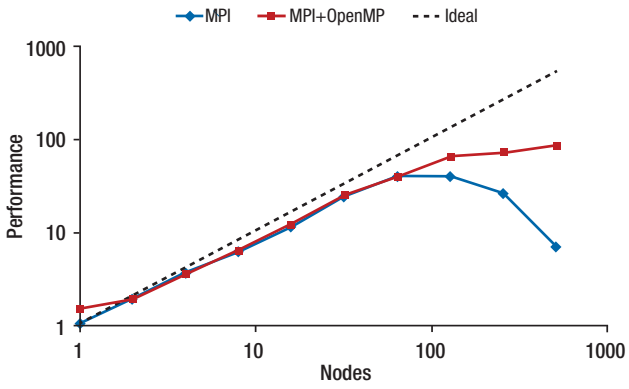
An example of a user application that has made use of the NERSC recommendations for running applications is Porous Media Adaptive Mesh Refinement (PMAMR), an application being used to model carbon sequestration and contaminant transport as part of the Advanced Simulation Capability for Environmental Management (ASCEM) project at Berkeley Lab. The goal of the ASCEM project is to better understand and quantify flow and contaminant transport behavior in complex geological systems. The PMAMR code is built on the BoxLib framework and was previously parallelized using MPI. In this work, hybrid MPI/OpenMP programming was added to the code to see what kind of performance gains could be achieved.

The figure below shows the results of adding OpenMP. The hybrid approach yields a net speedup of 2.6x compared to the MPI-only version, with substantially reduced memory usage, thus decreasing the time to solution by increasing the rate at which the calculation runs.

Facilitating Statistical Computing and Graphics Generation with Parallel R

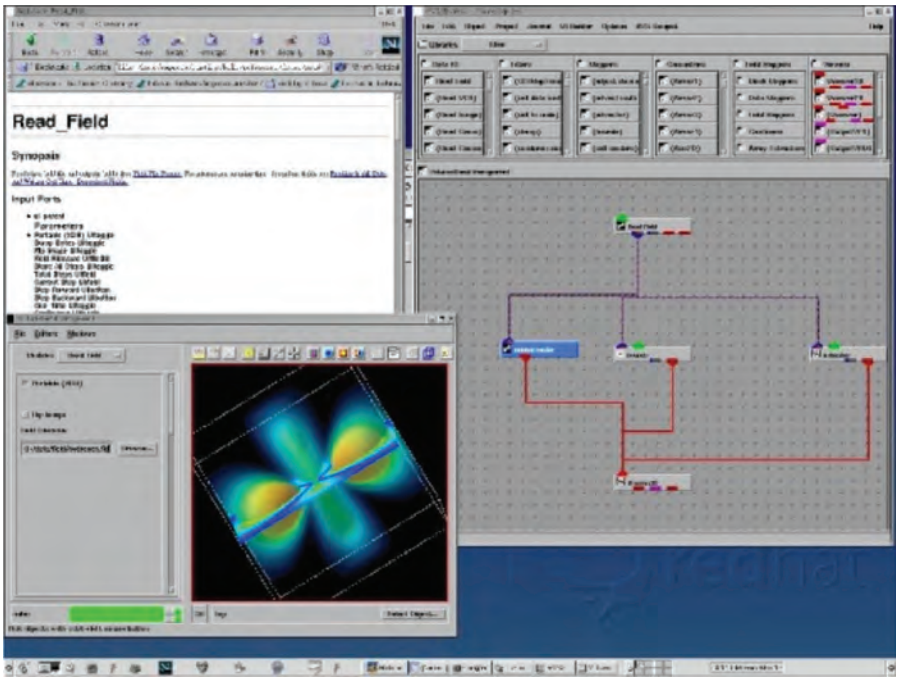
R is an integrated suite of software facilities for statistical computing, data analysis, and graphics. NERSC has provided R training at User Group meetings, maintains an on-line R tutorial and links to extensive documentation, and provides in-depth consulting services.

For example, David Romps of the UC Berkeley Department of Geology and Geophysics has been working with the RNetCDF and ncdf libraries, which are required for



Results of adding OpenMP to the PMAMR code.

[2] http://www.nersc.gov/users/computational-systems/hopper/performance-and-optimization/using-openmp-effectively-on-hopper/



AVS/Express is an X-Windows based application that is enabled for remote users by the NX X-Windows accelerator.

creation, access, and analysis of array-oriented scientific data in the NetCDF file format. Daniela Ushizima of the Analytics Group worked with Romps to customize R software installations on various NERSC machines.

Romps says, "Without the Data Analytics Team's help, I would not be able to run my cloud-resolving large-eddy simulations on Hopper. I use R to write the NetCDF initialization files (initial value data, forcing terms, etc.) that are read in by my LES model, Das Atmospherische Modell (DAM). I also use R+NetCDF to analyze the resulting NetCDF output and to generate figures for manuscripts. With the large number of cores available to me on Hopper, I was able to run simulations on 12,288 processors at a resolution that I could not dream of on the other machines available to me. These simulations performed on Hopper will be part of a paper soon to be submitted to the Journal of Atmospheric Sciences, which is tentatively titled 'Convective momentum transport and its parameterization.'"

Visualizing Electronic Structures from Quantum Monte Carlo Simulations

Prabhat of the NERSC Analytics Group has supported Michael Philpott of UC Berkeley in visualization and analysis of quantum Monte Carlo simulations of electronic structures. Philpott was interested in studying the magnetic properties of graphene as it was selectively modified by adding a hydrogen atom. Depending on the placement of the atom in various geometric locations of the structure, radically different magnetic properties were revealed by computer simulations. Prabhat assisted Philpott in making sense of the simulation output by producing visuals that would immediately and effectively convey the resulting spatial patterns of spin and bonding. The resulting work, titled "Magnetism and Bonding in Graphene Nanodots with H modified Interior, Edge and Apex," was recently accepted by the Journal of Chemical Physics.

Improving Access to X-Windows-Based Tools and Applications

AVS/Express is an X-Windows-based application that is enabled for remote users by the NX X-Windows accelerator.

X-Windows is a display system that enables many popular tools to have a graphical user interface for ease of use and interactivity. However, X uses is a very chatty protocol that is very sensitive to network latency. That makes applications of every moderate complexity difficult or impossible to use outside a local area network (LAN). But virtually every NERSC user accesses the center over a wide area network (WAN), which previously severely restricted their access to debuggers, performance tools, and data analysis and visualization applications.

In early 2011 NERSC put into production a service that enables excellent performance for X applications from virtually any geographic location. Very quickly after installation of a server and an enterprise version of a software package called NX, NERSC users were taking full advantage of the service. NERSC had been experimenting with NX for more than a year before putting it into full production.

This service opens an entirely new realm of visualization applications to all NERSC users, such as AVS/Express. In addition, users can remotely employ GUI-heavy, state-of-the-art debuggers like DDT and Totalview.

In the first weeks of service, more than 200 different NERSC users had connected to the NX server. At any given time, about 30 users had active sessions during June 2011.

User comments from the 2011 user survey included:

> *"The new NX is amazing, it is already more than I hoped to have."*
>
> *"The new NX server is wonderful; it changed my way of working."*
>
> *"The NX Client for visualization software is critical to make doing analysis at NERSC tractable."*

NERSC users have connected to the NX server from around the world, enabling them to use X applications that were previously unusable.

Improving the Efficiency of a User Code

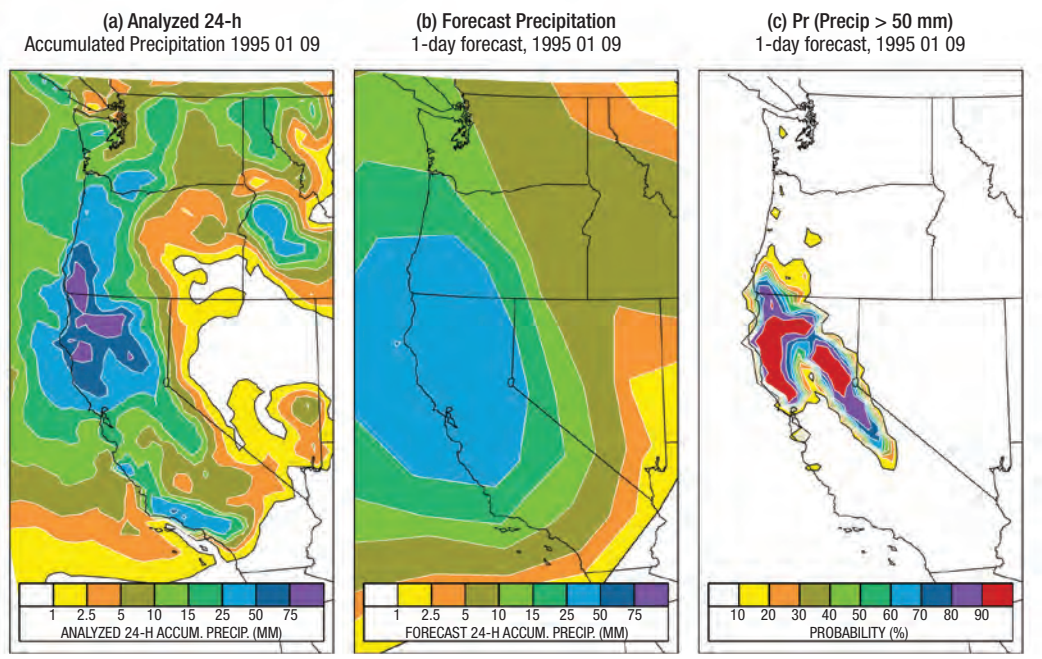
In addition to creating online tutorials and working with users one-to-one, NERSC consultants rapidly solve problems and collaborate with users to improve the efficiency of user codes. Collaborations with users can last a few hours or a few months, depending on the type of question the user has. Here is one example.

A group of users from the University of Utah and Argonne National Laboratory were testing MPI2's RMA (Remote Memory Access), often referred to as MPI's one-sided communication. Their code remaps local arrays to a global array, partitioned among a selected group of MPI aggregator tasks, using MPI's RMA functions. The users reported that the code ran fine up to 2,048 MPI tasks, but when 4,096 tasks were used on Hopper, the code failed.

A NERSC consultant extracted the communication information from more than 3,000 lines of code to create a simple test case that reproduced the problem and verified that the users had not made an error. After examining the communication pattern, the consultant suspected that there may have been insufficient internal MPI buffer space available. In particular, he noticed that many messages were being aggregated on the processor PE 0. During the remapping of data to a global array, all the MPI messages were directed to PE 0 during the first 12 out of 16 aggregation steps. Based on this observation, the consultant suggested a workaround where an additional synchronization was introduced so that the data sent up to that point could be cleared from PE 0's buffers. This approach was tested successfully on 4,096 cores. The consultant also filed a report with Cray to get the vendor's view on the problem and to determine if the issue should be filed as a bug in the implementation of the MPI library.

This example of a common interaction between a user and a NERSC consultant occurred over the period of a week. The user reported a problem running an application, and the consultant responded quickly with a workaround

NERSC users have connected to the NX server from around the world, enabling them to use X applications that were previously unusable.



One day's weather reforecast: (a) 24-hour observed precipitation amounts for January 9, 1995; (b) average one-day precipitation forecasts; (c) today's forecast calibrated with old reforecasts and precipitation analyses.

that allowed the user to continue running simulations. In the meantime, the consultant communicated with a vendor to analyze the problem and consider developing a longer-term solution.

Improving Account Support

NERSC has built a reputation as the computer center you can call on the phone and speak to someone who can help you right away. While continuing that tradition in 2011, we looked for ways to increase what users can do to help themselves via web interfaces. NERSC users may now reset their expired or forgotten passwords themselves by answering a set of predetermined security questions. NERSC also made it easier for first-time users to get up and running with their accounts. The initial account passwords which had been distributed by phone are now handled through secure links in an email. Another step in making services easier for web-savvy users was to allow scanned or photographed versions of the computer use policy form to be uploaded via the web. Previously only faxes were accepted.

Streamlined self-service web interfaces for NERSC users have added value to the center and reduced the number of emails, tickets, and staff hours required to help users be productive. In addition to being a convenience for NERSC users, these efficiency improvements help us provide quality service to a large numbers of users. If users find a self-service interface to account management useful, then NERSC staff have more time available to address problems for which human expertise is required.

Providing Comprehensive Support for an ALCC Project

The ASCR Leadership Computing Challenge (ALCC) project "A Multi-Decadal Reforecast Data Set to Improve Weather Forecasts for Renewable Energy" is an example of a project that uses a wide range of support services. Led by Thomas Hamill of NOAA, the project was awarded 14.5 million hours and to date has used over 8.5 million hours on Franklin and Hopper.

A reforecast is a retrospective forecast of the weather. Since numerical weather predictions are affected both by the chaotic growth of errors during the forecast and by model deficiencies, the use of direct model output for making weather-related decisions can lead to poor decisions. However, model forecast errors can be diagnosed from past forecasts and statistically corrected in post-processing, dramatically improving the forecast accuracy.

This project will generate a next-generation 30-year reforecast dataset, using the same modeling system that will be operational for the National Weather Service next year. The reforecast dataset will spur the development of novel longer-lead weather forecast applications for renewable energy (wind, solar, hydro), as well as improved probabilistic forecasts of rare extreme events.

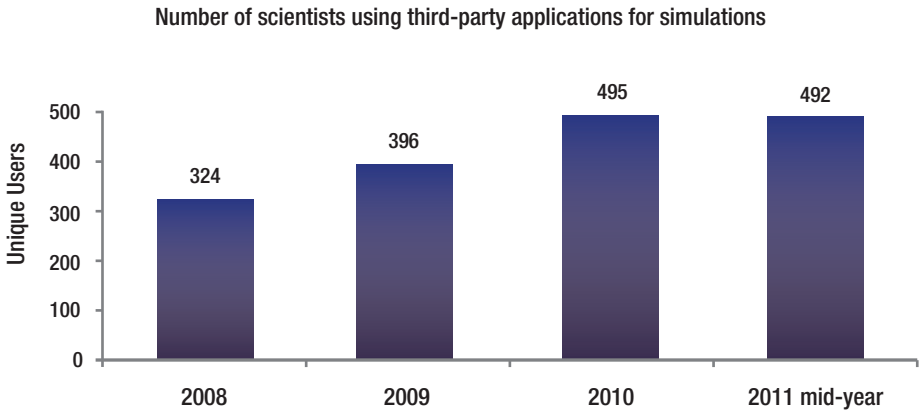
NERSC provided significant support in optimizing transfer of the initial data from the National Climatic Data Center in Asheville, North Carolina, to NERSC, and from NERSC to NOAA in Boulder, Colorado. NERSC's HPSS now serves as the master storage site for the full dataset.

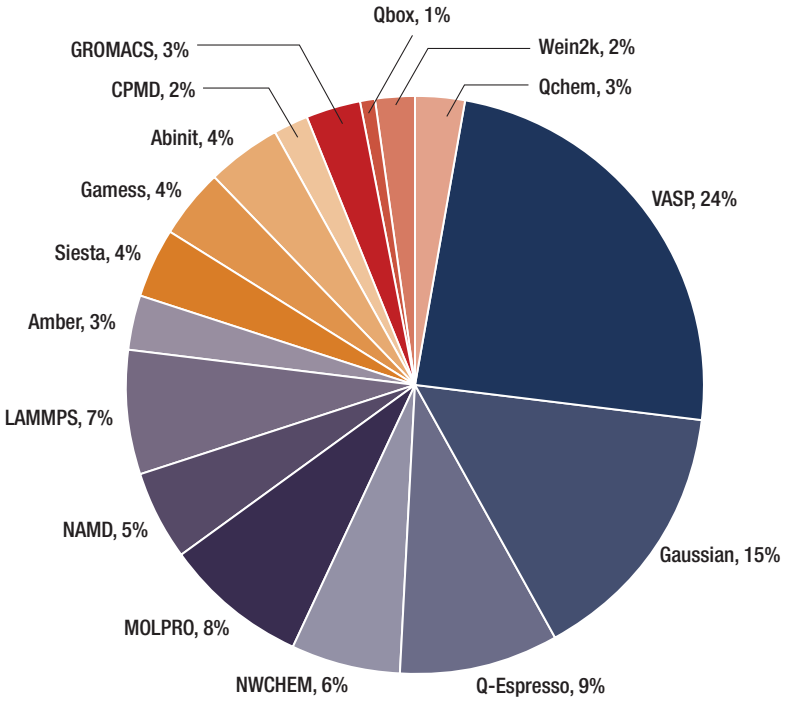
NERSC consultants rewrote part of the weather modeling code to enable porting from NOAA's IBMs to NERSC's Cray systems; provided tools to code developers to help them debug a memory problem; and improved throughput by assisting in redesigning the post-processing script and by providing a queue boost. After a slow start because of the code porting problem, the project was brought back on schedule.

Software Support for Third-Party Applications

NERSC compiles and supports a large number of software packages for our users, totaling over 13.5 million lines of source code. These software packages include debugging tools, analysis and post-processing software, and third-party applications scientists use to run their primary simulations. In the past few years, NERSC has seen an increase in the number of scientists using third-party applications. This is primarily due to the relative increase in the proportion of materials science and chemistry time allocated to NERSC projects and the popularity of third-party applications in these science areas.

Instead of each user installing his or her own version of a third-party software package, NERSC installs and supports optimized versions of many applications, allowing users to focus on their research rather than on installing and supporting software. The charts below show the growth in the number of users running third-party applications and the percent of usage for each application.





Third-party applications running at NERSC: growth since 2008 and percent of usage.

NERSC's new home page provides easy access to a wealth of information.

New Website Makes Information More Easily Accessible

NERSC's new home page provides easy access to a wealth of information.

NERSC's new, modern website (www.nersc.gov) was designed to improve how NERSC communicates with its users, DOE managers, peers in the HPC community, and the general public. Prominent sections include Live Status, Documentation for Users, Announcements, Science News, and more. The new website also contains presentations from training classes, video recordings of those classes, popular presentations given by NERSC staff, and workshop reports.

The 2011 NERSC user survey gives one example of the users' favorable response:

> *"NERSC's website has an enormous amount of good information and Google frequently takes me to it when I am looking for generic answers. The new website has been very easy to navigate."*

Research & Development by NERSC Staff

Staying ahead of the technological curve, anticipating problems, and developing proactive solutions are part of NERSC's culture. Many staff members collaborate on computer science research projects, scientific code development, and domain-specific research, as well as participating in professional organizations and conferences and contributing to journals and proceedings. The NERSC user community benefits from the results of these activities as they are applied to systems, software, and services at NERSC and throughout the HPC community.

This section describes the findings and recommendations of the Magellan Project, a two-year Department of Energy research effort into the potential benefits of cloud computing for science, followed by a list of NERSC staff publications and presentations.

Magellan at NERSC

Magellan Answers Questions about Cloud Computing for Science

After a two-year study of the feasibility of cloud computing systems for meeting the ever-increasing computational needs of scientists, Department of Energy researchers have issued a report stating that the cloud computing model is useful, but should not replace the centralized supercomputing centers operated by DOE national laboratories.

Cloud computing's ability to provide flexible, on-demand and cost-effective resources has found acceptance for enterprise applications, and as the largest funder of basic scientific research in the U.S., DOE was interested in whether this capability could translate to the scientific side. Launched in 2009, the study was carried out by the Argonne Leadership Computing Facility in Illinois and NERSC at Berkeley Lab. Called Magellan, the project used similar IBM computing clusters at the two labs. Scientific applications were run on the systems, as well as on commercial cloud offerings for comparison.

At the end of the two years, staff members from the two centers produced a 169-page report, *The Magellan Report on Cloud Computing for Science*, with a number of findings and recommendations.[3] Overall, the project members found that while commercial clouds are well suited for enterprise applications, scientific applications are more computationally demanding, and therefore the computing systems require more care and feeding. In short, the popular "plug and play" concept of cloud computing does not carry over to scientific computing.

To thoroughly evaluate the cloud computing model, the team ran scientific applications for studying particle physics, climate, quantum chemistry, plasma physics and astrophysics on the Magellan system, as well as on a Cray XT4 supercomputer and a Dell cluster system. The applications were also run on Amazon's EC2 commercial cloud offering, including the HPC offering with Cluster Compute instances, for comparison purposes.

[3] http://www.nersc.gov/assets/StaffPublications/2012/MagellanFinalReport.pdf

Many of the cost benefits from clouds result from the increased consolidation and higher average utilization. Because existing DOE centers are already consolidated and typically have high average utilization, they are usually cost effective when compared with public clouds.

"Our analysis shows that DOE centers are often three to four times less expensive than typical commercial offerings," the authors wrote in their report. "These cost factors include only the basic, standard services provided by commercial cloud computing, and do not take into consideration the additional services such as user support and training that are provided at supercomputing centers today and are essential for scientific users who deal with complex software stacks and require help with optimizing their codes."

A key characteristic of many scientific applications is that processes or phenomena are modeled in three dimensions over time, such as fuel burning in an engine, fluid flowing over different surfaces, or climate changing over years or decades. In order to create realistic simulations, the applications often run on hundreds or thousands of processors in parallel and constantly communicate with one another, sharing data. Enterprise applications typically analyze data in sequence without much inter-processor communication, and many commercial cloud systems are designed to handle this type of computing.

The Magellan project also looked at a number of other concerns specific to the Department of Energy and scientific computing:

- Can DOE cyber security requirements be met within a cloud?
- Can DOE HPC applications run efficiently in the cloud? What applications are suitable for clouds?
- How usable are cloud environments for scientific applications?
- When is it cost effective to run DOE HPC science in a cloud?

Over the course of the project, a number of scientific applications successfully used the Magellan system at NERSC. These applications include the STAR (Solenoid Tracker at RHIC) at Brookhaven National Laboratory, the Materials Project led by Berkeley Lab, the National Science Foundation's Laser Interferometer Gravitational-Wave Observatory (LIGO), and the Integrated Microbial Genomes database at the Joint Genome Institute. As an example, access to the two Magellan systems helped physicists at STAR assess whether cloud computing was a cost-effective alternative to using systems with much longer lead times. In their quest to understand proton spin, the scientists were able to create a real-time data processing system on the cloud, providing their research community with faster access to experimental data.

In the end, the Magellan research teams came up with a set of key findings:

- Scientific applications have special requirements that require cloud solutions that are tailored to these needs.
- The scientific applications currently best suited for clouds are those with minimal communication and I/O (input/output).
- Clouds can require significant programming and system administration support.
- Significant gaps and challenges exist in current open-source virtualized cloud software stacks for production science use.
- Clouds expose a different risk model, requiring different security practices and policies.
- The MapReduce programming model shows promise in addressing scientific needs, but current implementations have gaps and challenges.
- Public clouds can be more expensive than in-house large systems. Many of the cost benefits from clouds result from the increased consolidation and higher average utilization.
- DOE supercomputing centers already achieve energy efficiency levels comparable to commercial cloud centers.
- Cloud is a business model and can be applied at DOE supercomputing centers.

The progress of the Magellan study was closely watched by the scientific computing community. Berkeley Lab computer scientists who worked on the project presented their work at workshops over the course of the Magellan project. The team received "best paper" awards for "I/O Performance of Virtualized Cloud Environments,"[4]

[4] http://datasys.cs.iit.edu/events/DataCloud-SC11/p08.pdf

presented at the Second International Workshop on Data Intensive Computing in the Clouds (DataCloud-SC11) in November 2011; "Performance Analysis of High Performance Computing Applications on the Amazon Web Services Cloud,"[5] presented at the IEEE International Conference on Cloud Computing Technology and Science (CloudCom 2010) in November 2010; and "Seeking Supernovae in the Clouds: A Performance Study,"[6] presented at ScienceCloud 2010, the 1st Workshop on Scientific Cloud Computing, in June 2010.

In February 2012, NERSC Division Director Kathy Yelick gave a presentation on the Magellan project findings and recommendations to the Computer Science and Telecommunications Board (CSTB) of the National Academies, which includes the National Academy of Engineering, National Academy of Sciences, and the Institute of Medicine. CSTB is composed of nationally recognized experts from across the information technology fields and complementary fields germane to the Board's interests in IT and society.

X-Ray Image Bank Open for Business

Filipe Maia is building a data bank where scientists from around the world can deposit and share images generated by coherent x-ray light sources. A post-doctoral researcher with NERSC, Maia hopes the Coherent X-Ray Imaging Data Bank, or CXIDB,[7] can help researchers make the most of their valuable data.

CXIDB
Coherent X-ray Imaging Data Bank

Scientists use light sources to shoot intense x-ray beams into molecules, such as proteins, in order to understand their shapes and structures. The resulting diffraction patterns are painstakingly reconstructed to deduce an image. "It kind of works like a microscope, but it has no lens," Maia says. He works on the analysis software that reassembles diffraction patterns into an image—a software lens, of sorts.

"Until now, very few groups have had access to this kind of data," says Maia. He hopes the data bank will "maximize the impact of imaging data by making it easy to distribute to the community." As a result, "researchers who may not have access to experiments can use it for testing new ideas, improving on an analysis, or other things we can't imagine now," Maia explains.

To get the ball rolling, the NERSC Petascale Postdoc became the first depositor. He uploaded data from one of two papers he coauthored in the February 3, 2011 issue of the journal Nature. These first images involved the historic, direct x-ray imaging of a living organism, a virus, at the Linear Coherent Light Source (LCLS) at Stanford.

As the next generation of x-ray sources goes online, sharing data will become a necessity, says Maia. "Free electron laser facilities, like the LCLS, can produce up to 20 terabytes of data in one day, and the coming European XFEL (X-Ray Free Electron Laser) is expected to do 500 times that," he says. "We're going to have to distribute the analysis of that much data to make it manageable, and this repository could be a key part of that process."

NERSC Staff Publications

Publications and presentations by NERSC staff in 2011 are listed below. (Not all co-authors are from NERSC.) Many of these papers are available online at http://www.nersc.gov/news-publications/publications-reports/nersc-staff-publications-and-presentations/.

Jihan Kim, Alice Koniges, Richard Martin, Maciej Harancyzk, Joseph Swisher, Berend Smit, **"GPU Computational Screening of Carbon Capture Materials,"** Proceedings of the 2011 SciDAC Conference, 2011.

Jihan Kim, Jocelyn Rodgers, Manuel Athenes, Berend Smit, **"Molecular Monte Carlo Simulations Using Graphics Processing Units: To Waste Recycle or Not?"** Journal of Chemical Theory and Computation, 2011.

Thomas, R. C.; Aldering, G.; Antilogus, P.; Aragon, C.; Bailey, S.; Baltay, C.; Bongard, S.; Buton, C.; Canto, A.; Childress, M.; Chotard, N.; Copin, Y.; Fakhouri, H. K.; Gangler, E.; Hsiao, E. Y.; Kerschhaggl, M.; Kowalski, M.; Loken, S.; Nugent, P.; Paech, K.; Pain, R.; Pecontal, E.; Pereira, R.; Perlmutter, S.; Rabinowitz, D.; Rigault, M.; Rubin, D.; Runge, K.; Scalzo, R.; Smadja, G.; Tao, C.; Weaver, B. A.; Wu, C.; (The Nearby Supernova Factory); Brown, P. J.; Milne, P. A., **"Type Ia Supernova Carbon Footprints,"** Astrophysical Journal, December 2011, 743:27.

Zhengji Zhao and Helen He, **Cray Cluster Compatibility Mode on Hopper,** A Brown Bag Lunch talk at NERSC, Dec. 8, 2011, Oakland, CA, December 8, 2011.

[5] http://www.lbl.gov/cs/CSnews/cloudcomBP.pdf

[6] http://datasys.cs.iit.edu/events/ScienceCloud2010/p07.pdf

[7] http://www.cxidb.org

K. Antypas, "**Parallel I/O From a User's Perspective,"** HPC Advisory Council, December 6, 2011.

Li, Weidong; Bloom, Joshua S.; Podsiadlowski, Philipp; Miller, Adam A.; Cenko, S. Bradley; Jha, Saurabh W.; Sullivan, Mark; Howell, D. Andrew; Nugent, Peter E.; Butler, Nathaniel R.; Ofek, Eran O.; Kasliwal, Mansi M.; Richards, Joseph W.; Stockton, Alan; Shih, Hsin-Yi; Bildsten, Lars; Shara, Michael M.; Bibby, Joanne; Filippenko, Alexei V.; Ganeshalingam, Mohan; Silverman, Jeffrey M.; Kulkarni, S. R.; Law, Nicholas M.; Poznanski, Dovi; Quimby, Robert M.; McCully, Curtis; Patel, Brandon; Maguire, Kate; Shen, Ken J., **"Exclusion of a luminous red giant as a companion star to the progenitor of supernova SN 2011fe,"** Nature, December 2011, 480:348-350.

M. Di Pierro, J. Hetrick, S. Cholia, D. Skinner, **"Making QCD Lattice Data Accessible and Organized through Advanced Web Interfaces,"** Arxiv preprint arXiv:1112.2193, December 1, 2011, abs/1112.

Richard A. Gerber, Harvey J. Wasserman, **"Large Scale Computing and Storage Requirements for Fusion Energy Sciences,"** Workshop, December 2011.

Iair Arcavi, Avishay Gal-Yam, Ofer Yaron, Assaf Sternberg, Itay Rabinak, Eli Waxman, Mansi M. Kasliwal, Robert M. Quimby, Eran O. Ofek, Assaf Horesh, Shrinivas R. Kulkarni, Alexei V. Filippenko, Jeffrey M. Silverman, S. Bradley Cenko, Weidong Li, Joshua S. Bloom, Mark Sullivan, Peter E. Nugent, Dovi Poznanski, Evgeny Gorbikov, Benjamin J. Fulton, D. Andrew Howell, David Bersier, Amedee Riou, Stephane Lamotte-Bailey, Thomas Griga, Judith G. Cohen, Stephan Hachinger, David Polishook, Dong Xu, Sagi Ben-Ami, Ilan Manulis, Emma S. Walker, Kate Maguire, Yen-Chen Pan, Thomas Matheson, Paolo A. Mazzali, Elena Pian, Derek B. Fox, Neil Gehrels, Nicholas Law, Philip James, Jonathan M. Marchant, Robert J. Smith, Chris J. Mottram, Robert M. Barnsley, Michael T. Kandrashoff and Kelsey I. Clubb, **"SN 2011dh: Discovery of a Type IIb Supernova from a Compact Progenitor in the Nearby Galaxy M51,"** Astrophysical Journal, December 2011, 742:L18.

Nugent, Peter E.; Sullivan, Mark; Cenko, S. Bradley; Thomas, Rollin C.; Kasen, Daniel; Howell, D. Andrew; Bersier, David; Bloom, Joshua S.; Kulkarni, S. R.; Kandrashoff, Michael T.; Filippenko, Alexei V.; Silverman, Jeffrey M.; Marcy, Geoffrey W.; Howard, Andrew W.; Isaacson, Howard T.; Maguire, Kate; Suzuki, Nao; Tarlton, James E.; Pan, Yen-Chen; Bildsten, Lars; Fulton, Benjamin J.; Parrent, Jerod T.; Sand, David; Podsiadlowski, Philipp; Bianco, Federica B.; Dilday, Benjamin; Graham, Melissa L.; Lyman, Joe; James, Phil; Kasliwal, Mansi M.; Law, Nicholas M.; Quimby, Robert M.; Hook, Isobel M.; Walker, Emma S.; Mazzali, Paolo; Pian, Elena; Ofek, Eran O.; Gal-Yam, Avishay; Poznanski, Dovi, **"Supernova SN 2011fe from an exploding carbon-oxygen white dwarf star,"** Nature, December 2011, 480:344-347.

Katherine Yelick, Susan Coghlan, Brent Draney, Richard Shane Canon, Lavanya Ramakrishnan, Adam Scovel, Iwona Sakrejda, Anping Liu, Scott Campbell, Piotr T. Zbiegiel, Tina Declerck, Paul Rich, **"The Magellan Report on Cloud Computing for Science,"** U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR), December 2011.

Vladimir Getov, Adolfy Hoisie, Harvey J. Wasserman, **"Codesign for Systems and Applications: Charting the Path to Exascale Computing,"** guest editors' introduction, IEEE Computer, November 2011.

Scott Campbell, Jason Lee, **"Intrusion Detection at 100G,"** The International Conference for High Performance Computing, Networking, Storage, and Analysis, November 14, 2011.

M. Cary, J. Hick, A. Powers, **HPC Archive Solutions Made Simple,** Half-day Tutorial at Super Computing (SC11), November 13, 2011.

Glenn Brook, Donald Frederick, Richard Gerber, Jeff Larkin, **"Scaling to Petascale and Beyond: Performance Analysis and Optimization of Applications,"** (Tutorial), SC11, Seattle, WA, November 2011.

Samuel Williams, Oliker, Carter, John Shalf, **"Extracting ultra-scale Lattice Boltzmann performance via hierarchical and distributed auto-tuning,"** SC, November, 2011, 55.

Kamesh Madduri, Z. Ibrahim, Williams, Im, Ethier, Shalf, Leonid Oliker, **"Gyrokinetic toroidal simulations on leading multi- and manycore HPC systems,"** SC, November, 2011, 23.

Corsi, A.; Ofek, E. O.; Frail, D. A.; Poznanski, D.; Arcavi, I.; Gal-Yam, A.; Kulkarni, S. R.; Hurley, K.; Mazzali, P. A.; Howell, D. A.; Kasliwal, M. M.; Green, Y.; Murray, D.; Sullivan, M.; Xu, D.; Ben-ami, S.; Bloom, J. S.; Cenko, S. B.; Law, N. M.; Nugent, P.; Quimby, R. M.; Pal'shin, V.; Cummings, J.; Connaughton, V.; Yamaoka, K.; Rau, A.; Boynton, W.; Mitrofanov, I.; Goldsten, J., **"PTF 10bzf (SN 2010ah): A Broad-line Ic Supernova Discovered by the Palomar Transient Factory,"** Astrophysical Journal, November 2011, 741:76.

Robert Preissl, Wichmann, Long, Shalf, Ethier, Alice E. Koniges, **"Multithreaded global address space communication techniques for gyrokinetic fusion applications on ultra-scale platforms,"** SC, November 1, 2011, 15.

Jens Krueger, Donofrio, Shalf, Mohiyuddin, Williams, Oliker, Franz-Josef Pfreund, **"Hardware/software co-design for energy-efficient seismic modeling,"** SC, November 1, 2011, 73.

Ghoshal, Devarshi and Canon, Richard Shane and Ramakrishnan, Lavanya, **"Understanding I/O Performance of Virtualized Cloud Environments,"** The Second International Workshop on Data Intensive Computing in the Clouds (DataCloud-SC11), 2011.

Zhengji Zhao, Mike Davis, Katie Antypas, Rei Lee and Tina Butler, "**Shared Library Performance on Hopper,"** Oct. 26, 2011, Cray Quarterly Meeting at St Paul, MN, October 26, 2011.

Richard Shane Canon, **"Exploiting HPC Platforms for Metagenomics: Challenges and Opportunities,"** Metagenomics Informatics Challenges Workshop, October 12, 2011.

J. Hick, "**Digital Archiving and Preservation in Government Departments and Agencies,"** Oracle Open World 2011, October 6, 2011.

Lavanya Ramakrishnan & Shane Canon, **"NERSC, Hadoop and Pig Overview,"** October 2011.

Agüeros, Marcel A.; Covey, Kevin R.; Lemonias, Jenna J.; Law, Nicholas M.; Kraus, Adam; Batalha, Natasha; Bloom, Joshua S.; Cenko, S. Bradley; Kasliwal, Mansi M.; Kulkarni, Shrinivas R.; Nugent, Peter E.; Ofek, Eran O.; Poznanski, Dovi; Quimby, Robert M., **"The Factory and the Beehive. I. Rotation Periods for Low-mass Stars in Praesepe,"** Astrophysical Journal, October 2011, 740:110.

Lavanya Ramakrishnan, Richard Shane Canon, Krishna Muriki, Iwona Sakrejda, and Nicholas J. Wright., **"Evaluating Interconnect and Virtualization Performance for High Performance Computing,"** Proceedings of 2nd International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computing Systems (PMBS11), 2011.

Levitan, David; Fulton, Benjamin J.; Groot, Paul J.; Kulkarni, Shrinivas R.; Ofek, Eran O.; Prince, Thomas A.; Shporer, Avi; Bloom, Joshua S.; Cenko, S. Bradley; Kasliwal, Mansi M.; Law, Nicholas M.; Nugent, Peter E.; Poznanski, Dovi; Quimby, Robert M.; Horesh, Assaf; Sesar, Branimir; Sternberg, Assaf, **"PTF1 J071912.13+485834.0: An Outbursting AM CVn System Discovered by a Synoptic Survey,"** Astrophysical Journal, October 2011, 739:68.

Cano, Z.; Bersier, D.; Guidorzi, C.; Kobayashi, S.; Levan, A. J.; Tanvir, N. R.; Wiersema, K.; D'Avanzo, P.; Fruchter, A. S.; Garnavich, P.; Gomboc, A.; Gorosabel, J.; Kasen, D.; Kopač, D.; Margutti, R.; Mazzali, P. A.; Melandri, A.; Mundell, C. G.; Nugent, P. E.; Pian, E.; Smith, R. J.; Steele, I.; Wijers, R. A. M. J.; Woosley, S. E., **"XRF 100316D/SN 2010bh and the Nature of Gamma-Ray Burst Supernovae,"** Astrophysical Journal, October 2011, 740:41.

J. Hick, J. Hules, A. Uselton, **"DOE HPC Best Practices Workshop: File Systems and Archives,"** Workshop, September 27, 2011.

Richard A. Gerber, **"Introduction to HPC Systems,"** NERSC New User Training, September 13, 2011.

J. Hick, **"The NERSC Global Filesystem (NGF),"** Computing in Atmospheric Sciences 2011 (CAS2K11), September 13, 2011.

Richard A. Gerber, **"Experiences with Tools at NERSC, Programming weather, climate, and earth-system models on heterogeneous multi-core platorms,"** NCAR, Boulder, CO, September 7, 2011.

Krisciunas, Kevin; Li, Weidong; Matheson, Thomas; Howell, D. Andrew; Stritzinger, Maximilian; Aldering, Greg; Berlind, Perry L.; Calkins, M.; Challis, Peter; Chornock, Ryan; Conley, Alexander; Filippenko, Alexei V.; Ganeshalingam, Mohan; Germany, Lisa; González, Sergio; Gooding, Samuel D.; Hsiao, Eric; Kasen, Daniel; Kirshner, Robert P.; Howie Marion, G. H.; Muena, Cesar; Nugent, Peter E.; Phelps, M.; Phillips, Mark M.; Qiu, Yulei; Quimby, Robert; Rines, K.; Silverman, Jeffrey M.; Suntzeff, Nicholas B.; Thomas, Rollin C.; Wang, Lifan, **"The Most Slowly Declining Type Ia Supernova 2001ay,"** Astrophysical Journal, September 2011, 142:74.

Gal-Yam, Avishay; Kasliwal, Mansi M.; Arcavi, Iair; Green, Yoav; Yaron, Ofer; Ben-Ami, Sagi; Xu, Dong; Sternberg, Assaf; Quimby, Robert M.; Kulkarni, Shrinivas R.; Ofek, Eran O.; Walters, Richard; Nugent, Peter E.; Poznanski, Dovi; Bloom, Joshua S.; Cenko, S. Bradley; Filippenko, Alexei V.; Li, Weidong; Silverman, Jeffrey M.; Walker, Emma S.; Sullivan, Mark; Maguire, K.; Howell, D. Andrew; Mazzali, Paolo A.; Frail, Dale A.; Bersier, David; James, Phil A.; Akerlof, C. W.; Yuan, Fang; Law, Nicholas; Fox, Derek B.; Gehrels, Neil, **"Real-time Detection and Rapid Multiwavelength Follow-up Observations of a Highly Subluminous Type II-P Supernova from the Palomar Transient Factory Survey,"** Astrophysical Journal, August 2011, 736:159.

van Eyken, Julian C.; Ciardi, David R.; Rebull, Luisa M.; Stauffer, John R.; Akeson, Rachel L.; Beichman, Charles A.; Boden, Andrew F.; von Braun, Kaspar; Gelino, Dawn M.; Hoard, D. W.; Howell, Steve B.; Kane, Stephen R.; Plavchan, Peter; Ramírez, Solange V.; Bloom, Joshua S.; Cenko, S. Bradley; Kasliwal, Mansi M.; Kulkarni, Shrinivas R.; Law, Nicholas M.; Nugent, Peter E.; Ofek, Eran O.; Poznanski, Dovi; Quimby, Robert M.; Grillmair, Carl J.; Laher, Russ; Levitan,

David; Mattingly, Sean; Surace, Jason A., **"The Palomar "Transient Factory Orion Project: Eclipsing Binaries and Young Stellar Objects,"** Astrophysical Journal, August 2011, 142:60.

Lavanya Ramakrishnan, Piotr T. Zbiegel, Scott Campbell, Rick Bradshaw, Richard Shane Canon, Susan Coghlan, Iwona Sakrejda, Narayan Desai, Tina Declerck, Anping Liu, **"Magellan: Experiences from a Science Cloud,"** Proceedings of the 2nd International Workshop on Scientific Cloud Computing, ACM ScienceCloud '11, Boulder, Colorado, and New York, NY, 2011, 49 - 58.

Levan, A. J.; Tanvir, N. R.; Cenko, S. B.; Perley, D. A.; Wiersema, K.; Bloom, J. S.; Fruchter, A. S.; Postigo, A. de Ugarte; O'Brien, P. T.; Butler, N.; van der Horst, A. J.; Leloudas, G.; Morgan, A. N.; Misra, K.; Bower, G. C.; Farihi, J.; Tunnicliffe, R. L.; Modjaz, M.; Silverman, J. M.; Hjorth, J.; Thöne, C.; Cucchiara, A.; Cerón, J. M. Castro; Castro-Tirado, A. J.; Arnold, J. A.; Bremer, M.; Brodie, J. P.; Carroll, T.; Cooper, M. C.; Curran, P. A.; Cutri, R. M.; Ehle, J.; Forbes, D.; Fynbo, J.; Gorosabel, J.; Graham, J.; Hoffman, D. I.; Guziy, S.; Jakobsson, P.; Kamble, A.; Kerr, T.; Kasliwal, M. M.; Kouveliotou, C.; Kocevski, D.; Law, N. M.; Nugent, P. E.; Ofek, E. O.; Poznanski, D.; Quimby, R. M.; Rol, E.; Romanowsky, A. J.; Sánchez-Ramírez, R.; Schulze, S.; Singh, N.; van Spaandonk, L.; Starling, R. L. C.; Strom, R. G.; Tello, J. C.; Vaduvescu, O.; Wheatley, P. J.; Wijers, R. A. M. J.; Winters, J. M.; Xu, D., **"An Extremely Luminous Panchromatic Outburst from the Nucleus of a Distant Galaxy,"** Science, July 2011, 333:199-.

Kleiser, Io K. W.; Poznanski, Dovi; Kasen, Daniel; Young, Timothy R.; Chornock, Ryan; Filippenko, Alexei V.; Challis, Peter; Ganeshalingam, Mohan; Kirshner, Robert P.; Li, Weidong; Matheson, Thomas; Nugent, Peter E.; Silverman, Jeffrey M., **"Peculiar Type II supernovae from blue supergiants,"** Monthly Notices of the Royal Astronomical Society, July 2011, 415:372-382.

Shane Canon, **"Debunking Some Common Misconceptions of Science in the Cloud,"** ScienceCloud 2011, June 29, 2011.

"NERSC Overview for Environmental Energy Technologies," Berkeley Lab, Environmental Energy Technologies Division, Berkeley, CA, June 2011.

T. Nightingale, **"No Cost ZFS On Low Cost Hardware,"** NERSC High Performance Computing Seminar, June 14, 2011.

Richard A. Gerber, Harvey J. Wasserman, **"Large Scale Computing and Storage Requirements for Basic Energy Sciences,"** Workshop, June 10, 2011, LBNL LBNL-4809E

J. Hick, M. Andrews, **"Leveraging the Business Value of Tape,"** FujiFilm Executive IT Summit 2011, June 9, 2011.

"Getting Started at NERSC," NERSC Training Webinar, Berkeley Lab OSF, Oakland, CA, June 7, 2011.

Quimby, R. M.; Kulkarni, S. R.; Kasliwal, M. M.; Gal-Yam, A.; Arcavi, I.; Sullivan, M.; Nugent, P.; Thomas, R.; Howell, D. A.; Nakar, E.; Bildsten, L.; Theissen, C.; Law, N. M.; Dekany, R.; Rahmer, G.; Hale, D.; Smith, R.; Ofek, E. O.; Zolkower, J.; Velur, V.; Walters, R.; Henning, J.; Bui, K.; McKenna, D.; Poznanski, D.; Cenko, S. B.; Levitan, D., **"Hydrogen-poor superluminous stellar explosions,"** Nature, June 2011, 474:487-489.

Zhengji Zhao and Nick Wright, **"Performance of Density Functional Theory codes on Cray XE6,"** A paper presented in the Cray User Group meeting, May 23-26, 2011, Fairbanks, Alaska., May 24, 2011.

K. Antypas, Y. He, "**Transitioning Users from the Franklin XT4 System to the Hopper XE6 System,"** CUG Procceedings, Fairbanks, Alaska, May 23, 2011.

P. M. Stewart, Y. He, **"Benchmark Performance of Different Compilers on a Cray XE6,"** Fairbanks, AK, CUG Proceedings, May 23, 2011.

"Programming Environment, Applications, and Documentation" SIG Group, CUG 2011 Special Interest Group, Fairbanks, AK., May 23, 2011.

Michael Stewart, Yun (Helen) He, **"Benchmark Performance of Different Compilers on a Cray XE6,"** Cray User Group 2011, Fairbanks, AK, May 23, 2011.

Katie Antypas, Yun (Helen) He, **"Transitioning Users from the Franklin XT4 System to the Hopper XE6 System,"** Cray User Group 2011, Fairbanks, AK, May 23, 2011.

J. Hick, **"Storage Supporting DOE Science,"** Preservation and Archiving Special Interest Group (PASIG) 2011, May 12, 2011.

"NERSC Overview for the Joint Genome Institute," DOE Joint Genome Institute, Walnut Creek, CA, May 2, 2011.

N. Balthaser, D. Hazen, **"HSI Best Practices for NERSC Users,"** May 2, 2011.

Chotard, N.; Gangler, E.; Aldering, G.; Antilogus, P.; Aragon, C.; Bailey, S.; Baltay, C.; Bongard, S.; Buton, C.; Canto, A.; Childress, M.; Copin, Y.; Fakhouri, H. K.; Hsiao, E. Y.; Kerschhaggl, M.; Kowalski, M.; Loken, S.; Nugent, P.; Paech, K.; Pain, R.; Pecontal, E.; Pereira, R.; Perlmutter, S.; Rabinowitz, D.; Runge, K.; Scalzo, R.; Smadja, G.; Tao, C.;

Thomas, R. C.; Weaver, B. A.; Wu, C.; Nearby Supernova Factory, **"The reddening law of type Ia supernovae: separating intrinsic variability from dust using equivalent widths,"** Astronomy & Astrophysics, May 2011, 529:L4.

Sullivan, M.; Kasliwal, M. M.; Nugent, P. E.; Howell, D. A.; Thomas, R. C.; Ofek, E. O.; Arcavi, I.; Blake, S.; Cooke, J.; Gal-Yam, A.; Hook, I. M.; Mazzali, P.; Podsiadlowski, P.; Quimby, R.; Bildsten, L.; Bloom, J. S.; Cenko, S. B.; Kulkarni, S. R.; Law, N.; Poznanski, D., **"The Subluminous and Peculiar Type Ia Supernova PTF 09dav,"** Astrophysical Journal, May 2011, 732:118.

Barth, Aaron J.; Nguyen, My L.; Malkan, Matthew A.; Filippenko, Alexei V.; Li, Weidong; Gorjian, Varoujan; Joner, Michael D.; Bennert, Vardha Nicola; Botyanszki, Janos; Cenko, S. Bradley; Childress, Michael; Choi, Jieun; Comerford, Julia M.; Cucciara, Antonino; da Silva, Robert; Duchêne, Gaspard; Fumagalli, Michele; Ganeshalingam, Mohan; Gates, Elinor L.; Gerke, Brian F.; Griffith, Christopher V.; Harris, Chelsea; Hintz, Eric G.; Hsiao, Eric; Kandrashoff, Michael T.; Keel, William C.; Kirkman, David; Kleiser, Io K. W.; Laney, C. David; Lee, Jeffrey; Lopez, Liliana; Lowe, Thomas B.; Moody, J. Ward; Morton, Alekzandir; Nierenberg, A. M.; Nugent, Peter; Pancoast, Anna; Rex, Jacob; Rich, R. Michael; Silverman, Jeffrey M.; Smith, Graeme H.; Sonnenfeld, Alessandro; Suzuki, Nao; Tytler, David; Walsh, Jonelle L.; Woo, Jong-Hak; Yang, Yizhe; Zeisse, Carl, **"Broad-line Reverberation in the Kepler-field Seyfert Galaxy Zw 229-015,"** Astrophysical Journal, May 2011, 732:121.

Cano, Z.; Bersier, D.; Guidorzi, C.; Margutti, R.; Svensson, K. M.; Kobayashi, S.; Melandri, A.; Wiersema, K.; Pozanenko, A.; van der Horst, A. J.; Pooley, G. G.; Fernandez-Soto, A.; Castro-Tirado, A. J.; Postigo, A. De Ugarte; Im, M.; Kamble, A. P.; Sahu, D.; Alonso-Lorite, J.; Anupama, G.; Bibby, J. L.; Burgdorf, M. J.; Clay, N.; Curran, P. A.; Fatkhullin, T. A.; Fruchter, A. S.; Garnavich, P.; Gomboc, A.; Gorosabel, J.; Graham, J. F.; Gurugubelli, U.; Haislip, J.; Huang, K.; Huxor, A.; Ibrahimov, M.; Jeon, Y.; Jeon, Y.-B.; Ivarsen, K.; Kasen, D.; Klunko, E.; Kouveliotou, C.; Lacluyze, A.; Levan, A. J.; Loznikov, V.; Mazzali, P. A.; Moskvitin, A. S.; Mottram, C.; Mundell, C. G.; Nugent, P. E.; Nysewander, M.; O'Brien, P. T.; Park, W.-K.; Peris, V.; Pian, E.; Reichart, D.; Rhoads, J. E.; Rol, E.; Rumyantsev, V.; Scowcroft, V.; Shakhovskoy, D.; Small, E.; Smith, R. J.; Sokolov, V. V.; Starling, R. L. C.; Steele, I.; Strom, R. G.; Tanvir, N. R.; Tsapras, Y.; Urata, Y.; Vaduvescu, O.; Volnova, A.; Volvach, A.; Wijers, R. A. M. J.; Woosley, S. E.; Young, D. R., **"A tale of two GRB-SNe at a common redshift of z0.54,"** Monthly Notices of the Royal Astronomical Society, May 2011, 413:669-685.

Childress, M.; Aldering, G.; Aragon, C.; Antilogus, P.; Bailey, S.; Baltay, C.; Bongard, S.; Buton, C.; Canto, A.; Chotard, N.; Copin, Y.; Fakhouri, H. K.; Gangler, E.; Kerschhaggl, M.; Kowalski, M.; Hsiao, E. Y.; Loken, S.; Nugent, P.; Paech, K.; Pain, R.; Pecontal, E.; Pereira, R.; Perlmutter, S.; Rabinowitz, D.; Runge, K.; Scalzo, R.; Thomas, R. C.; Smadja, G.; Tao, C.; Weaver, B. A.; Wu, C., **"Keck Observations of the Young Metal-poor Host Galaxy of the Super-Chandrasekhar-mass Type Ia Supernova SN 2007if,"** Astrophysical Journal, May 2011, 733:3.

D. Hazen, J. Hick, W. Hurlbert, M. Welcome, **"Media Information Record (MIR) Analysis,"** LTUG 2011, April 19, 2011.

Zhengji Zhao, **"Introduction to Materials Science and Chemistry Applications at NERSC,"** NERSC User Training, April 5, 2011, Oakland, CA, April 5, 2011.

Zhengji Zhao, **"Tips to Compile Materials Science and Chemistry Codes at NERSC,"** NERSC User Training, April 5, 2011, Oakland, CA., April 5, 2011.

J. Hick, **"I/O Requirements for Exascale,"** Open Fabrics Alliance 2011, April 4, 2011.

Miller, Adam A.; Hillenbrand, Lynne A.; Covey, Kevin R.; Poznanski, Dovi; Silverman, Jeffrey M.; Kleiser, Io K. W.; Rojas-Ayala, Bárbara; Muirhead, Philip S.; Cenko, S. Bradley; Bloom, Joshua S.; Kasliwal, Mansi M.; Filippenko, Alexei V.; Law, Nicholas M.; Ofek, Eran O.; Dekany, Richard G.; Rahmer, Gustavo; Hale, David; Smith, Roger; Quimby, Robert M.; Nugent, Peter; Jacobsen, Janet; Zolkower, Jeff; Velur, Viswa; Walters, Richard; Henning, John; Bui, Khanh; McKenna, Dan; Kulkarni, Shrinivas R.; Klein, Christopher R.; Kandrashoff, Michael; Morton, Alekzandir, **"Evidence for an FU Orionis-like Outburst from a Classical T Tauri Star,"** Astrophysical Journal, April 1, 2011, 730:80.

K. Antypas, **"The Hopper XE6 System: Delivering High End Computing to the Nation's Science and Research Community,"** Cray Quarterly Review, April 1, 2011.

N. Balthaser, **"Introduction to Archival Storage at NERSC,"** March 8, 2011.

Richard Gerber, **"Computers—BSA Merit Badge,"** March 2011.

Richard Gerber, **"Debugging and Optimization Tools,"** Presented to CS267 class at UC-Berkeley, February 11, 2011.

Rob Cunningham, Zhengji Zhao, **"Compiling Code on the XE6,"** A talk at NERSC Cray XE6 user training, Feb 7-8, 2011, Oakland, CA., February 7, 2011.

"Introduction to OpenMP," Using the Cray XE6 Workshop, NERSC., February 7, 2011.

Scott Campbell, Steve Chan and Jason Lee, **"Detection of Fast Flux Service Networks,"** Australasian Information Security Conference 2011, January 17, 2011.

John Shalf, Donofrio, Rowen, Oliker, Michael F. Wehner, **"Green Flash: Climate Machine (LBNL),"** Encyclopedia of Parallel Computing, Pages: 809-819 January 1, 2011.

Matthias Reinsch, Brian Austin, John Corlett, Lawrence Doolittle, Gregory Penn, Donald Prosnitz, Ji Qiang, Andrew Sessler, Marco Menturini, Jonathan Wurtele, **"System Trade Analysis for an FEL Facility,"** Free Electron Laser Conference FEL 2011, Shanghai, China, January 1, 2011.

Maguire, K.; Sullivan, M.; Thomas, R. C.; Nugent, P.; Howell, D. A.; Gal-Yam, A.; Arcavi, I.; Ben-Ami, S.; Blake, S.; Botyanszki, J.; Buton, C.; Cooke, J.; Ellis, R. S.; Hook, I. M.; Kasliwal, M. M.; Pan, Y.-C.; Pereira, R.; Podsiadlowski, P.; Sternberg, A.; Suzuki, N.; Xu, D.; Yaron, O.; Bloom, J. S.; Cenko, S. B.; Kulkarni, S. R.; Law, N.; Ofek, E. O.; Poznanski, D.; Quimby, R. M., **"PTF10ops - a subluminous, normal-width light curve Type Ia supernova in the middle of nowhere,"** Monthly Notices of the Royal Astronomical Society, January 1, 2011, 418:747-758.

Keren Bergman, Hendry, Hargrove, Shalf, Jacob, Scott Hemmert, Rodrigues, David Resnick, **"Let there be light!: the future of memory systems is photonics and 3D stacking,"** MSPC, January 1, 2011, 43-48.

R.L.C. Starling, K. Wiersema, A.J. Levan, T. Sakamoto, D. Bersier, P. Goldoni, S.R. Oates, A. Rowlinson, S. Campana, J. Sollerman, N.R. Tanvir, D. Malesani, J.P.U. Fynbo, S. Covino, P. D'Avanzo, P.T. O'Brien, K.L. Page, J.P. Osborne, S.D. Vergani, S. Barthelmy, D.N. Burrows, Z. Cano, P.A. Curran, M. De Pasquale, V. D'Elia, P.A. Evans, H. Flores, A.S. Fruchter, P. Garnavich, N. Gehrels, J. Gorosabel, J. Hjorth, S.T. Holland, A.J. van der Horst, C.P. Hurkett, P. Jakobsson, A.P. Kamble, C. Kouveliotou, N.P.M. Kuin, L. Kaper, P.A. Mazzali, P.E. Nugent, E. Pian, M. Stamatikos, C.C. Thoene, S.E. Woosley, **"Discovery of the nearby long, soft GRB 100316D with an associated supernova,"** MNRAS, Pages: 2792-2803 January 1, 2011.

J. Chou, M. Howison, B. Austin, K. Wu, J. Qiang, E. Bethel, A. Shoshani, O. R\ ubel, R.D. Ryne, others, **"Parallel index and query for large scale data analysis,"** Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, January 1, 2011, 30.

J.N. Corlett, B. Austin, K.M. Baptiste, J.M. Byrd, P. Denes, R. Donahue, L. Doolittle, R.W. Falcone, D. Filippetto, S. Fournier, D. Li, H.A. Padmore, C. Papadopoulos, C. Pappas, G. Penn, M. Placidi, S. Prestemon, D. Prosnitz, J. Qiang, A. Ratti, M. Reinsch, F. Sannibale, R. Schlueter, R.W. Schoenlein, J.W. Staples, T. Vecchione, M. Venturini, R. Wells, R. Wilcox, J. Wurtele, A. Charman, E. Kur, A.A. Zholents, **"A Next Generation Light Source Facility at LBNL,"** PAC 11 Conference Proceedings, January 1, 2011.

B.M. Austin, D.Y. Zubarev, W.A. Lester, **"Quantum Monte Carlo and Related Approaches,"** Chemical Reviews, January 1, 2011.

M. Kerschhaggl, G. Aldering, P. Antilogus, C. Aragon, S. Bailey, C. Baltay, S. Bongard, C. Buton, A. Canto, M. Childress, N. Chotard, Y. Copin, H.K. Fakhouri, E. Gangler, E.Y. Hsiao, M. Kowalski, S. Loken, P. Nugent, K. Paech, R. Pain, E. Pecontal, R. Pereira, S. Perlmutter, D. Rabinowitz, K. Runge, R. Scalzo, G. Smadja, C. Tao, R.C. Thomas, C. Wu, **"Cosmology with the Nearby Supernova Factory, Progress in Particle and Nuclear Physics,"** Pages: 335-339 January 1, 2011.

J. Dongarra, P. Beckman, T. Moore, P. Aerts, G. Aloisio, J.C. Andre, D. Barkai, J.Y. Berthou, T. Boku, B. Braunschweig, others, **"The international exascale software project roadmap,"** International Journal of High Performance Computing Applications, January 2011, 25:3--60.

A. M. Smith, S. Lynn, M. Sullivan, C. J. Lintott, P. E. Nugent, J. Botyanszki, M. Kasliwal, R. Quimby, S. P. Bamford, L. F. Fortson, K. Schawinski, I. Hook, S. Blake, P. Podsiadlowski, J. Joensson, A. Gal-Yam, I. Arcavi, D. A. Howell, J. S. Bloom, J. Jacobsen, S. R. Kulkarni, N. M. Law, E. O. Ofek, R. Walters, **"Galaxy Zoo Supernovae,"** MNRAS, Pages: 1309-1319 January 1, 2011.

Brian Austin, Ji Qiang, Jonathan Wurtele, Alice Koniges, **"Influences of architecture and threading on the MPI communication strategies in an accelerator simulation code,"** SciDAC 2011, Denver, CO, January 1, 2011.

Kasliwal, Mansi M.; Kulkarni, Shri R.; Arcavi, Iair; Quimby, Robert M.; Ofek, Eran O.; Nugent, Peter; Jacobsen, Janet; Gal-Yam, Avishay; Green, Yoav; Yaron, Ofer; Fox, Derek B.; Howell, Jacob L.; Cenko, S. Bradley; Kleiser, Io; Bloom, Joshua S.; Miller, Adam; Li, Weidong; Filippenko, Alexei V.; Starr, Dan; Poznanski, Dovi; Law, Nicholas M.; Helou, George; Frail, Dale A.; Neill, James D.; Forster, Karl; Martin, D. Christopher; Tendulkar, Shriharsh P.; Gehrels, Neil; Kennea, Jamie; Sullivan, Mark; Bildsten, Lars; Dekany, Richard; Rahmer, Gustavo; Hale, David; Smith, Roger; Zolkower, Jeff; Velur, Viswa; Walters, Richard; Henning, John; Bui, Kahnh; McKenna, Dan; Blake, Cullen, **"PTF 10fqs: A Luminous Red Nova in the Spiral Galaxy Messier 99,"** APJ 730, Pages: 134 January 1, 2011.

James D. Neill, Mark Sullivan, Avishay Gal-Yam, Robert Quimby, Eran Ofek, Ted K. Wyder, D. Andrew Howell, Peter Nugent, Mark Seibert, D. Christopher Martin, Roderik Overzier , Tom A. Barlow, Karl Foster, Peter G. Friedman, Patrick Morrissey, Susan G. Neff, David Schiminovich, Luciana Bianchi , José Donas, Timothy M. Heckman, Young-Wook Lee, Barry F. Madore, Bruno Milliard, R. Michael Rich, Alex S. Szalay, **"The Extreme Hosts of Extreme Supernovae,"** APJ, Pages: 15 January 1, 2011.

S. Campbell, **"Local System Security via SSHD Instrumentation,"** LISA, January 1, 2011.

Hemant Shukla, Hsi-Schive, Tak-Pong Woo, Tzihong Chiueh, **"Multi-science applications with single codebase — GAMER — for massively parallel architectures,"** SC, January 1, 2011, 37.

Conference on High Performance Computing Networking, **"Storage Analysis,"** SC 2011, Seattle, WA, USA, November 12-18, 2011," SC, ACM, January 1, 2011.

H. Agakishiev, others, **"High pT non-photonic electron production in p+p collisions at √s=200 GeV,"** Phys.Rev., January 1, 2011, D83:052006.

G. Agakishiev, others, **"Anomalous centrality evolution of two-particle angular from Au-Au collisions at $\sqrt{s_{\rm NN}}$ = 62 and 200 GeV,"** January 1, 2011.

E. Deines, G.H. Weber, C. Garth, B. Van Straalen, S. Borovikov, D.F. Martin, K. Joy, **"On the Computation of Integral Curves in Adaptive Mesh Refinement Vector Fields,"** Proceedings of Dagstuhl Seminar on Scientific Visualization 2009, January 1, 2011.

Gunther H. Weber, Peer-Timo Bremer, Valerio Pascucci, **"Topological Cacti: Visualizing Contour-based Statistics,"** Topological Methods in Data Analysis and Visualization II, January 1, 2011.

Gunther H. Weber, Peer-Timo Bremer, Attila Gyulassy and, **"Topology-based Feature Definition and Analysis,"** Numerical Modeling of Space Plasma Flows—ASTRONUM-2010, Astronomical Society of the Pacific, January 1, 2011, 444:292--297.

H. Agakishiev, others, **"Observation of the antimatter helium-4 nucleus,"** Nature, January 1, 2011, 473:353.

G. Agakishiev, others, **"Measurement of the $W \to e \nu$ and $Z/\gamma^* \to e^+e^-$ Production Cross Sections at Mid-rapidity in Proton-Proton Collisions at $\sqrt{s}$ = 500 GeV,"** January 1, 2011.

H. Agakishiev, others, **"Evolution of the differential transverse momentum function with centrality in Au+Au collisions,"** Phys.Lett., January 1, 2011, B704:467-473.

H. Agakishiev, others, **"Studies of di-jet survival and surface emission bias in collisions via angular correlations with respect to leading hadrons,"** Phys.Rev., January 1, 2011, C83:061901.

A. Dubey, K. B. Antypas, C. Daley, **"Parallel Algorithms for Moving Lagrangian Data on Block Structured Eulerian Meshes,"** Parallel Computing, 2011, 37 (2):101 - 113.

John Shalf, D. Quinlan, Curtis Janssen, **"Rethinking Hardware-Software Codesign for Exascale Systems,"** IEEE Computer, January 1, 2011, 44:22-30.

M.M. Aggarwal, others, **"Scaling properties at freeze-out in relativistic heavy collisions,"** Phys.Rev., January 1, 2011, C83:034910.

K. Furlinger, N.J. Wright, D. Skinner, **"Comprehensive Performance Monitoring for GPU Cluster Systems,"** Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on, 2011, 1377--1386.

Kenes Beketayev, Gunther H. Weber, Maciej Haranczyk, Bremer, Mario Hlawitschka, Bernd Hamann, **"Visualization of Topology of Transformation Pathways in Complex Systems,"** Computer Graphics Forum (Special Issue, Proceedings Symposium on Visualization), January 2011, 663--672.

Kevin R. Covey, Lynne A. Hillenbrand, Adam A. Miller, Dovi Poznanski, S. Bradley Cenko, Jeffrey M. Silverman, Joshua S. Bloom, Mansi M. Kasliwal, William Fischer, John Rayner, Luisa M. Rebull, Nathaniel R. Butler, Alexei V. Filippenko, Nicholas M. Law, Eran O. Ofek, Marcel Agueros, Richard G. Dekany, Gustavo Rahmer, David Hale, Roger Smith, Robert M. Quimby, Peter Nugent, Janet Jacobsen, Jeff Zolkower, Viswa Velur, Richard Walters, John Henning, Khanh Bui, Dan McKenna, Shrinivas R. Kulkarni, Christopher Klein, **"PTF10nvg: An Outbursting Class I Protostar in the Pelican/North American Nebula,"** AJ, January 1, 2011, 141:40.

Jeff Cooke, Richard S. Ellis, Mark Sullivan, Peter Nugent, D. Andrew Howell, Avishay Gal-Yam, Chris Lidman, Joshua S. Bloom, S. Bradley Cenko, Mansi M. Kasliwal, Shrinivas R. Kulkarni, Nicholas M. Law, Eran O. Ofek, Robert M. Quimby, **"Hubble Space Telescope Studies of Nearby Type Ia Supernovae: The Mean Maximum Light Ultraviolet Spectrum and its Dispersion,"** APJL, Pages: L35 January 1, 2011.

R.C,Thomas, P.E. Nugent, J.C. Meza, **"SYNAPPS: Data-Driven Analysis for Supernova Spectroscopy,"** PASP, Pages: 237-248 January 1, 2011.

R.L.C. Starling, K. Wiersema, A.J. Levan, T. Sakamoto, D. Bersier, P. Goldoni, S.R. Oates, A. Rowlinson, S. Campana, J. Sollerman, N.R. Tanvir, D. Malesani, J.P.U. Fynbo, S. Covino, P. D'Avanzo, P.T. O'Brien, K.L. Page, J.P. Osborne, S.D. Vergani, S. Barthelmy, D.N. Burrows, Z. Cano, P.A. Curran, M. De Pasquale, V. D'Elia, P.A. Evans, H. Flores, A.S. Fruchter, P. Garnavich, N. Gehrels, J. Gorosabel, J. Hjorth, S.T. Holland, A.J. van der Horst, C.P. Hurkett, P. Jakobsson, A.P. Kamble, C. Kouveliotou, N.P.M. Kuin, L. Kaper, P.A. Mazzali, P.E. Nugent, E. Pian, M. Stamatikos, C.C. Thoene, S.E. Woosley, **"Discovery of the nearby long soft GRB 100316D with an associated supernova,"** MNRAS, Pages: 2792-2803 January 1, 2011.

M.M. Aggarwal, others, **"Measurement of the parity-violating longitudinal single-spin asymmetry for $W^ {\pm}$ boson production in polarized proton-proton collisions at $\sqrt{s} = 500 $GeV,"** Phys.Rev.Lett., January 1, 2011, 106:062002.

Adamczyk, others, **"Directed Flow of Identified Particles in Au + Au Collisions at $\sqrtsNN = 200$ GeV at RHIC,"** January 1, 2011.

M.M. Aggarwal, others, **"Strange and Multi-strange Particle Production in Au+Au Collisions at $\sqrt{s_{NN}}$ = 62.4 GeV,"** Phys.Rev., January 1, 2011, C83:024901.

Appendix A
NERSC Client Statistics

In support of the DOE Office of Science's mission, NERSC served 4,934 scientists throughout the United States in 2011. These researchers work in DOE laboratories, universities, industry, and other Federal agencies. Figure 1 shows the proportion of NERSC usage by each type of institution, while Figures 2 and 3 show laboratory, university, and other organizations that used large allocations of computer time. Computational science conducted at NERSC covers the entire range of scientific disciplines, but is focused on research that supports the DOE's mission and scientific goals, as shown in Figure 4.

On their Allocation Year 2012 proposal forms, principal investigators reported 1,770 refereed publications (published or in press)—as well as 93 publications submitted to refereed journals—for the preceding 12 months, based, at least in part, on using NERSC resources. Lists of publications resulting from use of NERSC resources are available at http://www.nersc.gov/news-publications/publications-reports/nersc-user-publications/.

The MPP hours reported here are Cray XT4 equivalent hours.

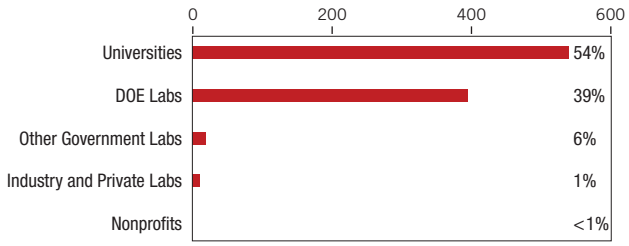


Figure 1. NERSC usage by institution type, 2011 (MPP hours in millions).

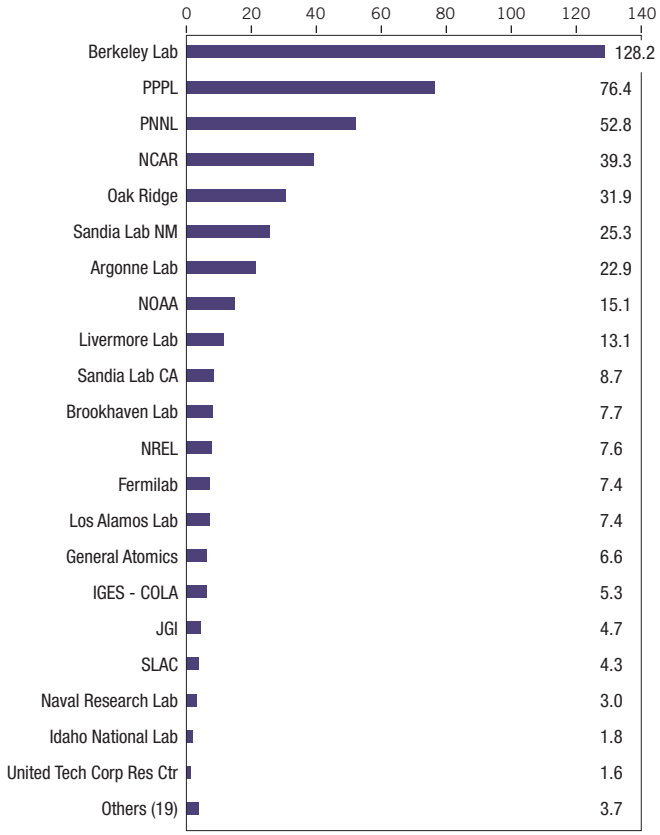


Figure 2. DOE, other laboratory, and industry usage at NERSC, 2011 (MPP hours in millions).

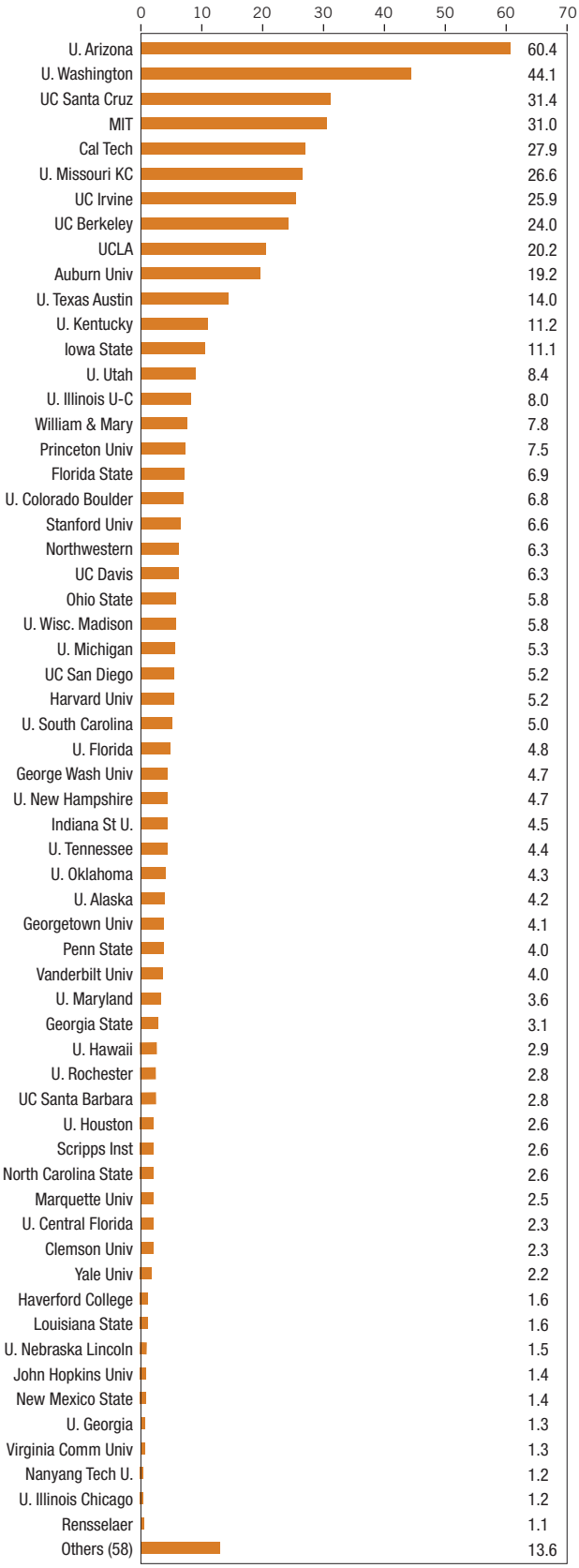


Figure 3. Academic usage at NERSC, 2011 (MPP hours in millions).

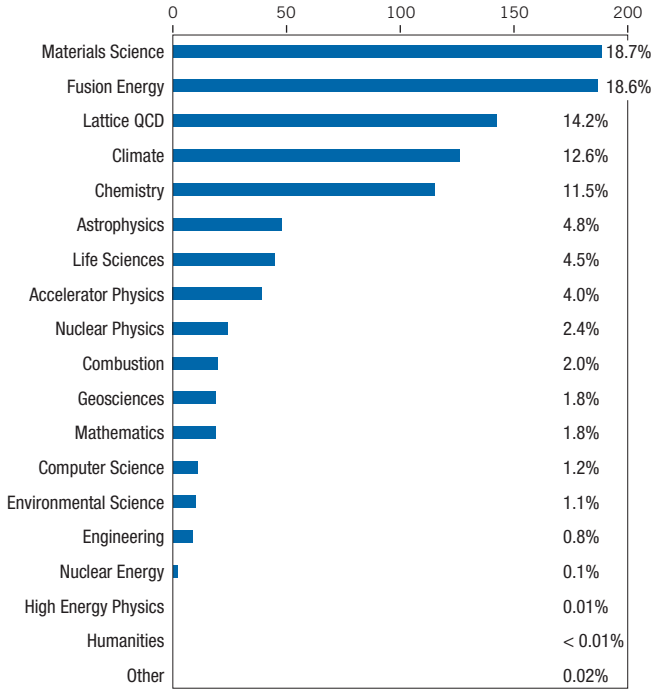


Figure 4. NERSC usage by scientific discipline, 2011 (MPP hours in millions).

Appendix B
NERSC Users Group Executive Committee

Office of Advanced Scientific Computing Research
Mark Adams, *Columbia University*
Sergei Chumakov, *Stanford University*
Mike Lijewski, *Lawrence Berkeley National Laboratory*

Office of Basic Energy Sciences
Eric Bylaska, *Pacific Northwest National Laboratory*
Paul Kent, *Oak Ridge National Laboratory*
Brian Moritz, *SLAC National Accelerator Laboratory*

Office of Biological and Environmental Research
David Beck, *University of Washington*
Barbara Collignon, *Oak Ridge National Laboratory*
Adrianne Middleton, *National Center for Atmospheric Research*

Office of Fusion Energy Sciences
Stephane Ethier (Chair), *Princeton Plasma Physics Laboratory*
Linda Sugiyama, *Massachusetts Institute of Technology*
Jean-Luc Vay, *Lawrence Berkeley National Laboratory*

Office of High Energy Physics
Julian Borrill, *Lawrence Berkeley National Laboratory*
Steven Gottlieb, *Indiana University*
Frank Tsung (Vice Chair), *University of California, Los Angeles*

Office of Nuclear Physics
Daniel Kasen, *Lawrence Berkeley National Laboratory*
Tomasz Plewa, *Florida State University*
Martin Savage, *University of Washington*

Members at Large
Gregory Newman, *Lawrence Berkeley National Laboratory*
Jason Nordhaus, *Princeton University*
Maxim Umansky, *Lawrence Livermore National Laboratory*

Appendix C

Office of Advanced Scientific Computing Research

The mission of the Advanced Scientific Computing Research (ASCR) program is to discover, develop, and deploy computational and networking capabilities to analyze, model, simulate, and predict complex phenomena important to the Department of Energy (DOE). A particular challenge of this program is fulfilling the science potential of emerging computing systems and other novel computing architectures, which will require numerous significant modifications to today's tools and techniques to deliver on the promise of exascale science.

To accomplish its mission and address those challenges, the ASCR program is organized into two subprograms: Mathematical, Computational, and Computer Sciences Research; and High Performance Computing and Network Facilities.

- The Mathematical, Computational, and Computer Sciences Research subprogram develops mathematical descriptions, models, methods, and algorithms to describe and understand complex systems, often involving processes that span a wide range of time and/or length scales. The subprogram also develops the software to make effective use of advanced networks and computers, many of which contain thousands of multi-core processors with complicated interconnections, and to transform enormous data sets from experiments and simulations into scientific insight.

- The High Performance Computing and Network Facilities subprogram delivers forefront computational and networking capabilities and contributes to the development of next-generation capabilities through support of prototypes and testbeds.

Berkeley Lab thanks the program managers with direct responsibility for the NERSC program and the research projects described in this report:

Daniel Hitchcock, *Associate Director, ASCR*
Robert Lindsay, *Special Projects*
Julie Scott, *Financial Management Specialist*
Lori Jernigan, *Program Support Specialist*
Melea Baker, *Administrative Specialist*

Facilities Division

Vincent Dattoria, *Computer Scientist*
Dave Goodwin, *Physical Scientist*
Sally McPherson, *Program Assistant*
Betsy Riley, *Computer Scientist*
Yukiko Sekine, *Computer Scientist*

Computational Science Research and Partnerships (SciDAC) Division

William Harrod, *Division Director*
Teresa Beachley, *Program Assistant*
Rich Carlson, *Computer Scientist*
Christine Chalk, *Physical Scientist*
Sandy Landsberg, *Mathematician*
Carolyn Lauzon, *AAAS S&T Fellow*
Randall Laviolette, *Physical Scientist*
Steven Lee, *Physical Scientist*
Thomas Ndousse-Fetter, *Computer Scientist*
Lucy Nowell, *Computer Scientist*
Karen Pao, *Mathematician*
Sonia Sachs, *Computer Scientist*
Ceren Susut, *Physical Scientist*
Angie Thevenot, *Program Assistant*

Appendix D
Advanced Scientific Computing Advisory Committee

The Advanced Scientific Computing Advisory Committee (ASCAC) provides valuable, independent advice to the Department of Energy on a variety of complex scientific and technical issues related to its Advanced Scientific Computing Research program. ASCAC's recommendations include advice on long-range plans, priorities, and strategies to address more effectively the scientific aspects of advanced scientific computing including the relationship of advanced scientific computing to other scientific disciplines, and maintaining appropriate balance among elements of the program. The Committee formally reports to the Director, Office of Science. The Committee primarily includes representatives of universities, national laboratories, and industries involved in advanced computing research. Particular attention is paid to obtaining a diverse membership with a balance among scientific disciplines, institutions, and geographic regions.

Roscoe C. Giles, *Chair, Boston University*

Marsha Berger, *Courant Institute of Mathematical Sciences*

Marjory S. Blumenthal, *Georgetown University*

Vinton G. Cerf, *Google, Inc.*

Vincent Chan, *General Atomics*

Barbara M.P. Chapman, *University of Houston*

Jackie Chen, *Sandia National Laboratories*

Jack J. Dongarra, *University of Tennessee*

Sharon C. Glotzer, *University of Michigan*

Susan L. Graham, *University of California, Berkeley*

Anthony Hey, *Microsoft Research*

Gwendolyn L. Huntoon, *Pittsburgh Supercomputing Center*

Juan Meza, *University of California, Merced*

John Negele, *Massachusetts Institute of Technology*

Linda R. Petzold, *University of California, Santa Barbara*

Vivek Sarkar, *Rice University*

Victoria White, *Fermi National Accelerator Laboratory*

Dean N. Williams, *Lawrence Livermore National Laboratory*

Appendix E
Acronyms and Abbreviations

ACS
American Chemical Society

Adv. Mater.
Advanced Materials

ALCC
ASCR Leadership Computing Challenge

AMD
Advanced Micro Devices

Appl. Phys. Lett.
Applied Physics Letters

APS
American Physical Society

ASCEM
Advanced Simulation Capability for Environmental Management

ASCR
Office of Advanced Scientific Computing Research (DOE)

BAO
Baryon acoustic oscillation

BELLA
Berkeley Lab Laser Accelerator

BER
Office of Biological and Environmental Research (DOE)

BES
Office of Basic Energy Sciences (DOE)

BLAST
Basic Local Alignment Search Tool

BOSS
Baryon Oscillation Spectroscopic Survey

$CaCl_2$
Calcium chloride

CAF
Co-Array Fortran

CAM5
Community Atmospheric Model version 5

CCM
Cluster Compute Mode

CFL
Compact fluorescent light

CO_2
Carbon dioxide

COLA
Center for Ocean-Land-Atmosphere Studies

CSTB
Computer Science and Telecommunications Board

DNA
Deoxyribonucleic acid

DNTT
Dinaphthothienothiophene

DOE
U.S. Department of Energy

DOI
Digital object identifier

DTN
Data transfer node

EBI
Energy Biosciences Institute

EDL
Electrical double layer

EFRC
Energy Frontiers Research Center

EMSL
Environmental Molecular Science Laboratory at Pacific Northwest National Laboratory

ERCAP
Energy Research Computing Allocations Process

ESnet
Energy Sciences Network

FES
Office of Fusion Energy Sciences (DOE)

GB
Gigabyte

Geophys. Res. Lett.
Geophysical Research Letters

GPU
Graphics processing unit

HEP
Office of High Energy Physics (DOE)

HPC
High performance computing

HPSS
High Performance Storage System

IEEE
Institute of Electrical and Electronics Engineers

IGES
Institute of Global Environment and Society

I/O
Input/output

IT
Information technology

ITER
Latin for "the way"; an international fusion energy experiment in southern France

J. Geophys. Res.
Journal of Geophysical Research

JGI
Joint Genome Institute (DOE)

LAN
Local area network

LBNL
Lawrence Berkeley National Laboratory

LED
Light-emitting diode

LES
Large eddy simulation

LIGO
Laser Interferometer Gravitational-Wave Observatory

LSM
Library storage module

MHD
Magnetohydrodynamic

MPI
Message Passing Interface

MPP
Massively parallel processing

NCAR
National Center for Atmospheric Research

NCGC
Center for Nanoscale Control of Geologic CO_2

NaCl
Sodium chloride

NASA
National Aeronautics and Space Administration

NERSC
National Energy Research Scientific Computing Center

NGF
NERSC Global Filesystem

NICS
National Institute for Computational Sciences at the University of Tennessee/Oak Ridge National Laboratory

NISE
NERSC Initiative for Scientific Exploration

NOAA
National Oceanic and Atmospheric Administration

NP
Office of Nuclear Physics (DOE)

NSD
Nuclear Science Division at Lawrence Berkeley
National Laboratory

NSF
National Science Foundation

NSTX
National Spherical Torus Experiment

NTM
Neoclassical tearing mode

Nucl. Fusion
Nuclear Fusion

NUG
NERSC Users Group

OLCF
Oak Ridge Leadership Computing Facility

OLED
Organic light-emitting diode

PDSF
Parallel Distributed Systems Facility (NERSC)

PE
Processing element

Phys. Plasmas
Physics of Plasmas

PIBS
Parallel Incremental Backup System

PIC
Particle-in-cell

PMAMR
Porous Media Adaptive Mesh Refinement

PNAS
Proceedings of the National Academy of Sciences

PNNL
Pacific Northwest National Laboratory

PPPL
Princeton Plasma Physics Laboratory

RHIC
Relativistic Heavy Ion Collider

RMA
Remote Memory Access

SciDAC
Scientific Discovery through Advanced Computing (DOE)

SDSS
Sloan Digital Sky Survey

SGE
Sun Grid Engine

SIMD
Single Instruction, Multiple Data

SLAC
SLAC National Accelerator Laboratory

SMP
Symmetric multiprocessor

SN Ia
Type Ia supernova

STAR
Solenoidal Tracker at RHIC

TACC
Texas Advanced Computing Center

TB
Terabyte

TPC
Time Projection Chamber

UCSB
University of California, Santa Barbara

UIUC
University of Illinois at Urbana-Champaign

UNEDF
Universal Nuclear Energy Density Functional Project

UPC
Unified Parallel C (programming language)

VPC
Virtual private cluster

WAN
Wide area network

For more information about NERSC, contact:

Jon Bashor
NERSC Communications
Berkeley Lab, MS 50B2215
1 Cyclotron Road
Berkeley, CA 94720-8148

email jbashor@lbl.gov
phone (510) 486-5849
fax (510) 486-4300

NERSC's web site
http://www.nersc.gov/

NERSC Annual Report Editor
John Hules

Contributing writers
Jon Bashor, John Hules, Linda Vu, and Margie Wylie
Berkeley Lab Computing Sciences Communications Group

Katie Antypas, Shane Canon, Richard Gerber, Jason Hick, Peter Nugent, Hemant Shukla, David Skinner, Francesca Verdier, and Harvey Wasserman
NERSC

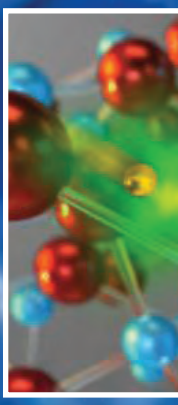
Dan Krotz and Paul Preuss
Berkeley Lab Communications Department

Janet Lathrop
University of Massachusetts Amherst

Jennifer Ovink and Greg Exarhos
Pacific Northwest National Laboratory

Design, layout, illustrations, photography, and printing coordination provided by the Creative Services Office of Berkeley Lab's Public Affairs Department.

CSO 25059



NERSC

BERKELEY LAB
Lawrence Berkeley
National Laboratory

U.S. DEPARTMENT OF
ENERGY
Office of Science