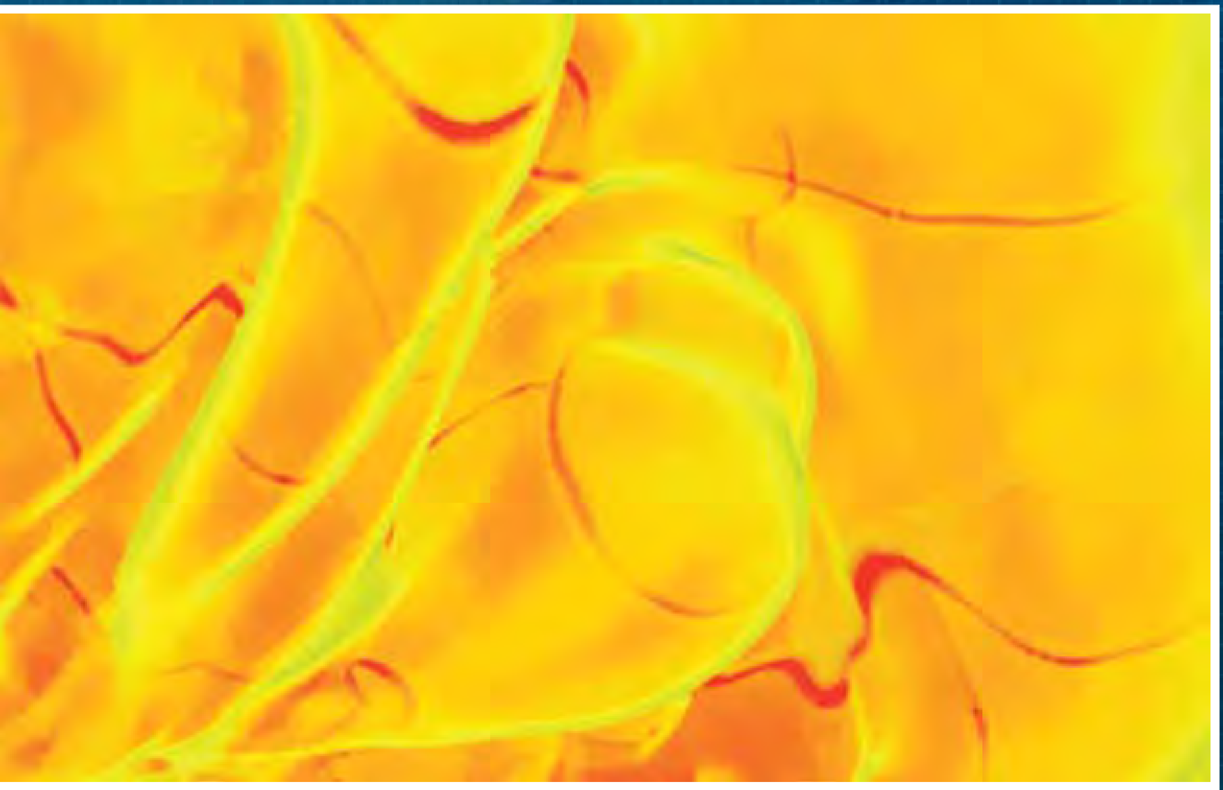
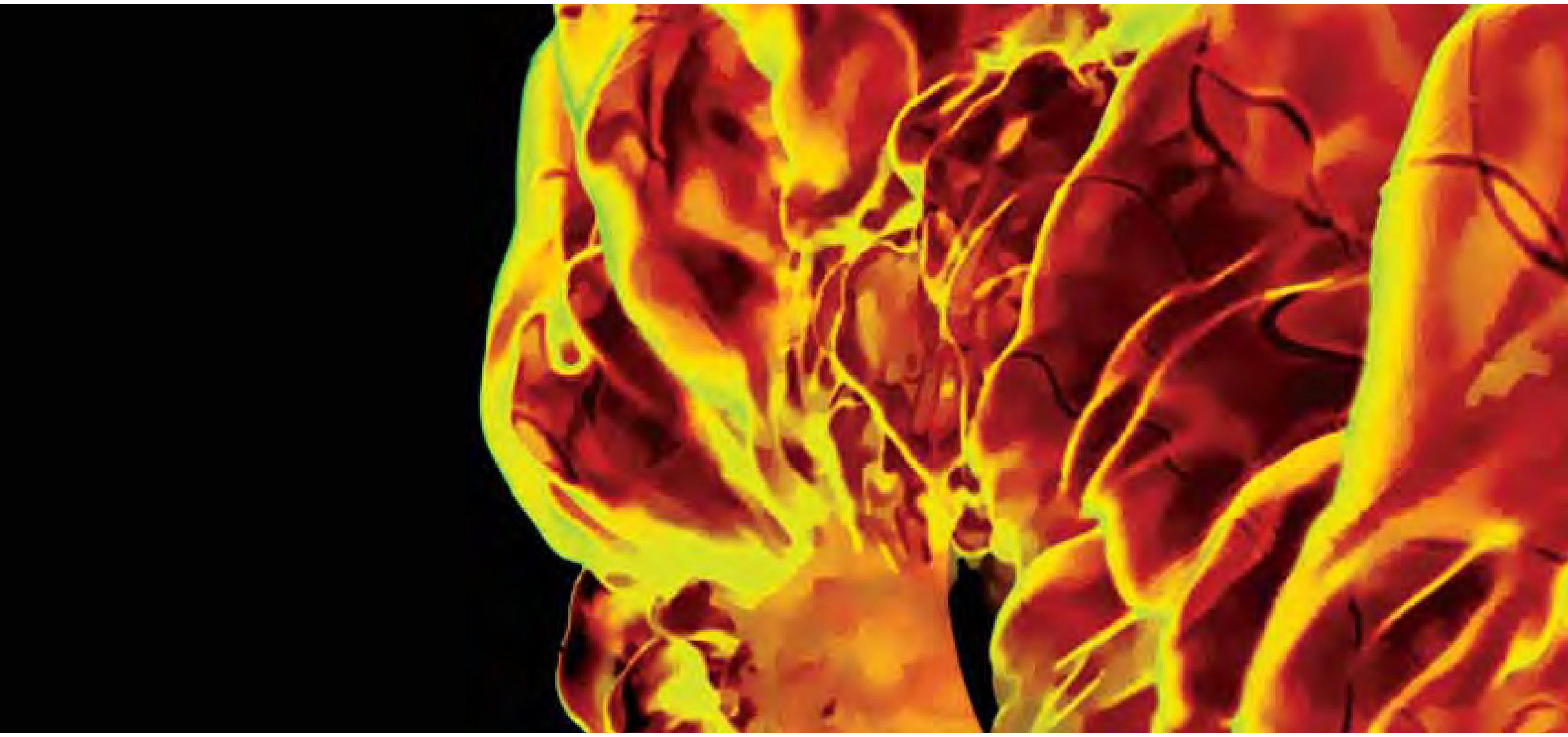


NERSC

National Energy Research Scientific Computing Center

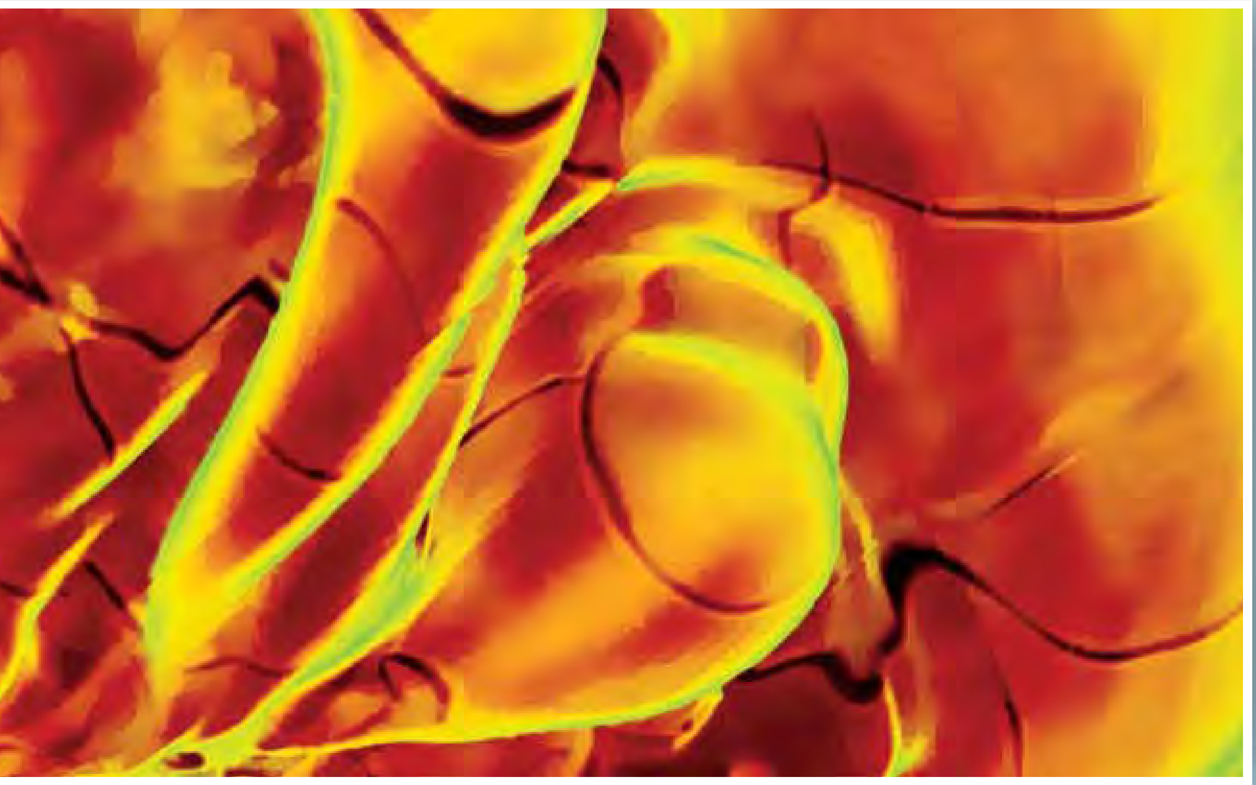
2007 Annual Report



NATIONAL ENERGY RESEARCH
SCIENTIFIC COMPUTING CENTER

National Energy Research Scientific Computing Center

2007 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.

LBNL-1143E, October 2008

Table of Contents

The Year in Perspective

As a computer scientist, I have always been interested in making computer systems more efficient and easier to use through better architectures, programming languages, algorithms, and tools that connect the hardware to the applications. So when Berkeley Lab offered me the position of NERSC Division Director beginning in January 2008, I saw it as an opportunity to help scientists make new discoveries in domains ranging from basic scientific understanding, such as the origins of the universe, to some of the most critical issues facing the world today, including climate modeling and the development of new sources of energy.

The computing industry is facing its greatest challenge ever, as it shifts from single-core to multicore processing chips, driven by power density limits and the recognition that hidden forms of instruction-level parallelism have been tapped out. This change was apparent at NERSC in 2007 as a major computational system, a Cray XT4 built with dual-core AMD chips, was introduced. The deployment of this system, known as Franklin, was a major milestone in NERSC's history — as the first major system based on multicore technology, it sustained the performance increases that the community has come to expect, increasing the computational power available to NERSC users by a factor of six.

With close to 20,000 cores, Franklin has a theoretical peak speed of over 100 teraflops (100 trillion floating point operations per second). In the installation, testing, and acceptance of Franklin, NERSC staff demonstrated their expertise in standing up large systems that provide excellent performance across a diverse set of applications. With nearly 3000 users and 500 application codes from across the DOE science disciplines, the NERSC workload is one of the most challenging workloads supported by any center. To support this user base, NERSC and Cray developed a plan to test Cray's CLE operating system, an ultra-lightweight version of Linux, and Franklin became the first Cray XT4 system to run CLE in production.

The availability of Franklin released a pent-up demand for large-scale computing as users quickly adapted their codes to run on the multicore-based machine with CLE. Within a week of its formal acceptance in October 2007, Franklin was 80–95% utilized, and the users, on average, consumed five times more compute time on Franklin than they had initially been allocated for 2007 — fourteen times more for the largest users. They used this opportunity to scale their codes to new levels, to experiment with new algorithms, and to produce new scientific results.

The real impact of the Franklin system is measured by the science produced by NERSC users. One dramatic example was the project "Structure and Reactions of Hadrons and Nuclei," led by James Vary of Iowa State University, which investigates longstanding problems in nuclear physics such as the nature of the strong interactions and the origins of the spin-orbit force. These researchers originally had an allocation of only 200,000 hours, but were able to use 4 million hours on Franklin for their realistic *ab initio* calculations of nucleon–nucleon interactions of oxygen-16. By increasing the scaling of their calculations from 2000 to 12,000 compute cores and diagonalizing a matrix of dimension 1 billion, they achieved the most accurate calculations to date on this size nucleus. These results can be used to parameterize new density functionals for nuclear structure simulations.

Another outstanding achievement this year was the “20th Century Reanalysis” INCITE project, led by Gil Compo of the University of Colorado and the NOAA Earth System Research Lab, which is using an algorithm called an Ensemble Kalman Filter to reconstruct missing climate data from 1892 to the present. Compo’s team has successfully reproduced historical weather phenomena like the 1922 Knickerbocker Storm, and the comprehensive three-dimensional database they are producing will be used to validate climate and weather models. With a 3.1-million-hour allocation and what they described as “fabulous support” from NERSC consultants, the researchers ran their code on all four of NERSC’s large-scale computing systems, switched to a higher-resolution algorithm when they moved to the Cray XT4, and parallelized virtually their entire workflow.

The year 2007 represents both a beginning and an end in the history of NERSC’s major computational systems, as the IBM SP RS/6000 system, Seaborg, was in its seventh and final year of production use. Over the course of its lifetime, Seaborg provided over 250 million CPU hours to the users and resulted in an estimated 7000 published scientific results in astrophysics, climate research, fusion energy, chemistry, and other disciplines. One of the last breakthroughs enabled by Seaborg was the first spontaneous detonation of a white dwarf star into a supernova in a three-dimensional simulation, achieved by Don Lamb and a team at the University of Chicago Flash Center using resources at NERSC and Lawrence Livermore National Laboratory. NERSC HPC consultants helped get the Flash Center team’s 512-processor job up and running on short notice to help them meet a hard deadline and a longstanding scientific goal.

Despite the availability of the new Franklin system, there is still a huge unmet demand from users for more compute time. We will upgrade Franklin from dual cores to quad cores during the second half of 2008 and are beginning a project for the procurement of the next major computational system at NERSC.

As we look to the future, two critical issues arise for NERSC and the users it serves: the growing energy requirement of large-scale systems, which could dwarf the cost of hardware purchases if it goes unchecked, and the virtual tsunami of scientific data arising from simulations as well as experimental and measurement devices. In the discussion of NERSC’s transition to the exascale in the last section of this annual report, we describe these challenges in more detail along with two longer-term goals we will be pursuing to address them.

The first goal comes from a holistic look at system design, including hardware, algorithms, software, and applications; it involves leveraging low-power embedded processing technology, massively multicore compute nodes to reduce power, and scalable algorithms. As a demonstration, we are focusing on a system design driven by global climate modeling with the goal of demonstrating an affordable system for kilometer-scale modeling.

The second goal will look at the growing data requirements of DOE science areas, and the spectrum of storage, communication, and computing technology needed to preserve, manage, and analyze the data. Just as web search engines have revolutionized nearly every aspect of our lives, we believe that better access to and ability to search and manipulate scientific data sets will revolutionize science in the next decade.

I am looking forward to working with NERSC’s dedicated staff, Associate Lab Director (and former NERSC Director) Horst Simon, and the diverse community of researchers we support to continue finding innovative solutions to these and other challenging scientific problems.

Katherine Yelick
NERSC Division Director

Quantum Secrets of Photosynthesis Revealed

Computational models guide the design of the experiment and help interpret the results

Through photosynthesis, green plants and cyanobacteria are able to transfer sunlight energy to molecular reaction centers for conversion into chemical energy with nearly 100-percent efficiency. Speed is the key — the transfer of the solar energy takes place almost instantaneously so little energy is wasted as heat. How photosynthesis achieves this near instantaneous energy transfer is a longstanding mystery that may have finally been solved.

A study led by researchers with Berkeley Lab and the University of California (UC) at Berkeley reports that the answer lies in quantum mechanical effects. Results of the study were presented in the April 12, 2007 issue of the journal Nature.[1]

"We have obtained the first direct evidence that remarkably long-lived wavelike electronic quantum coherence plays an important part in energy transfer processes during photosynthesis," said Graham Fleming, the principal investigator for the study. "This wavelike characteristic can explain the extreme efficiency of the energy transfer because it enables the system to simultaneously sample all the potential energy pathways and choose the most efficient one."

Fleming, a former Deputy Director of Berkeley Lab, is a researcher in the Lab's Physical Biosciences Division, a professor of chemistry at UC Berkeley, and an internationally acclaimed leader in spectroscopic studies of the photosynthetic process. Co-authoring the Nature paper were Gregory Engel, who was first author, Tessa Calhoun, Elizabeth Read, Tae-Kyu Ahn, Tomáš Mančal, and Yuan-Chung Cheng, all of whom held joint appointments with Berkeley Lab's Physical Biosciences Division and the UC Berkeley Chemistry Department at the time of the study, plus Robert Blankenship, from Washington University in St. Louis.

In the paper, Fleming and his collaborators report the detection of "quantum beating" signals, coherent electronic oscillations in both donor and acceptor molecules, generated by light-induced energy excitations, like the ripples formed when stones are tossed into a pond (Figure 1). Electronic spectroscopy measurements made on a femtosecond (millionths of a billionth of a second) time-scale showed the oscillations meeting and interfering constructively, forming wavelike motions of energy (superposition states) that can explore all potential energy pathways simultaneously and reversibly, meaning

Project: Simulations of Nonlinear Optical Spectra and Energy Transfer Dynamics of Photosynthetic Light-Harvesting Complexes

PI: Graham Fleming, University of California, Berkeley, and Lawrence Berkeley National Laboratory

Senior investigators: Elizabeth Read and Yuan-Chung Cheng, University of California, Berkeley, and Lawrence Berkeley National Laboratory

Funding: BES, MIBRS, KRFG

[1] G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mančal, Y.-C. Cheng, R. E. Blankenship, and G. R. Fleming, "Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems," Nature **446,** 782 (2007).

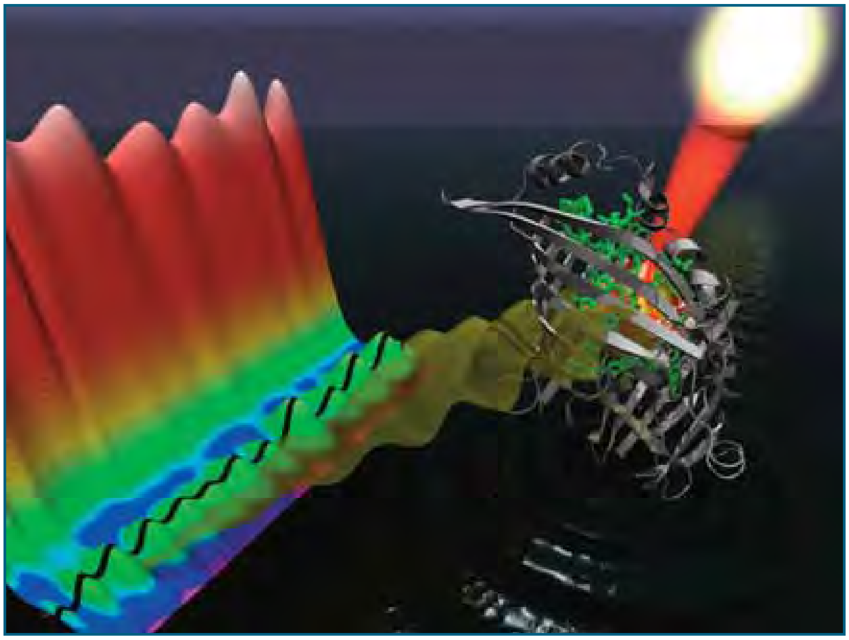


Figure 1. Sunlight absorbed by bacteriochlorophyll (green) within the FMO protein (gray) generates a wavelike motion of excitation energy whose quantum mechanical properties can be mapped through the use of two-dimensional electronic spectroscopy. (Image courtesy of Greg Engel)

they can retreat from wrong pathways with no penalty.

This finding contradicts the classical description of the photosynthetic energy transfer process as one in which excitation energy hops from light-capturing pigment molecules to reaction center molecules step-by-step down the molecular energy ladder.

"The classical hopping description of the energy transfer process is both inadequate and inaccurate," said Fleming. "It gives the wrong picture of how the process actually works, and misses a crucial aspect of the reason for the wonderful efficiency."

Following the flow of energy

The photosynthetic technique for transferring energy from one molecular system to another should make any short-list of Mother Nature's spectacular accomplishments. If we can learn enough to emulate this process, we might be able to create artificial versions of photosynthesis that would help us effectively tap into the sun as a clean, efficient, sustainable and carbon-neutral source of energy.

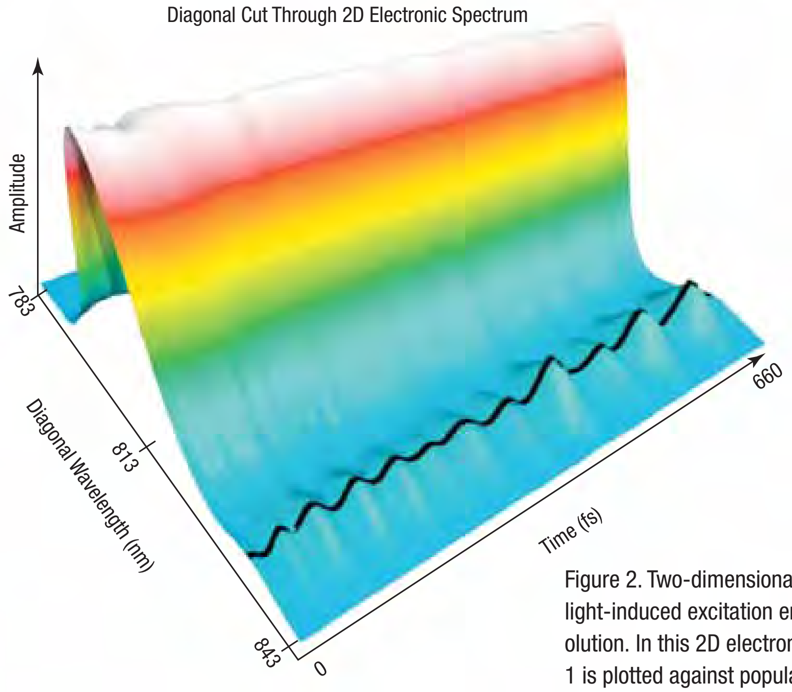
Towards this end, Fleming and his research group have developed a technique called two-dimensional electronic spectroscopy that enables them to follow the flow of light-induced excitation energy through molecular complexes with femtosecond temporal resolution. The technique involves sequentially flashing a sample with femtosecond pulses of light from three laser beams. A fourth beam is used as a local oscillator to amplify and detect the resulting spectroscopic signals as the excitation energy from the laser lights is transferred from one molecule to the next. (The excitation energy changes the way each molecule absorbs and emits light.)

Fleming has compared 2D electronic spectroscopy to the technique used in the early super-heterodyne radios, where an incoming high-frequency radio signal was converted by an oscillator to a lower frequency for more controllable amplification and better reception. In the case of 2D electronic spectroscopy, scientists can track the transfer of energy between molecules that are coupled (connected) through their electronic and vibrational states in any photoactive system, macromolecular assembly, or nanostructure.

Fleming and his group first described 2D electronic spectroscopy in a 2005 *Nature* paper, when they used the technique to observe electronic couplings in the Fenna-Matthews-Olson (FMO) photosynthetic light-harvesting protein, a molecular complex in green sulfur bacteria.[2]

Said Engel, "The 2005 paper was the first biological application of this technique; now we have used 2D electronic spectroscopy to discover a new phenomenon in photosynthetic systems. While the

[2] T. Brixner, J. Stenger, H. M. Vaswani, M. Cho, R. E. Blankenship, and G. R. Fleming, "Two-dimensional spectroscopy of electronic couplings in photosynthesis," Nature **434,** 625 (2005).


Diagonal Cut Through 2D Electronic Spectrum

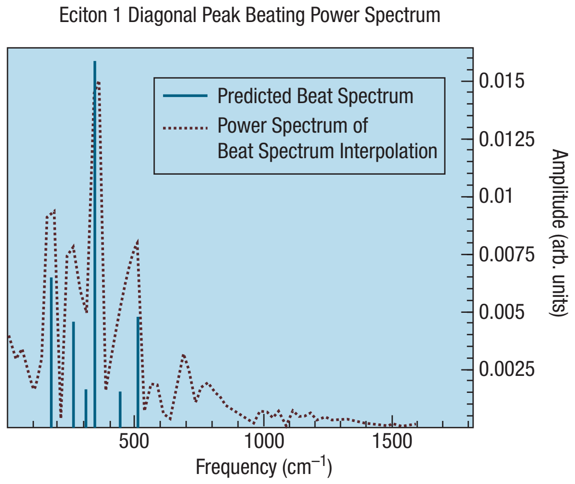

Eciton 1 Diagonal Peak Beating Power Spectrum

Figure 2. Two-dimensional electronic spectroscopy enables scientists to follow the flow of light-induced excitation energy through molecular complexes with femtosecond temporal resolution. In this 2D electronic spectrum, the amplitude of the quantum beating signal for exciton 1 is plotted against population time. The black line covers the exciton 1 peak amplitude. The experimental data's agreement with the computational simulation is shown on the right.

possibility that photosynthetic energy transfer might involve quantum oscillations was first suggested more than 70 years ago, the wavelike motion of excitation energy had never been observed until now."

As in the 2005 paper, the FMO protein was again the target. FMO is considered a model system for studying photosynthetic energy transfer because it consists of only seven pigment molecules, and its chemistry has been well characterized.

"To observe the quantum beats, 2D spectra were taken at 33 population times, ranging from 0 to 660 femtoseconds," said Engel. "In these spectra, the lowest-energy exciton (a bound electron-hole pair formed when an incoming photon boosts an electron out of the valence energy band into the conduction band) gives rise to a diagonal peak near 825 nanometers that clearly oscillates (Figure 2). The associated cross-peak amplitude also appears to oscillate. Surprisingly, this quantum beating lasted the entire 660 femtoseconds."

Engel said the duration of the quantum beating signals was unexpected because the general scientific assumption had been that the electronic coherences responsible for such oscillations are rapidly destroyed.

"For this reason, the transfer of electronic coherence between excitons during relaxation has usually been ignored," Engel said. "By demonstrating that the energy transfer process does involve electronic coherence and that this coherence is much stronger than we would ever have expected, we have shown that the process can be much more efficient than the classical view could explain. However, we still don't know to what degree photosynthesis benefits from these quantum effects."

Simulations provide a preview

"Computational modeling and simulation play a critical role in this kind of research," Fleming said. "Nobody has ever done experiments like this before, so simulations are essential to tell us what the information content of the measurement actually is. I often say to students, it's normal for theory to come in after you've done an experiment, but in our case, the model actually guides the design of the experiment. It tells us what to look for, how to measure it, and how to know what we're looking at."

Yuan-Chung Cheng explained further: "At first we didn't understand why the signal oscillates. To figure that out, we had to run a simulation on a smaller model, which clearly showed that the oscillations correlate with the quantum coherence in the system. It actually represents the energy transfer back and forth in those molecules. So when we figured that out, we could then carry out more large-scale simulations.

"It's not just a single system that we have to simulate," Cheng continued, "because in those protein systems there is an intrinsic static disorder, so each individual protein complex is slightly different from the others."

"So we have to repeat the calculation thousands of times to get the averaged behavior of the real set of molecules," Fleming pointed out.

Elizabeth Read added, "Once we have a decent model of what this protein is doing, that enables us to estimate what the frequencies of oscillation will be, and then that tells us how many time points we need to

Figure 3. Half of the green visible on Earth is the molecule LHC2, which is the next subject of the Fleming group's research. (Composite image created by Reto Stöckli, Nazmi El Saleous, and Marit Jentoft-Nilsen, NASA GSFC.)

measure in order to be able to observe the quantum coherence, and so it all goes into the design of the experiments. Without these models we would have no basis for doing that."

This kind of modeling requires the capacity of a massively parallel computer. Calculating a 2D spectrum for just one set of initial conditions can take several hours because of the large number of molecules and states involved. The results from thousands of independent calculations with different initial parameters must be averaged, requiring up to 20,000 processor-hours for a complete 2D spectrum. The research team used Jacquard, NERSC's 712-processor Opteron cluster, and Bassi, the 888-processor IBM p575 POWER5 system, and were granted an increase in their disk quota to accommodate all the data.

In addition to the 2D spectrum modeling, the Fleming group also performs quantum dynamics simulations to elucidate the energy transfer pathways inside the network of photosynthetic pigment-protein complexes, and to study the effects of coherence transfer on the energy transfer dynamics. These simulations can take up to 1000 processor-hours per job, depending on the number of molecules being modeled. The researchers' goal is to understand the "design rules" that enable the extremely efficient energy transfer in photosynthesis.

One of the next steps for the group will be to look at the effects of temperature changes on the photosynthetic energy transfer process. They will also be looking at broader bandwidths of energy using different colors of light pulses to map out everything that is going on, not just energy transfer. And they plan to begin studying light-harvesting complex 2 (LHC2), which is the most abundant pigment-protein complex in green plants (Figure 3).

"The contribution of green sulfur bacteria to the entire world's energy supply is probably not all that large," Fleming said, "but half the chlorophyll in the world lives in LHC2. It's the single most important light-harvesting protein on earth. We're planning experiments that are somewhat harder to do and even more complicated to interpret because there are 14 cholorophylls of two different kinds in this protein."

Ultimately, the idea is to gain a much better understanding of how Nature not only transfers energy from one molecular system to another, but is also able to convert it into useful forms.

"Nature has had about 2.7 billion years to perfect photosynthesis, so there are huge lessons that remain for us to learn," Engel said. "The results we're reporting in this latest paper, however, at least give us a new way to think about the design of future artificial photosynthesis systems."

This article written by: Lynn Yarris and John Hules (Berkeley Lab).

Bridging the Gap between Climate and Weather

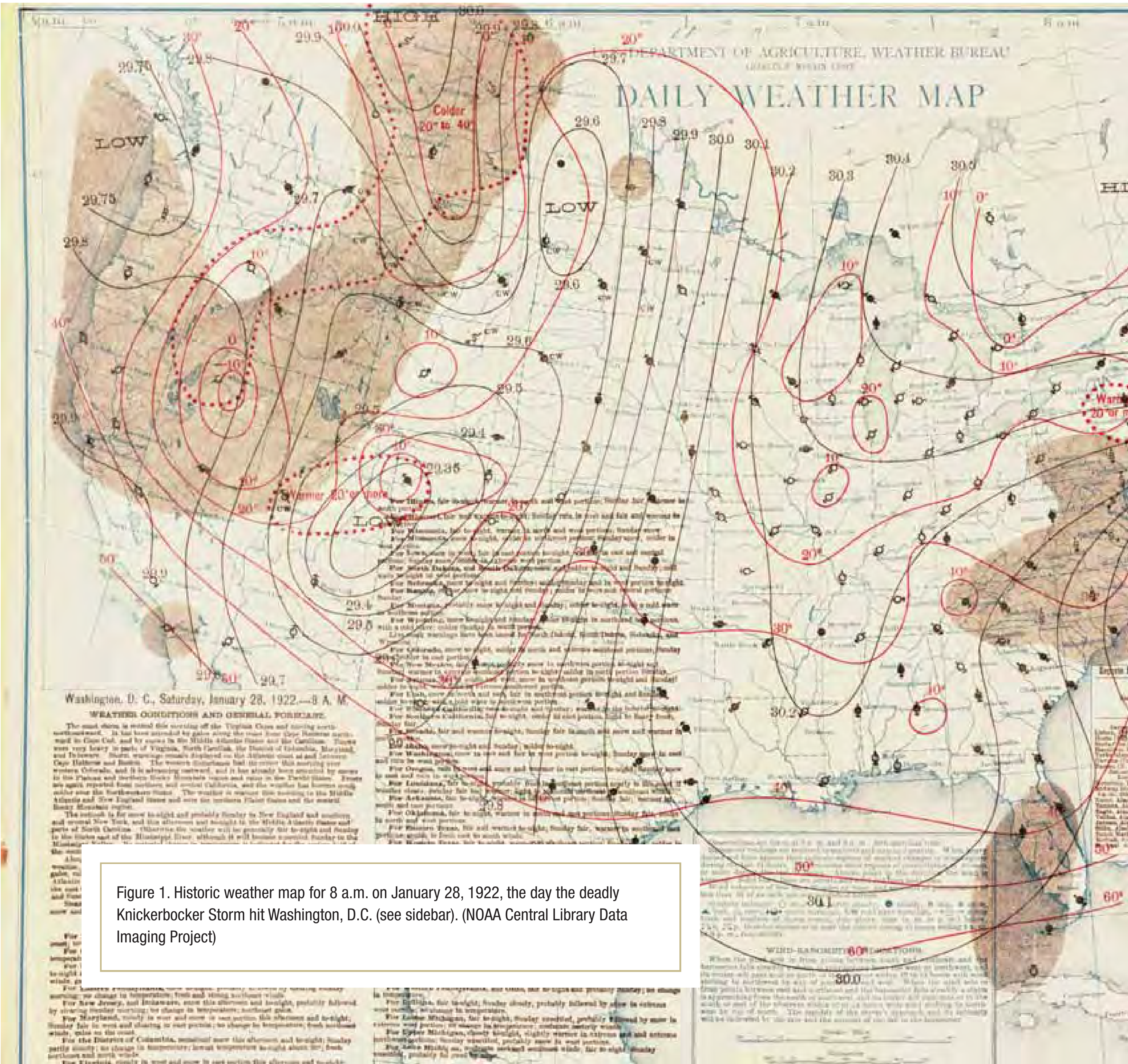

DEPARTMENT OF AGRICULTURE, WEATHER BUREAU
DAILY WEATHER MAP
HIGH
LOW
LOW
LOW
Colder 20° to 40°
Warmer 20° or more
Washington, D. C., Saturday, January 28, 1922.—8 A. M.
WEATHER CONDITIONS AND GENERAL FORECAST.
WIND-BAROMETER INDICATIONS
29.6
29.8
29.9
30.0
30.1
30.2
30.3
30.4
30.5
29.7
29.5
29.4
29.35
10°
20°
30°
40°
50°
60°
0°

Figure 1. Historic weather map for 8 a.m. on January 28, 1922, the day the deadly Knickerbocker Storm hit Washington, D.C. (see sidebar). (NOAA Central Library Data Imaging Project)

A century's worth of reconstructed weather data will provide a better baseline for climate change studies

GH
0°
10°
20°
30°
40°
29.9
29.8
29.7
30.0
LOW

The distinction between climate and weather was expressed most succinctly by science fiction writer Robert A. Heinlein: "Climate is what you expect; weather is what you get." But as global warming produces more noticeable changes on a planetary scale, how do we even know what to expect in a particular region?

Climate change studies are increasingly focused on understanding and predicting regional changes of daily weather statistics. But to predict the next century's statistical trends with confidence, researchers have to demonstrate that their forecasting tools can successfully recreate the conditions of the past century. That requires a detailed set of historical atmospheric circulation data — not just monthly averages, but statistics for at least every six hours, so that phenomena like severe storms can be analyzed.

Although there is scant atmospheric data from weather balloons and none from satellites for the first half of the 20th century, there is an enormous amount of observational data collected at the Earth's surface by a variety of sources, from meteorologists and military personnel to volunteer observers and ships' crews. Until recently, this two-dimensional data was widely available only on hand-drawn weather maps (Figure 1). Despite many errors, these maps are indispensable to researchers, and extensive efforts are being made to put these maps into a digital format and make them available on the Web.

Now, using the latest data integration and atmospheric modeling tools and a 2007 INCITE award of 2 million supercomputing hours at NERSC, scientists from the NOAA Earth System Research Lab and the Cooperative Institute for Research in Environmental Sciences (CIRES) are building the first complete database of three-dimensional global weather maps of the 20th century.

Called the 20th Century Reanalysis Project, the new dataset will double the number of years for which a complete record of three-dimensional atmospheric climate data is available, extending the usable digital dataset from 1948 back to 1892. The team expects to complete the dataset within two years, including observations currently

Project: The 20th Century Reanalysis Project

PI: Gil Compo, University of Colorado/CIRES/Climate Diagnostics Center and NOAA Earth System Research Lab

Senior investigators: Jeffrey Whitaker, NOAA Earth System Research Lab; Prashant Sardeshmukh, University of Colorado/CIRES

Funding: INCITE, CIRES, NOAA

being digitized around the world. The final maps will depict weather conditions every six hours from the Earth's surface to the level of the jet stream (about 11 km or 36,000 ft high), and will allow researchers to compare the patterns, magnitudes, means, and extremes of recent and projected climate changes with past changes.

"We expect the reanalysis of a century's worth of data will enable climate researchers to better address issues such as the range of natural variability of extreme events including floods, droughts, hurricanes, extratropical cyclones, and cold waves," said principal investigator Gil Compo of CIRES. Other team members are Jeff Whitaker of the NOAA Earth System Research Lab and Prashant Sardeshmukh, also of CIRES, a joint institute of NOAA and the University of Colorado.

"Climate change may alter a region's weather and its dominant weather patterns," Compo said. "We need to know if we can understand and simulate the variations in weather and weather patterns over the past 100 years to have confidence in our projections of changes in the future. The alternative — to wait for another 50 years of observations — is less appealing."

From two to three dimensions

Compo, Whitaker, and Sardeshmukh have discovered that using only surface air pressure data, it is possible to recreate a snapshot of other variables, such as winds and temperatures, throughout the troposphere, from the ground or sea level to the jet stream.[1] This discovery makes it possible to extend two-dimensional weather maps into three dimensions. "This was a bit unexpected," Compo said, "but it means that we can use the surface pressure measurements to get a very good picture of the weather back to the 19th century."

The computer code used to combine the data and reconstruct the third dimension has two components. The forecast model is the atmospheric component of the Climate Forecast System, which is used by the National Weather Service's National Centers for Environmental Prediction (NCEP) to make operational climate forecasts. The data assimilation component is the Ensemble Kalman Filter.

Data assimilation is the process by which raw data such as temperature and atmospheric pressure observations are incorporated into the physics-based equations that make up numerical weather models. This process provides the initial values used in the equations to predict how atmospheric conditions will evolve. Data assimilation takes place in a series of analysis cycles. In each analysis cycle, observational data is combined with the forecast results from the mathematical model to produce the best estimate of the current state of the system, balancing the uncertainty in the data and in the forecast. The model then advances several hours, and the results become the forecast for the next analysis cycle.

The Ensemble Kalman Filter is one of the most sophisticated tools available for data assimilation. Generically, a Kalman filter is a recursive algorithm that estimates the state of a dynamic system from a series of incomplete and noisy measurements. Kalman filters are used in a wide range of engineering applications, from radar to computer vision to aircraft and spacecraft navigation. Perhaps the most commonly used type of Kalman filter is the phase-locked loop, which enables radios, video equipment, and other communications devices to recover a signal from a noisy communication channel. Kalman filtering has only recently been applied to weather and climate applications, but the initial results have been so good that the Meteorological Service of Canada has incorporated it into their forecasting code. The 20th Century Reanalysis Project uses the Ensemble Kalman Filter to remove errors in the observations and to fill in the blanks where information is missing, creating a complete weather map of the troposphere.

Rather than making a single estimate of atmospheric conditions at

[1] G. P. Compo, J. S. Whitaker, and P. D. Sardeshmukh, "Feasibility of a 100-year reanalysis using only surface pressure data," Bulletin of the American Meteorological Society **87,** 175 (2006).

Recreating the Knickerbocker Storm of 1922

One of the deadliest snowstorms in U.S. history was the Knickerbocker Storm, a slow-moving blizzard that occurred on January 27–29, 1922 in the upper South and Middle Atlantic states. This storm was named after the collapse of the Knickerbocker Theater in Washington, D.C. shortly after 9 p.m. on January 28. The movie theater's flat roof collapsed under the weight of 28 inches of wet snow, bringing down the balcony and a portion of the brick wall and killing 98 people, including a Congressman.

An arctic air mass had been in place across the Northeast for several days before the storm, and Washington had been below freezing since the afternoon of January 23. The storm formed over Florida on January 26 and took three days to move up the Eastern Seaboard. Snow reached Washington and Philadelphia by noon on January 28 and continued into the morning of January 29. Winds gusting up to 50 mph created blizzard conditions, and heavy drifting blocked roads for days. Railroad lines between Philadelphia and Washington were covered by at least 36 inches of snow, with drifts as high as 16 feet.

Figure 2 presents data from the 20th Century Reanalysis Project's three-dimensional reanalysis of conditions at 7 p.m. on January 28, 1922. With data like this available for the entire 20th century, climate researchers hope to improve their models so that they can more confidently predict regional weather trends for the future.

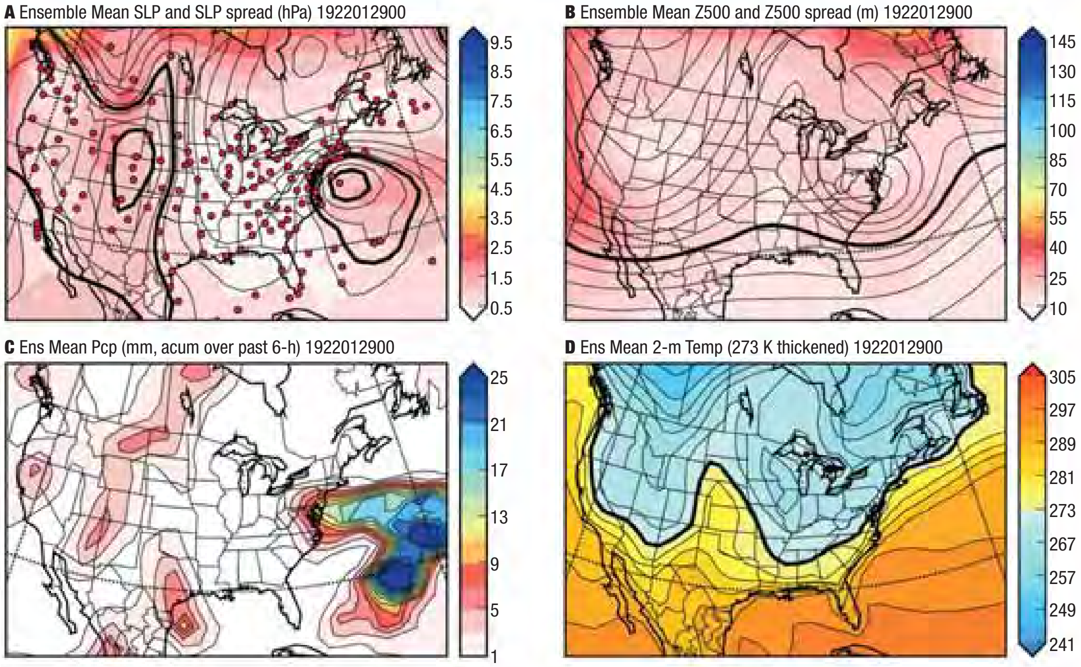

A Ensemble Mean SLP and SLP spread (hPa) 1922012900
9.5
8.5
7.5
6.5
5.5
4.5
3.5
2.5
1.5
0.5
B Ensemble Mean Z500 and Z500 spread (m) 1922012900
145
130
115
100
85
70
55
40
25
10
C Ens Mean Pcp (mm, acum over past 6-h) 1922012900
25
21
17
13
9
5
1
D Ens Mean 2-m Temp (273 K thickened) 1922012900
305
297
289
281
273
267
257
249
241

Figure 2. Reanalysis of conditions at 7 p.m. on January 28, 1922. (A) Sea level pressure (SLP) measured in hectopascals (hPa): contours show the ensemble mean SLP, with 1000 and 1010 hPa contours thickened; colors show the range of uncertainty; red dots indicate observation locations. (B) Height of 500 hPa pressure in meters: contours show the ensemble mean height, with the 5600 m contour thickened; colors show the range of uncertainty. (C) Ensemble mean precipitation accumulated over 6 hours, in millimeters. (D) Ensemble mean temperature (Kelvin) at 2 meters, with the 273 K (0° F) contour thickened. (J. Whitaker, NOAA Earth System Research Lab)

each time step, the Ensemble Kalman Filter reduces the uncertainty by covering a wide range — it produces 56 estimated weather maps (the “ensemble”), each slightly different from the others. The mean of the ensemble is the best estimate, and the variance within the ensemble indicates the degree of uncertainty, with less variance indicating higher certainty. The filter blends the forecasts with the observations, giving more weight to the observations when they are high quality, or to the forecasts when the observations are noisy. The NCEP forecasting system then takes the blended 56 weather maps and runs them forward six hours to produce the next forecast. Processing one month of global weather data takes about a day of computing, with each map running on its own processor. The Kalman filter is flexible enough to change continuously, adapting to the location and number of observations as well as meteorological conditions, thus enabling the model to correct itself in each analysis cycle.

“What we have shown is that the map for the entire troposphere is very good, even though we have only used the surface pressure observations,” Compo said. He estimates that the error for the 3D weather maps will be comparable to the error of modern two- to three-day weather forecasts.

Reanalysis: Reconstructing complete climate data

Traditionally the Earth’s climate has been studied by statistical analysis of weather data such as temperature, wind direction and speed, and precipitation, with the results expressed in terms of long-term averages and variability. But statistical summaries by themselves are inadequate for studies of climate changes; for one thing, many important atmospheric events happen too quickly to be captured in the averages. The ideal historical data set would provide continuous, three-dimensional weather data for the entire globe, collected using consistent methods for a period of at least a century, and more if possible. In reality, weather records are incomplete both spatially and temporally, skewed by changing methods of collecting data, and sprinkled with inaccuracies.

Reanalysis is a technique for reconstructing complete, continuous, and physically consistent long-term climate data. It integrates quality-controlled data obtained from disparate observing systems, then feeds these data into a numerical weather forecasting model to produce short-term forecasts. The output from these forecasts fills in the gaps in the recorded observations both in time and space, resulting in high-resolution, three-dimensional data sets.

Over the past decade, reanalysis data sets have been used in a wide range of climate applications and have provided a more detailed and comprehensive understanding of the dynamics of the Earth’s atmosphere, especially over regions where the data are sparse, such as the poles and the Southern oceans. Reanalysis has also alleviated the impacts of changing observation systems and reduced the uncertainty of climate modeling by providing consistent and reliable data sets for the development and validation of models.

The reanalysis team ran their code on all four of NERSC's large-scale computing systems — Bassi, Jacquard, Seaborg, and Franklin — and switched to a higher-resolution algorithm when they moved to Franklin. "We got fabulous support from the consultants," Compo said, "especially Helen He and David Turner, on porting code, debugging, disk quota increases, using the HPSS, and special software requests." They parallelized virtually their entire workflow on the Franklin architecture via job bundling, writing compute-node shell scripts, and using MPI sub-communicators to increase the concurrency of the analysis code.

Filling in and correcting the historical record

With the 2007 INCITE allocation, the researchers reconstructed weather maps for the years 1918 to 1949. In 2008, they plan to extend the dataset back to 1892 and forward to 2007, spanning the 20th century. In the future, they hope to run the model at higher resolution on more powerful computers, and perhaps extend the global dataset back to 1850.

One of the first results of the INCITE award is that more historical data are being made available to the international research community. This project will provide climate modelers with surface pressure observations never before released from Australia, Canada, Croatia, the United States, Hong Kong, Italy, Spain, and 11 West African nations. When the researchers see gaps in the data, they contact the country's weather service for more information, and the prospect of contributing to a global database has motivated some countries to increase the quality and quantity of their observational data.

The team also aims to reduce inconsistencies in the atmospheric climate record, which stem from differences in how and where atmospheric conditions are observed. Until the 1940s, for example, weather and climate observations were mainly taken from the Earth's surface. Later, weather balloons were added. Since the 1970s, extensive satellite observations have become the norm. Discrepancies in data resulting from these different observing platforms have caused otherwise similar climate datasets to perform poorly in determining the variability of storm tracks or of tropical and Antarctic climate trends. In some cases, flawed datasets have produced spurious long-term trends.

The new 3D atmospheric dataset will provide missing information about the conditions in which early-century extreme climate events occurred, such as the Dust Bowl of the 1930s and the Arctic warming of the 1920s to 1940s. It will also help to explain climate variations that may have misinformed early-century policy decisions, such as the prolonged wet period in central North America that led to overestimates of expected future precipitation and over-allocation of water resources in the Colorado River basin.

But the most important use of weather data from the past will be the validation of climate model simulations and projections into the future. "This dataset will provide an important validation check on the climate models being used to make 21st century climate projections in the recently released Fourth Assessment Report of the Intergovernmental Panel on Climate Change," Compo said. "Our dataset will also help improve the climate models that will contribute to the IPCC's Fifth Assessment Report."

This article written by: John Hules (Berkeley Lab).

A Perfect Sandwich

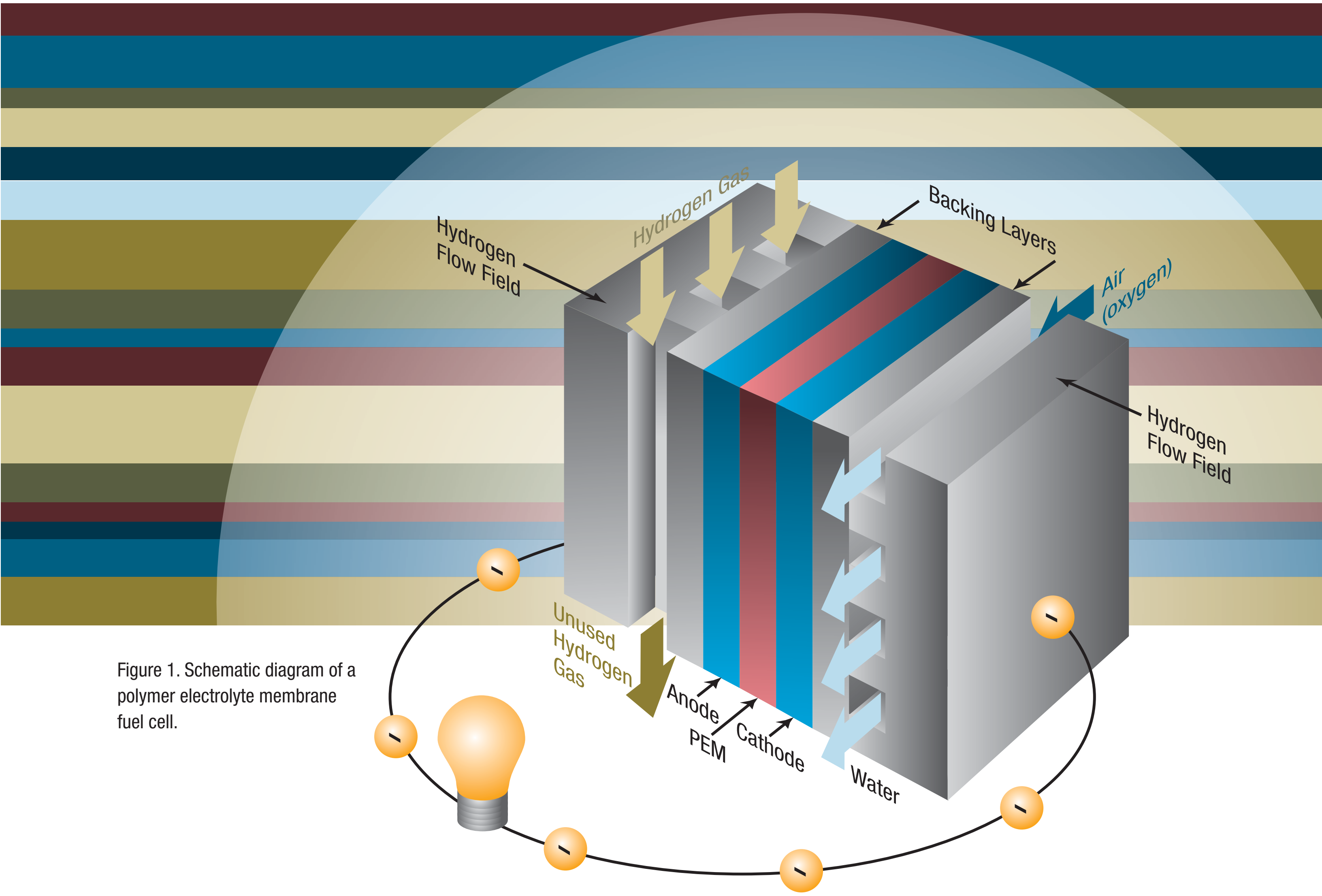
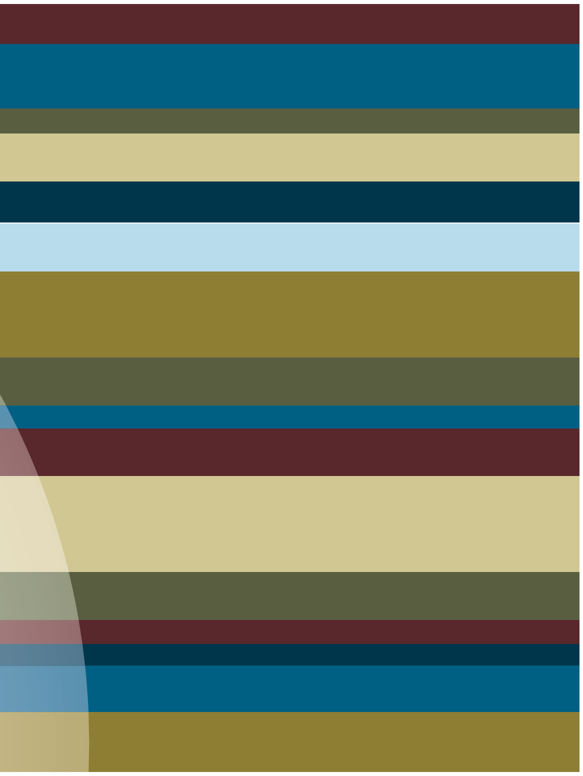


Figure 1. Schematic diagram of a polymer electrolyte membrane fuel cell.

Scientists discover why the right amount of moisture in the membrane plays a key role in fuel cell efficiency



What makes a perfect sandwich? Besides good bread and a tasty combination of fillings and condiments, you need the right amount of moisture to convey the flavor in your mouth. If the sandwich is too dry, it may seem less flavorful, and if it is too soggy, the flavor may seem watered down.

The art of sandwich making may be far removed from the science and technology of hydrogen fuel cells, but in both cases, the amount of moisture in the sandwich is important. In a polymer electrolyte membrane (PEM) fuel cell, the electrolyte membrane is sandwiched between an anode (negative electrode) and a cathode (positive electrode), as shown in Figure 1. After the catalyst in the anode splits the hydrogen fuel into protons and electrons, the PEM transports the protons to the cathode, allowing the separated electrons to flow along an external circuit as an electric current. But the PEM needs the right amount of moisture for efficient proton transport — with too much or too little water, power output will drop.

A fundamental understanding of the relationship between membrane nanostructure and the dynamics of water molecules is needed for the development of efficient, reliable, and cost-effective membranes to advance PEM fuel cell technology. The structure and dynamics of the polymer membranes under different levels of hydration cannot be directly observed in experiments, but they can be modeled in molecular dynamics simulations, as shown in a series of three papers published in the Journal of Physical Chemistry B by Ram Devanathan, Arun Venkatnathan, and Michel Dupuis of Pacific Northwest National Laboratory (PNNL).[1]

"Experimental studies are inadequate to understand proton dynamics, because it occurs below nanoscale," said Devanathan. "This is where NERSC's computing power becomes indispensable. By using advanced computer models, we are getting a grasp of the complex processes at the molecular level in polymer membranes." The simulations for these three papers were run on Jacquard and Bassi.

The research is part of President Bush's Hydrogen Fuel Initiative, which

Project: Charge Transfer, Transport, and Reactivity in Complex Environments

PI: Michel Dupuis, Pacific Northwest National Laboratory

Senior investigators: Ram Devanathan and Arun Venkatnathan, Pacific Northwest National Laboratory

Funding: BES

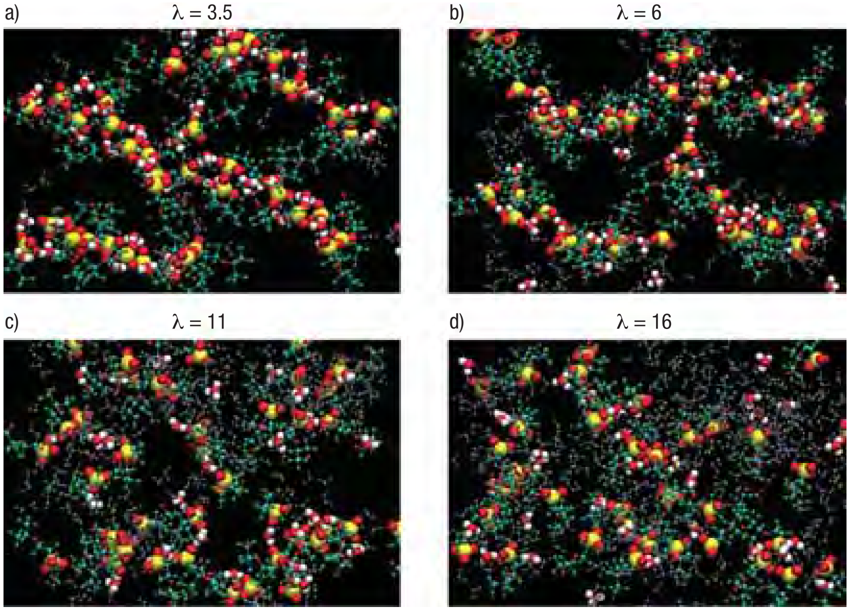
Computing resources: NERSC, MSCF/EMSL

[1] A. Venkatnathan, R. Devanathan, and M. Dupuis, "Atomistic simulations of hydrated Nafion and temperature effects on hydronium ion mobility," J. Phys. Chem. B **111,** 7234 (2007).
R. Devanathan, A. Venkatnathan, and M. Dupuis, "Atomistic simulation of Nafion membrane: 1. Effect of hydration on membrane nanostructure," J. Phys. Chem. B **111,** 8069 (2007).
R. Devanathan, A. Venkatnathan, and M. Dupuis, "Atomistic simulation of Nafion membrane. 2. Dynamics of water molecules and hydronium ions," J. Phys. Chem. B **111,** 13006 (2007).

aims to develop commercially viable hydrogen fuel cells. Using this clean and efficient technology would help to reduce the world's reliance on fossil fuels and lessen greenhouse gas emissions.

The PNNL researchers' three Journal of Physical Chemistry B papers all studied a polymer membrane manufactured by DuPont called Nafion, which has been the subject of numerous experiments and is considered a good starting point for the development of next-generation polymer electrolytes. "Nafion 117 has excellent proton conductivity and good chemical and mechanical stability, but the atomic-level details of its structure at various degrees of hydration are not well characterized or understood," the authors wrote in their first paper, which was featured on the cover of the journal's June 28, 2007 issue (Figure 2).

Figure 2. The June 28, 2007 cover of the Journal of Physical Chemistry B showed snapshots of ionized Nafion and hydronium ions at various degrees of hydration: λ = 3.5 (a), λ = 6 (b), λ = 11 (c), and λ = 16 (d) at 350 K. The black area corresponds to the polymer backbone that is not shown. The pendant side chain (green), sulfonate (yellow and red), hydronium ions (red and white), and water molecules (steel gray) show the structural changes associated with changes in hydration.



Hydration and temperature

In this paper the researchers set out to create simulations that examined the impact of hydration and temperature on the positively charged hydrated protons and water molecules. Understanding these dynamics could lead to polymer membranes that are better engineered for transporting protons while controlling electrode flooding by the water molecules. One of the goals is to develop PEM membranes that need little water.

Using classical molecular dynamics simulations, the research team investigated the impact of four levels of hydration and two different temperatures. The scientists calculated structural properties such as radial distribution functions, coordination numbers, and dynamical properties such as diffusion coefficients of hydronium ions (H_3O^+) and water molecules.

The results of their calculations showed that protons and water molecules are bound to sulfonate groups in the membrane at low hydration levels. As the hydration level increases, the water molecules become free and form a network along which protons can hop (Figure 2). This leads to a dramatic increase in proton conductivity. Temperature was found to have a significant effect on the absolute value of the diffusion coefficients for both water and hydronium ions. These findings have helped in interpreting experimental results that indicate a major structural change taking place in the membrane with increasing hydration.

In the second paper, the authors used all-atom molecular dynamics simulations to systematically examine eight different levels of membrane hydration to closely mirror two experimental studies. They also simulated bulk water to develop a novel criterion to identify free water in Nafion. This enabled them to quantify the fraction of free, weakly bound, and bound water molecules in the membrane as a function of hydration.

The researchers found that at low hydration levels, strong binding of hydronium ions to sulfonate groups prevents transport of protons. Multiple sulfonate groups surrounding the hydronium ions in bridging configurations hinder the hydration of the hydronium and the structural diffusion of protons (Figure 3). As the hydration level increases, the water molecules mediate the interaction between hydronium ions and sulfonate groups, moving them farther apart. These results provide atomic-level insights into structural changes observed in Nafion by infrared spectroscopy.

Structure and dynamics

In the third report, Devanathan, Venkatnathan, and Dupuis computed the dynamical properties of water molecules and hydronium ions in Nafion and related them to the structural changes reported previously. They confirmed other researchers' finding that the behavior of water molecules within nanoscale pores and channels of PEMs, especially at low hydration levels, is remarkably different from that of molecules in bulk water.

At low hydration, fewer than 20% of the water molecules are free (bulklike). With increasing hydration, the diffusion coefficients of hydronium ions and water molecules increase, and the mean residence time of water molecules around sulfonate groups decreases. These results provide a molecular-level explanation for the proton and water dynamics observed in neutron scattering experiments.

Because the structure and dynamics of the membrane under different levels of hydration cannot be directly observed in experiments, there is no universally accepted model of the structure of Nafion. This research makes a significant step toward that goal and toward the development of the next generation of PEMs.

Characteristics of the ideal PEM include high proton conductivity at low hydration levels; thermal, mechanical, and chemical stability; durability under prolonged operation; and low cost. None of the existing membranes meet all these requirements, and developing new membranes requires a molecular-level understanding of membrane chemistry and nanostructure. Molecular dynamics simulations like these, together with experiments, are laying the foundation for future breakthroughs in fuel cells.

This article written by: John Hules and Ucilia Wang, Berkeley Lab.

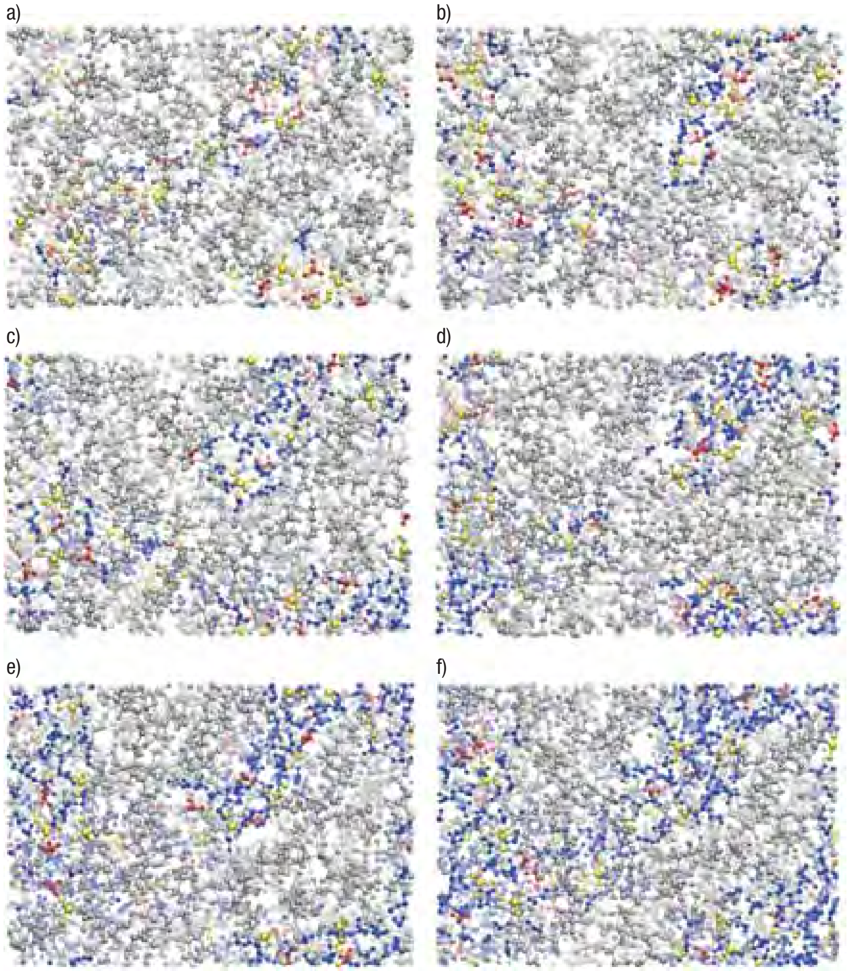
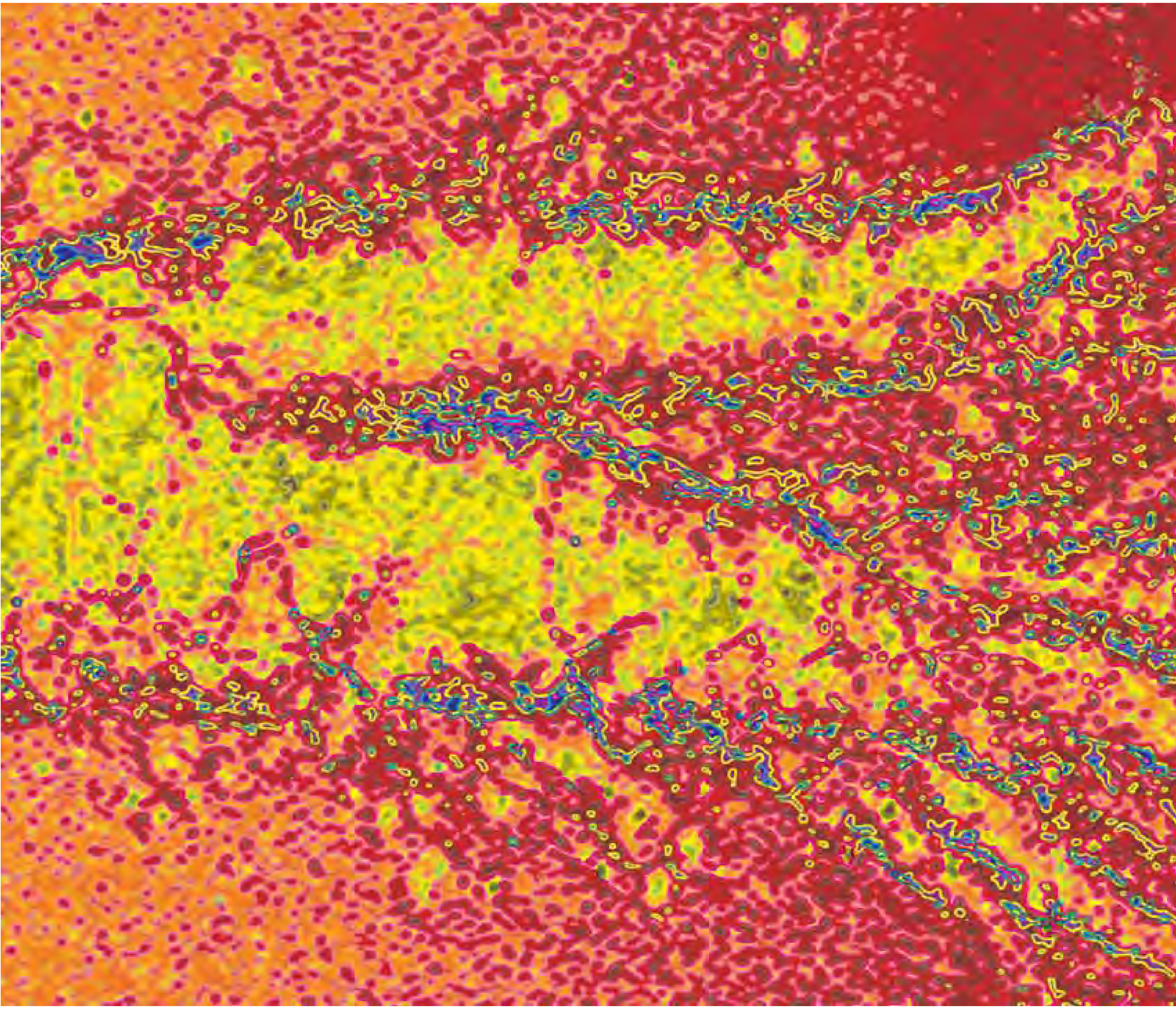
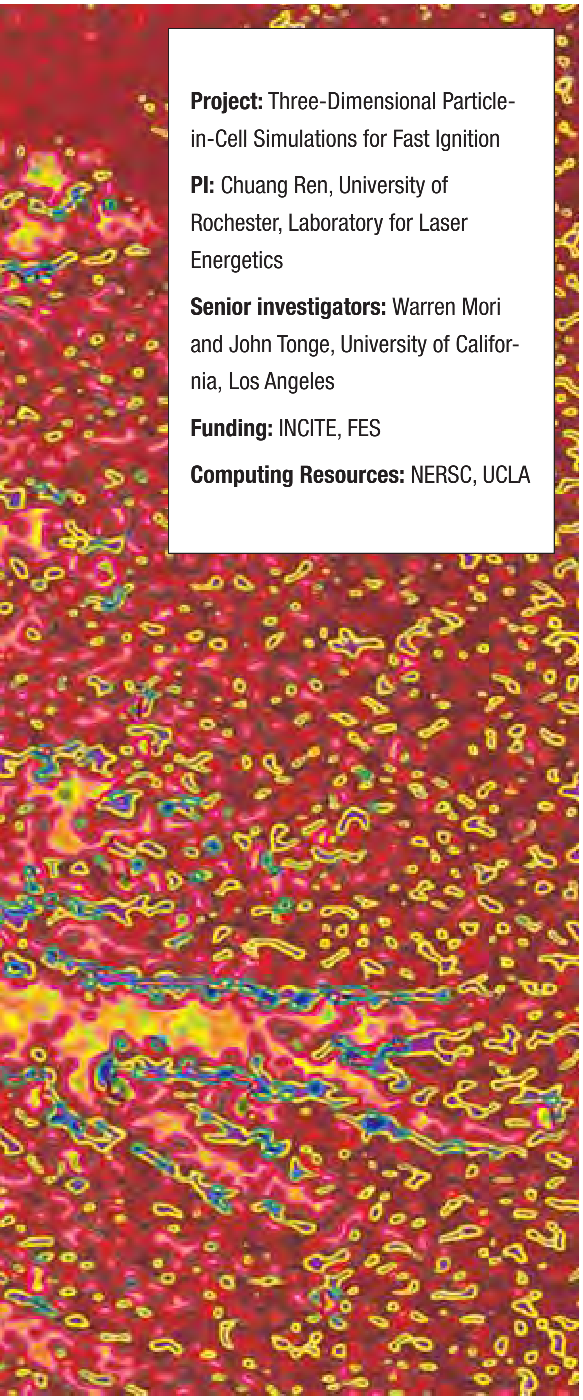


Figure 3. Orthographic projection (~42 Å × 30 Å) of hydrated Nafion for the following λ values: (a) 3; (b) 5; (c) 7; (d) 9; (e) 11; and (f) 13.5. Water molecules, hydronium ions, sulfonate groups, and the rest of the membrane are represented in blue, red, yellow, and gray, respectively.

The Fusion Two-Step



Simulations elucidate the physics of fast ignition



Project: Three-Dimensional Particle-in-Cell Simulations for Fast Ignition

PI: Chuang Ren, University of Rochester, Laboratory for Laser Energetics

Senior investigators: Warren Mori and John Tonge, University of California, Los Angeles

Funding: INCITE, FES

Computing Resources: NERSC, UCLA

To a dance aficionado, the term *two-step* may refer to the ballroom dance that evolved into the foxtrot, or to country/western dances like the Texas two-step and the Cajun two-step. But in the realm of alternative energy sources, one of the hottest new trends is the two-step fast ignition concept for inertial confinement fusion (ICF).

ICF is the process of initiating a nuclear fusion reaction by heating and compressing a fuel target, usually a pellet of deuterium-tritium (DT) ice. If a 10 milligram DT fuel pellet was completely consumed by fusion, it would release energy equivalent to more than half a barrel of oil.

Until recently, the most common approach to ICF has been *hot-spot ignition,* in which the fuel pellet is compressed and heated in one step by a multi-beam laser. This is the concept around which the National Ignition Facility (NIF), scheduled to be completed in 2009 at Lawrence Livermore National Laboratory, was designed.

In the newer *fast ignition* concept, compression and ignition are separated into two steps: first, a compression laser compresses a spherical shell of DT ice to high density at low temperature, without a central hot spot; then a second very high-intensity laser delivers an extremely short pulse of energy that ignites the compressed fuel.

The two ignition concepts are sometimes compared to a diesel engine (pure compression) and a gasoline engine (where the fast ignitor is equivalent to a spark plug). Compared to hot-spot ignition, fast ignition promises much higher gain (the ratio of energy output to energy input) for the same driver energy, possible reduction of the driver energy necessary to achieve ignition, and less stringent compression symmetry requirements.

The ignition step is the least understood aspect of fast ignition and is the subject of the INCITE project "Three-Dimensional Particle-in-Cell Simulations for Fast Ignition," led by Chuang Ren, Assistant Professor of Mechanical Engineering and Physics at the University of Rochester, and Warren Mori, Professor of Physics and Electrical Engineering at UCLA.

The hole-boring scheme

Since the ignition laser cannot directly reach the dense core region, the laser energy needs to be converted into an energetic (super-hot) electron beam that can penetrate to the core and deposit its energy there. The electron beam needs to be generated as close to the core as possible to reduce energy loss along its path to the core. One way to do that is the *hole-boring scheme,* in which the ignition laser pulse is preceded by a channeling laser pulse to create a channel through the underdense corona and into a critical density surface, beyond which it cannot penetrate (Figure 1). The ignition pulse is then sent in tandem to reach the critical surface and may continue to push forward into the overdense plasma, in the meantime heating the plasma to generate the energetic electron beam. This beam will penetrate through the dense plasma and deposit the energy in the core, heating a small area to a high-enough temperature to ignite the fusion reaction.

"Compressing the fuel to high density is relatively easy to achieve, but high temperature is more challenging," Ren said. "You need to convert the laser energy into electron energy, and these electrons need to be collimated [focused], because if they spread to a large area, you need much more energy to heat a large area. You need to focus the laser down to a very small spot, say a 20 micron radius, and make the electrons go forward, not just spread."

Ren continued: "The ignition laser has to heat the pellet in a very short time, 10 to 20 picoseconds, so its intensity is a lot higher than the compression laser, which works on the time scale of 1 nanosecond and has intensity of 10^{14} watts per square centimeter. But the ignition laser will need intensity of 10^{19} to 10^{20} watts per square centimeter. So we understand less about the interactions between the more intense ignition beam and a plasma. Our computation is about this process."

Particle-in-cell (PIC) methods provide the best available simulation tool to understand the highly nonlinear and kinetic physics in the ignition phase. Ren's project covers almost all the physics in the ignition phase with the goal of answering the following questions:

1. Can a clean channel be created by a channeling pulse so that the ignition pulse can arrive at the critical surface without significant energy loss?
2. What are the amount and spectrum of the laser-generated energetic electrons?
3. What is the energetic electron transport process beyond the laser-plasma interface in a plasma with densities up to 10^{23} cm^{-3}?

Millimeter-scale simulations

Most of the previous channeling experiments and simulations were done in 100 micrometer-scale plasmas, but the underdense region of an actual fast ignition target is 10 times longer. "If you extend the experiment 10 times longer in both size and time, you see new phenomena," Ren said.

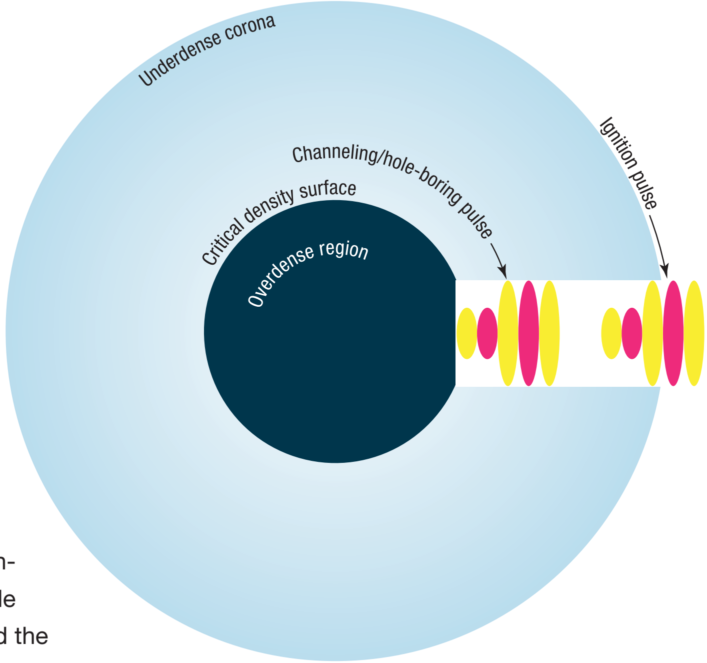


Figure 1. A sketch of the hole-boring scheme for fast ignition.

Using the OSIRIS code, which can run on over 1000 processors with more than 80% efficiency, Ren's team ran the first 2D simulations of channeling at the millimeter scale.[1] These simulations, which ran on Seaborg and Bassi, employed up to 8 × 10^7 grids, 10^9 particles, and 10^6 time steps. NERSC's User Services Group increased the researchers' disk quota and queue priority to accommodate the scale of these calculations. The Analytics Team also assisted by reducing network latency for remote performance of Xlib-based applications.

The results of the simulations showed important new details of the channeling process, including plasma buildup in front of the laser, laser hosing (an undulating motion like water coming out of a garden hose), and channel bifurcation and

[1] G. Li, R. Yan, C. Ren, T.-L. Wang, J. Tonge, and W. B. Mori, "Laser channeling in millimeter-scale underdense plasmas of fast-ignition targets," Physical Review Letters **100,** 125002 (2008).

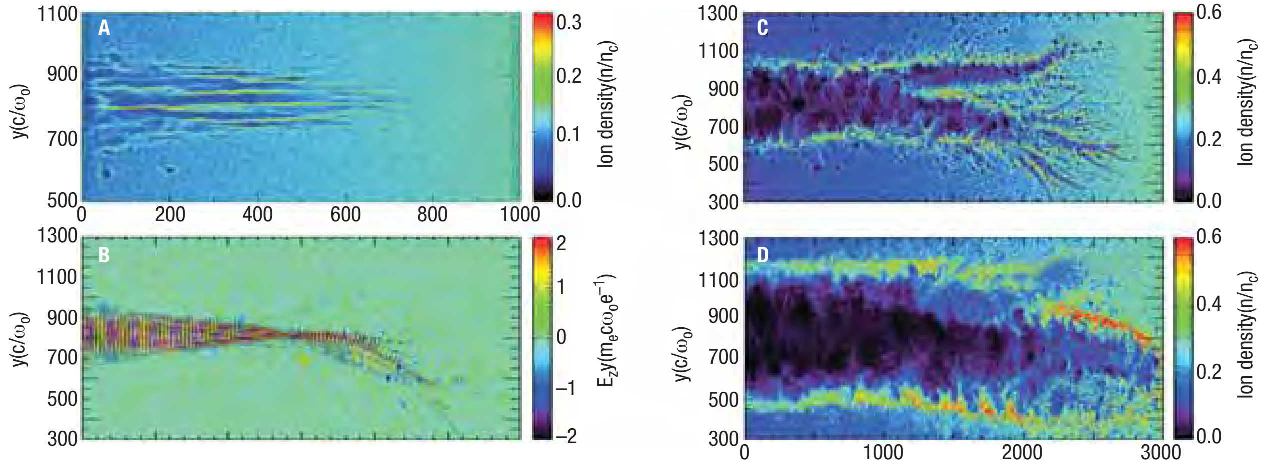


Figure 2. Simulation of laser channeling through an underdense plasma. (a) Ion density at t = 0.8 ps showing micro channels formed; (b) laser E-field showing laser hosing and (c) ion density showing channel bifurcation at t = 3.4 ps; (d) ion density at t = 7.2 ps showing channel self-correction.

self-correction (Figure 2). The simulations demonstrated electron heating to relativistic temperatures, a channeling speed much less than the linear group velocity of the laser, and increased transmission of an ignition pulse in a preformed channel.

The simulation results also shed light on the question of how to save energy during the channeling process. “You want to spend as little energy as possible on creating the channel and save it for ignition,” Ren said, “so what kind of laser intensity do you use for the channeling? High intensity will create a channel more quickly, but you may spend more energy. You can also use a low intensity laser but it takes longer, so it was not clear before how to minimize that. We showed that a lower intensity laser takes more time to produce the channel but does it effectively using less energy than a high intensity laser. This result will provide some guidance for designing experiments.”

The group’s 2D simulations of hot electron generation and transport were the largest ever in target size and the first with isolated targets. These simulations employed 5×10^8 grids, 10^9 particles, and 10^5 time steps. The results showed that the temperature of electrons emitted at high laser intensities is only half of that predicted by an empirical formula used in many fast ignition feasibility studies. The simulations also found that the laser absorption rate increases with the laser intensity. “Higher-intensity lasers are desirable since they bore a deeper hole and deliver their energy to a smaller area, creating a hot spot of higher temperature,” Ren explained. The combination of these effects means that ultra-high-intensity lasers can produce an electron flux with a majority of electrons in the usable energy range.

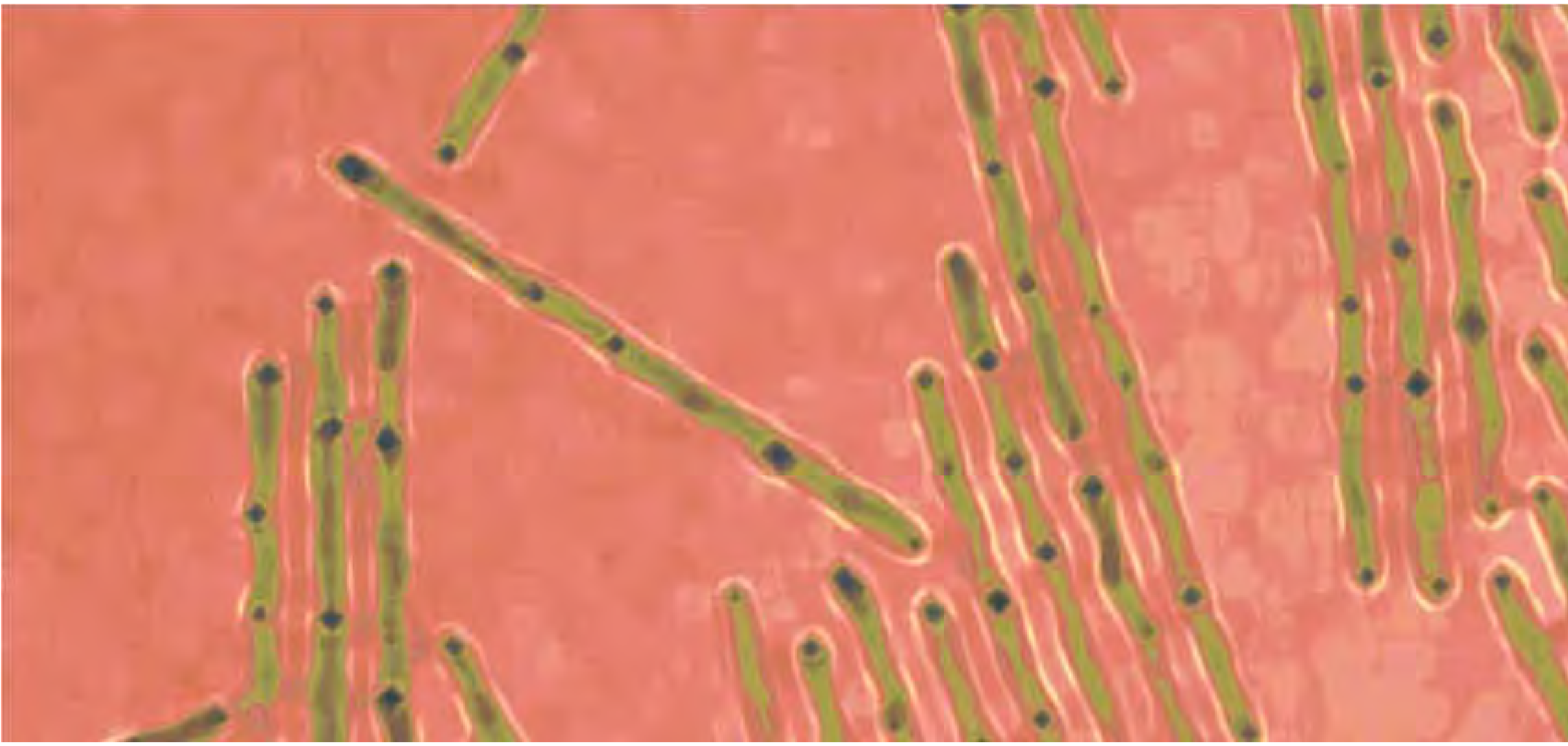
“But there are important effects that cannot be simulated in two dimensions,” Ren pointed out. “Simply scaling up our 2D simulations to 3D would require more than a 4000-fold computation increase. That is not feasible even on the largest computers available. So we will combine 3D simulations at reduced scales with full-scale 2D results and theory to figure out what happens.”

Ren is a researcher at the University of Rochester’s Laboratory for Laser Energetics, where the OMEGA EP laser system, the world’s leading system for fast ignition experiments, is scheduled to be completed in April 2008. He also collaborates with other investigators in the DOE’s Fusion Science Center for Extreme States of Matter and Fast Ignition Physics, which coordinates research in all aspects of fast ignition. Fast ignition is also going to be tested at the FIREX facility in Japan and the Z machine at Sandia National Laboratories. The National Ignition Facility could be adapted for full-scale fast ignition experiments, and the proposed HiPER facility in Europe is being designed for just that purpose. All of these experiments will benefit from the insights gained in Ren’s simulations, which gives his work a sense of urgency.

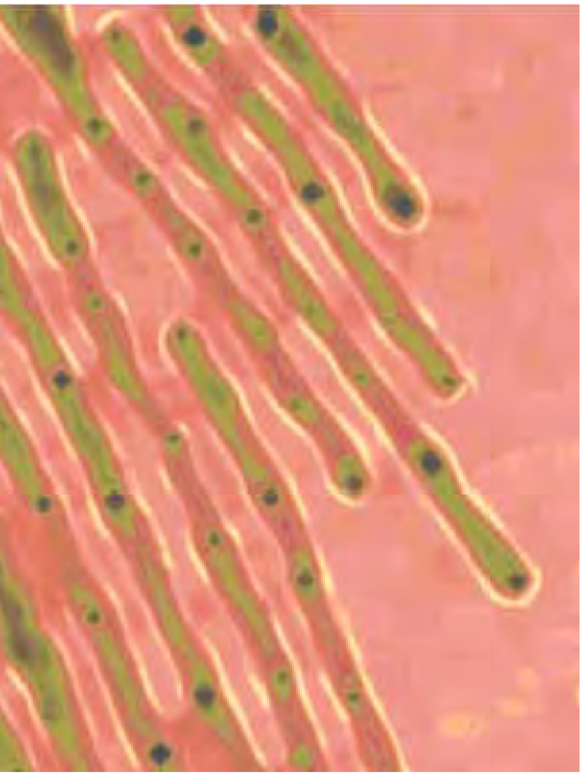
“This research will help toward the realization of fusion as a controllable energy source, and can help solve the energy crisis facing the world today,” he concluded.

This article written by:
John Hules, Berkeley Lab.

Spontaneous Superlattice



Ab initio calculations and modeling contribute to the discovery of a new way to fabricate striped nanorods



Superlatticed or "striped" nanorods — crystalline materials only a few molecules in thickness and made up of two or more semiconductors — are highly valued for their potential to serve in a variety of nanodevices, including transistors, biochemical sensors, and light-emitting diodes (LEDs). Until now the potential of superlatticed nanorods has been limited by the relatively expensive and exacting process required to make them. That paradigm may be shifting.

A team of researchers with Lawrence Berkeley National Laboratory (Berkeley Lab) and the University of California (UC) at Berkeley, has found a way to make striped nanorods in a colloid — a suspension of particles in solution. Previously, striped nanorods were made through epitaxial processes, in which the rods were attached to or embedded within a solid medium.

"We have demonstrated the application of strain engineering in a colloidal quantum-dot system by introducing a method that spontaneously creates a regularly spaced arrangement of quantum dots within a colloidal quantum rod," said chemist Paul Alivisatos, who led this research. "A linear array of quantum dots within a nanorod effectively creates a one-dimensional superlattice, or striped nanorod."

Alivisatos, an internationally recognized authority on colloidal nanocrystal research, is the Director of the Materials Sciences Division and Associate Laboratory Director for Physical Sciences at Berkeley Lab, and is the Larry and Diane Bock Professor of Nanotechnology at UC Berkeley. Collaborators on this project, which culminated in a paper published in the journal Science[1], were Richard Robinson of Berkeley Lab's Materials Sciences Division (lead author), Denis Demchenko and Lin-Wang Wang of Berkeley Lab's Computational Research Division; and Bryce Sadtler and Can Erdonmez, of the UC Berkeley Department of Chemistry.

One-dimensional fabrication

Today's electronics industry is built on two-dimensional semiconductor materials that feature carefully controlled doping and interfaces. Tomorrow's

Project: Large Scale Nanostructure Electronic Structure Calculations

PI: Lin-Wang Wang, Lawrence Berkeley National Laboratory

Senior investigators: Byounghak Lee, Joshua Schrier, Denis Demchenko, Nenad Vukmirovic, Sefa Dag, Lawrence Berkeley National Laboratory

Funding: BES, ASCR

[1] R. D. Robinson, B. Sadtler, D. O. Demchenko, C. K. Erdonmez, L. W. Wang, and A. P. Alivisatos, "Spontaneous superlattice formation in nanorods through partial cation exchange," Science **317,** 355 (2007).

industry will be built on one-dimensional materials, in which controlled doping and interfaces are achieved through superlatticed structures. Formed from alternating layers of semiconductor materials with wide and narrow band gaps, superlatticed structures, such as striped nanorods, can display not only outstanding electronic properties, but photonic properties as well.

"A target of colloidal nanocrystal research has been to create superlatticed structures while leveraging the advantages of solution-phase fabrication, such as low-cost synthesis and compatibility in disparate environments," Alivisatos said. "A colloidal approach to making striped nanorods opens up the possibility of using them in biological labeling, and in solution-processed LEDs and solar cells."

Previous research by Alivisatos and his group had shown that the exchange of cations could be used to vary the proportion of two semiconductors within a single nanocrystal without changing the crystal's size and shape, so long as the crystal's minimum dimension exceeded four nanometers. This led the group to investigate the possibility of using a partial exchange of cations between two semiconductors in a colloid to form a superlattice. Working with previously formed cadmium-sulfide (CdS) nanorods, they engineered a cation exchange with free-standing quantum dots of the semiconductor silver sulfide (Ag_2S) (Figure 1).

"We found that a linear arrangement of regularly spaced silver sulfide contained within a cadmium-sulfide nanorod forms spontaneously at a cation exchange rate of approximately 36 percent," said Alivisatos. "The resulting striped nanorods display properties expected of an epitaxially prepared array of silver-sulfide quantum dots separated by confining regions of cadmium sulfide. This includes the ability to emit near-infrared light, which opens up potential applications such as nanometer-scale optoelectronic devices."

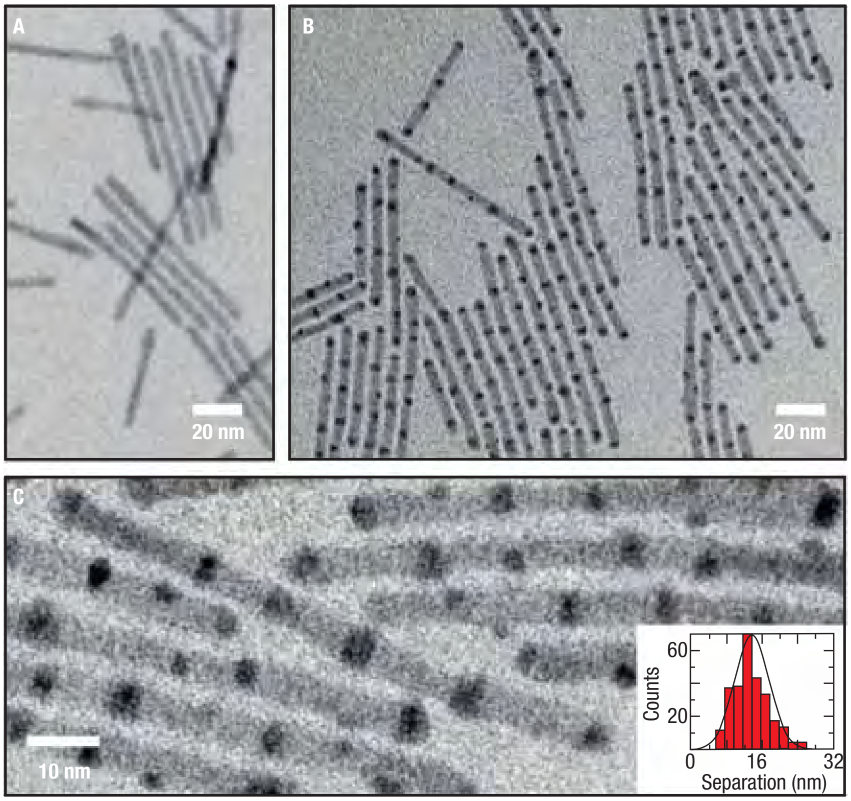


Figure 1. In these transmission electron microscope images of superlatticed or striped nanorods formed through partial cation exchange, (A) shows the original cadmium-sulfide nanorods; (B and C) show cadmium-sulfide nanorods striped with silver-sulfide. The inset is a histogram showing the pattern spacing of the silver-sulfide stripes.

Strain engineering

One of the key difference between quantum dots epitaxially grown on a substrate and freestanding colloidal quantum dots is the presence of strain. The use of temperature, pressure, and other forms of stress to place a strain on material structures that can alter certain properties is called "strain engineering." This technique is used to enhance the performance of today's electronic devices, and has recently been used to spatially pattern epitaxially grown striped nanorods.

However, strain engineering in epitaxially produced striped nanorods requires clever tricks, whereas Demchenko and Wang discovered — through *ab initio* calculations of the interfacial energy and computer modeling of strain energies — that naturally occurring strain in the colloidal process would be the driving force that induced

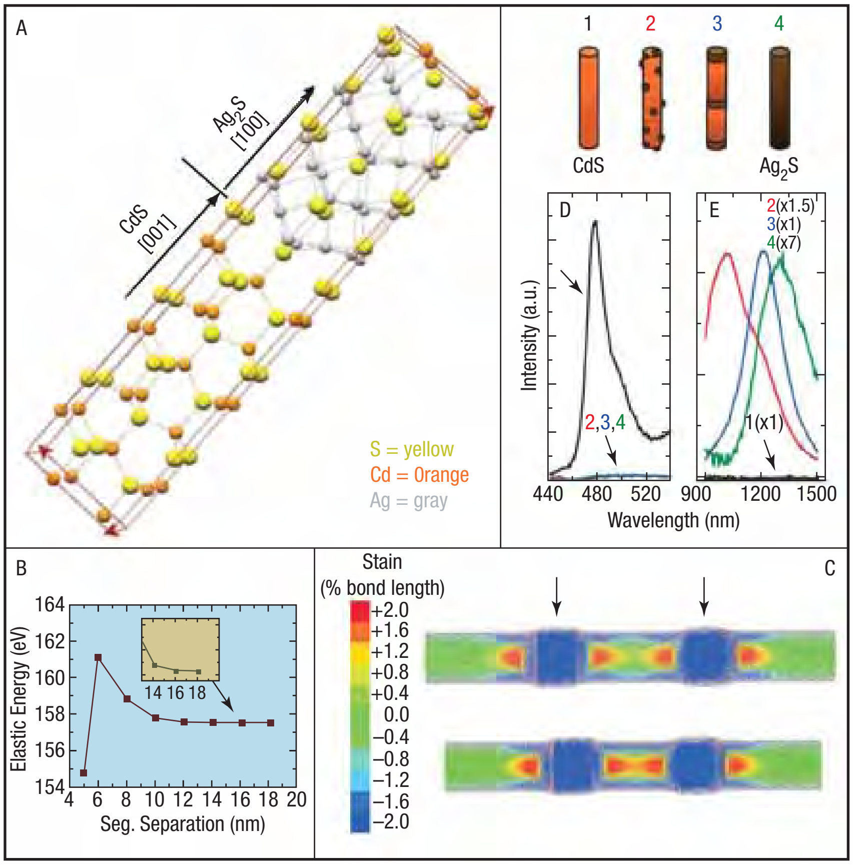


Figure 2. Theoretical modeling and experimental optical characterization. (A) Cubic-cutout representation of cells used for *ab initio* energy calculations. A distorted monoclinic Ag$_2$S (100) plane connects with the wurtzite CdS (001) plane. (B) Elastic energy of the rod as a function of segment separation (center-to-center). (C) Z-axis strain for the case of two mismatched segments at a center-to-center separation distance of 14.1 nm (top) and 12.1 nm (bottom). The elastic interaction between segments is greatly reduced for separations >12.1 nm. Arrows show the placement of mismatched segments. (D) Visible and (E) near-infrared photoluminescence spectra at λ = 400- and 550-nm excitation, respectively. Coupling between the CdS and Ag$_2$S is evident by the complete quenching of the visible photoluminescence (D) in the heterostructures. The shift in near-infrared photoluminescence (E) is due to quantum confinement of the Ag$_2$S.

the spontaneous formation of the superlattice structures (Figure 2). This is the first time that the elastic energy has been shown to be responsible for pattern formation in a colloidal nanostructure.

Demchenko and Wang performed the *ab initio* calculations of the electronic structure of Ag$_2$S and CdS on Seaborg and Bassi using the Vienna Ab-Initio Simulation Package (VASP) and Parallel Total Energy (PEtot) codes, utilizing the local density approximation and generalized gradient approximation to the density functional theory. These techniques were used to estimate stability of various Ag$_2$S phases, find the optimal geometry for the epitaxial attachment, calculate the formation energies of the CdS–Ag$_2$S interfaces, and calculate the corresponding band alignment. Elastic energies and strains were estimated using the valence force field (VFF) method, which is an atomistic bond stretching and bending model. The researchers received assistance with code installation and testing from Zhengji Zhao, a materials science and chemistry specialist in NERSC's User Services Group.

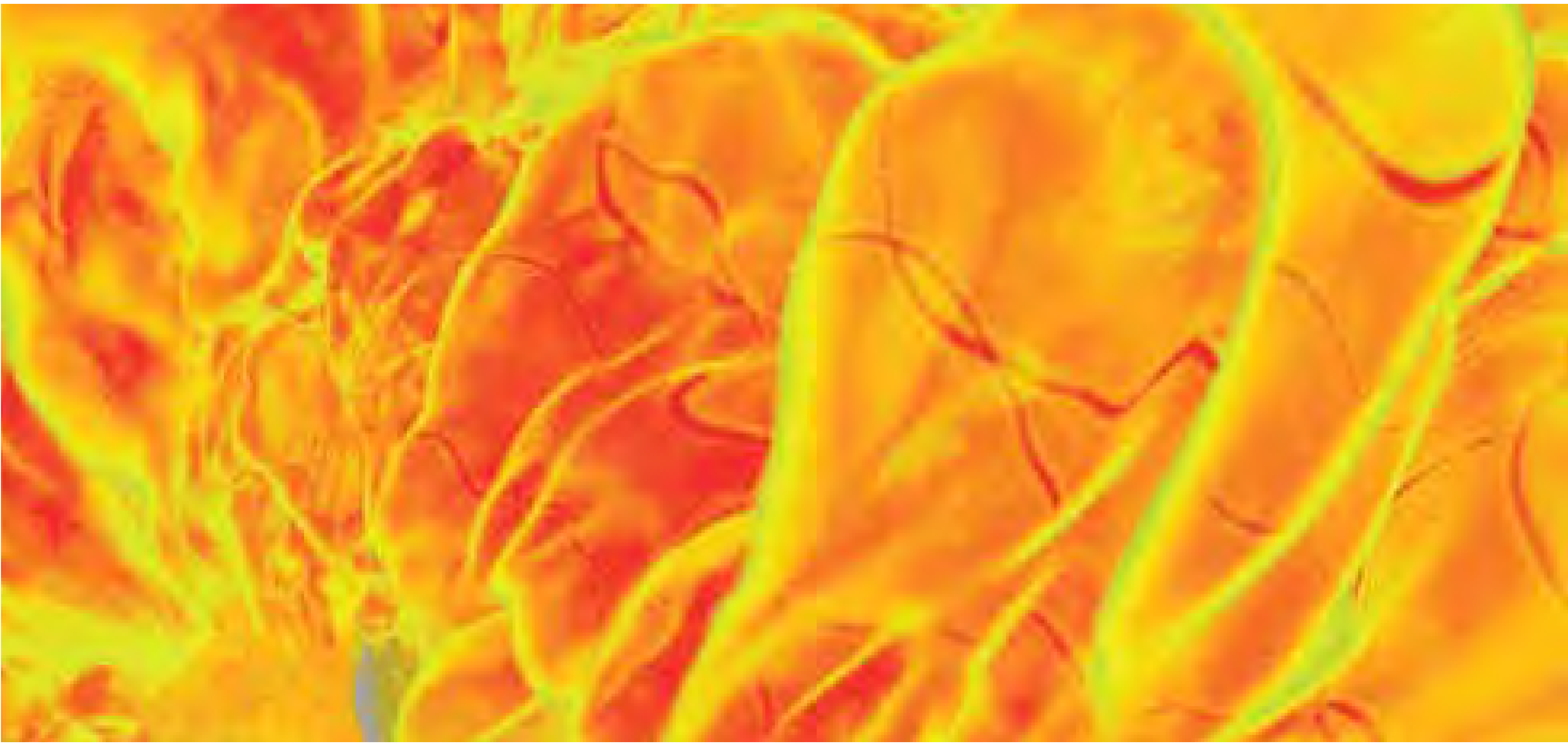
"This project has involved tight coordination between computer simulations and experiment, and the results obtained here would not have been possible to achieve without the contributions of our computational scientists, Demchenko and Wang," Alivisatos said. "It is another clear example where we see that theoretical simulations are not just being used to explain materials growth after the fact, but are now an integral part of the materials design and creation process from the very start."

Even though the colloidal striped nanorods form spontaneously, Alivisatos said it should be possible to control their superlatticed pattern — hence their properties — by adjusting the length, width, composition, etc., of the original nanocrystals. However, much more work remains to be done before the colloidal method of fabricating striped nanorods can match some of the "spectacular results" that have been obtained from epitaxial fabrication.

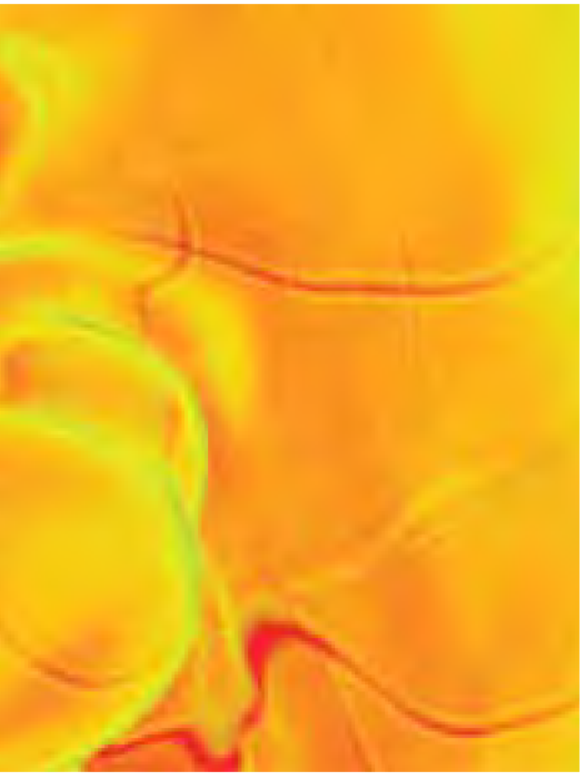
"For now, the value of our work lies in the unification of concepts between epitaxial and colloidal fabrication methods," he said.

This article written by: Lynn Yarris and John Hules (Berkeley Lab).

Igniting a Stellar Explosion



Flash Center achieves the first 3D simulation of the spontaneous detonation of a white dwarf star



University of Chicago scientists demonstrated how to incinerate a white dwarf star in unprecedented detail at the "Paths to Exploding Stars" conference on March 22, 2007, in Santa Barbara, Calif.

White dwarf stars pack one and a half times the mass of the sun into an object the size of Earth. When they burn out, the ensuing explosion produces a type of supernova that astrophysicists believe manufactures most of the iron in the universe. These type Ia supernovas, as they are called, may also help illuminate the mystery of dark energy, an unknown force that dominates the universe.

"That will only be possible if we can gain a much better understanding of the way in which these stars explode," said Don Lamb, Director of the University of Chicago's Center for Astrophysical Thermonuclear Flashes.

Scientists for years have attempted to blow up a white dwarf star by writing the laws of physics into computer software and then testing it in simulations. At first the detonations would only occur if inserted manually into the programs. Then the Flash team naturally detonated white dwarf stars in simplified, two-dimensional tests, but "there were claims made that it wouldn't work in 3D," Lamb said.

But in January 2007, the Flash Center team for the first time naturally detonated a white dwarf in a more realistic three-dimensional simulation. The simulation confirmed what the team already suspected from previous tests: that the stars detonate in a supersonic process resembling diesel-engine combustion, which they call *gravitationally confined detonation.*[1]

Unlike a gasoline engine, in which a spark ignites the fuel, it is compression that triggers ignition in a diesel engine. "You don't want supersonic burning in a car engine, but the triggering is similar," said Dean Townsley, a Research Associate at the Joint Institute for Nuclear Astrophysics at Chicago.

The temperature attained by a detonating white dwarf star makes the 10,000-degree surface of the sun seem like a cold winter day in Chicago by comparison. "In nuclear explosions, you deal with temperatures on the order of a billion degrees," said Flash Center Research Associate Cal Jordan.

The new 3D white dwarf simulation shows the formation of a flame bubble near the center of the star. The bubble, initially measuring approximately 10 miles in diameter, rises more than 1,200 miles to the surface of the star in one

Project: Validation Study of Fundamental Properties of Type Ia Supernovae Models

PI: Don Lamb, University of Chicago Center for Astrophysical Thermonuclear Flashes

Senior investigators: Robert Fisher, Anshu Dubey, Jim Truran, Carlo Graziani, Dean Townsley, Cal Jordan, Casey Meakin, University of Chicago

Funding: INCITE, ASC, NSF

Computing Resources: NERSC, LLNL

[1] G. C. Jordan IV, R. T. Fisher, D. M. Townsley, A. C. Calder, C. Graziani, S. Asida, D. Q. Lamb, and J. W. Truran, "Three-dimensional simulations of the deflagration phase of the gravitationally confined detonation model of Type Ia supernovae," Astrophysical Journal **681,** 1448–1457 (2008).

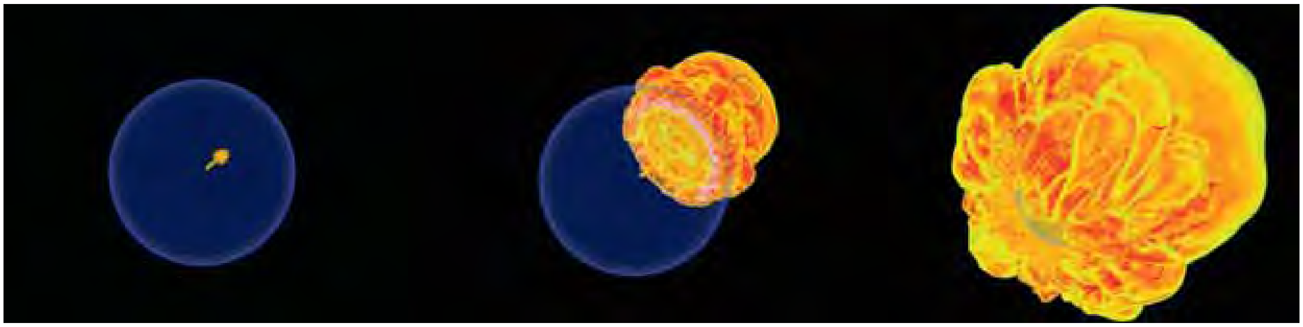


Figure 1. Three phases of the gravitationally confined detonation mechanism. The images show the flame surface (orange) and the star surface (blue) (a) at 0.5 s, soon after the bubble becomes unstable and develops into a mushroom shape, (b) at 1.0 s, as the bubble breaks through the surface of the star, and (c) at 1.7 s, shortly before the hot ash from the bubble collides at the opposite point on the surface of the star. These images are generated from volume-renderings of the flame surface and the density.

second. In another second, the flame crashes into itself on the opposite end of the star, triggering a detonation. "It seems that the dynamics of the collision is what creates a localized compression region where the detonation will manifest," Townsley said. Figure 1 shows stages of the flame erupting and enveloping the star, while Figure 2 shows the entire process from flame formation to detonation.

Extreme computing

This process plays out in no more than three seconds, but the simulations take considerably longer. The Flash Center team ran its massive simulation on two powerful supercomputers at Lawrence Livermore National Laboratory and at NERSC. Just one of the jobs ran for 75 hours on 768 computer processors, for a total of 58,000 hours.

"I cannot say enough about the support we received from the high-performance computing teams at Lawrence Livermore and NERSC," Lamb said. "Without their help, we would never have been able to do the simulations."

Katie Antypas, an HPC consultant in NERSC's User Services Group, worked closely with Lamb to run the simulations on Seaborg. "With help and input from many people at NERSC, from setting up accounts, allocating terabytes of disk space, granting file sharing permissions to analyzing output from failed runs, we were able to get the Flash team's 512-processor job up and running on short notice to help them meet a hard deadline," Antypas said.

The simulations are so demanding — the Flash team calls it "extreme computing" — that they monopolize powerful computers of the U.S. Department of Energy during the allocated time. To ensure that these computers are used to their maximum potential, the Flash team stands on alert to rapidly correct any glitches that may arise.

"We have it set up so that if something goes wrong, text messages are sent out instantaneously to everyone," said Flash Center Research Scientist Robert Fisher. "It's like being a doctor on call 24/7."

But the scientific payoff for logging these long, stressful hours is potentially huge. Astrophysicists value type Ia supernovas because they all seem to explode with approximately the same intensity. Calibrating these explosions according to their distance reveals how fast the universe has been expanding at various times during its long history.

In the late 1990s, supernova measurements revealed that the expansion of the universe is accelerating. Not knowing what force was working against gravity to cause this expansion, scientists began calling it "dark energy." The Flash Center simulations may help astrophysicists make better calibrations to adjust for the minor variation that they believe occurs from one supernova to the next.

"To make extremely precise statements about the nature of dark energy and cosmological expansion, you have to be able to understand the nature of that variation," Fisher said.

Telescopic images of the two supernovas closest to Earth seem

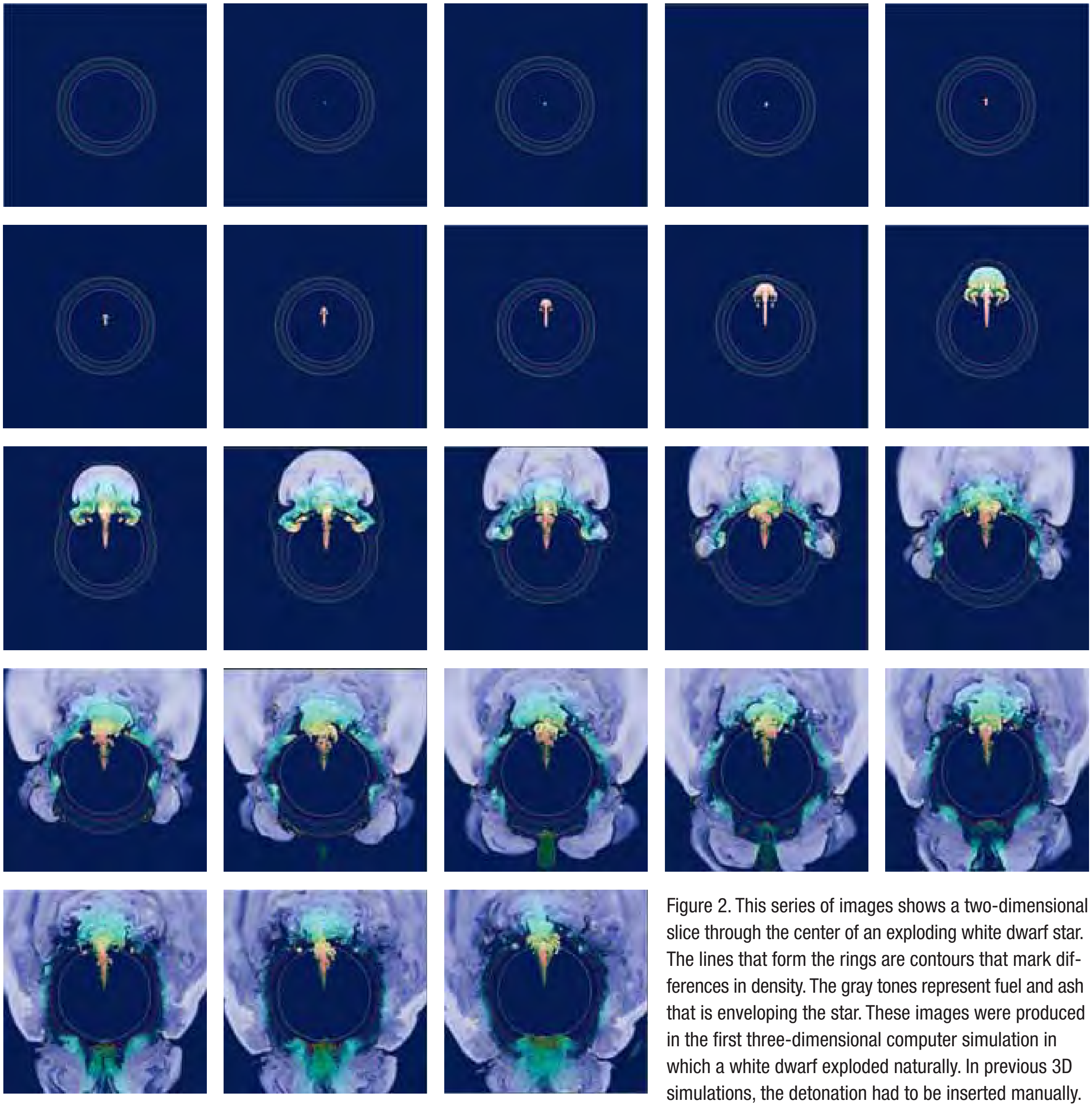


Figure 2. This series of images shows a two-dimensional slice through the center of an exploding white dwarf star. The lines that form the rings are contours that mark differences in density. The gray tones represent fuel and ash that is enveloping the star. These images were produced in the first three-dimensional computer simulation in which a white dwarf exploded naturally. In previous 3D simulations, the detonation had to be inserted manually.

match the Flash team's findings. The images of both supernovas show a sphere with a cap blown off the end.

"In our model, we have a rising bubble that pops out of the top. It's very suggestive," Jordan said.

Support for these simulations was provided by two separate DOE programs: the Advanced Simulation and Computing program, which has provided funding and computer time to the Flash Center for nearly a decade, and INCITE (Innovative and Novel Computation Impact on Theory and Experiment) of the Office of Science, which has provided computer time.

This article written by: Steve Koppes, University of Chicago; Ucilia Wang and John Hules, Berkeley Lab.

Science for Humanity

NERSC users share Nobel Peace Prize, among other honors

More than 20 NERSC users were contributing authors of the United Nations Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4), which was published in early 2007. Later in the year, the IPCC — a group of more than 2000 scientists and policy experts — shared the 2007 Nobel Peace Prize with former Vice President Al Gore "for their efforts to build up and disseminate greater knowledge about man-made climate change, and to lay the foundations for the measures that are needed to counteract such change," according to the Nobel announcement. Supercomputers at NERSC and the National Center for Computational Sciences (NCCS) at Oak Ridge National Laboratory provided more than half of the simulation data for the joint Department of Energy and National Science Foundation data contribution to AR4.

"Access to DOE leadership-class, high-performance computing assets at NERSC and ORNL significantly improved model simulations," said atmospheric scientist Lawrence Buja of the National Center for Atmospheric Research (NCAR), an NSF center. "These computers made it possible to run more realistic physical processes at higher resolutions, with more ensemble members, and longer historical validation simulations. We simply couldn't have done this without the strong DOE/NSF interagency partnership."

At NERSC, climate runs for the IPCC began in the late 1990s with the Parallel Climate Model (PCM). Results from these runs were stored in the PCM database at NERSC, the first truly public database for distributing climate data. "It is fair to say that without the PCM runs, made largely at NERSC in the late 1990s through 2002, the U.S. modeling effort would have not been the major factor it is in the IPCC report," said Michael Wehner, who managed the PCM database at the time. Since 2002, running PCM and the newer Community Climate System Model (CCSM), the IPCC project has used nearly 10 million processor hours at NERSC. Warren Washington of NCAR is principal investigator of the climate change simulation project at NERSC, working with senior investigators Jerry Meehl and Lawrence Buja.

Other current NERSC users who contributed to the IPCC AR4 report include:

Krishna Achutarao, Lawrence Livermore National Laboratory

Natalia Andronova, University of Michigan

Julie Arblaster, National Center for Atmospheric Research and Bureau of Meteorology Research Centre, Australia

William Collins, Lawrence Berkeley National Laboratory

Curt Covey, Lawrence Livermore National Laboratory

Chris Forest, Massachusetts Institute of Technology

Inez Fung, University of California, Berkeley and Lawrence Berkeley National Laboratory

Nathan Gillett, University of East Anglia

Jonathan Gregory, University of Reading and Hadley Centre for Climate Prediction and Research

William Gutowski, Iowa State University

Aixue Hu, National Center for Atmospheric Research

Hugo Lambert, University of California, Berkeley

Eric Leuliette, University of Colorado, Boulder

Ruby Leung, Pacific Northwest National Laboratory and National Oceanic and Atmospheric Administration

Michael Mastrandrea, Stanford University

Carl Mears, Remote Sensing Systems

Surabi Menon, Lawrence Berkeley National Laboratory

Joyce Penner, University of Michigan

Thomas Phillips, Lawrence Livermore National Laboratory

David Randall, Colorado State University

Haiyan Teng, National Center for Atmospheric Research

Minghuai Wang, University of Michigan

Li Xu, University of Michigan

A number of other awards and honors were bestowed on NERSC users in 2007:

Member of the National Academy of Sciences

American Chemical Society Ahmed Zewail Award in Ultrafast Science and Technology

Graham Fleming, University of California, Berkeley and Lawrence Berkeley National Laboratory

Fellow of the American Academy of Arts & Sciences

Saul Perlmutter, University of California, Berkeley and Lawrence Berkeley National Laboratory

Gruber Cosmology Prize

Saul Perlmutter, Greg Aldering, Alex Kim, and Peter Nugent (Supernova Cosmology Project), Lawrence Berkeley National Laboratory

Fellows of the American Association for the Advancement of Science

James Chelikowsky, University of Texas, Austin
Peter Cummings, Vanderbilt University
Fritz Prinz, Stanford University

Fellows of the American Physical Society

Michael Borland, Argonne National Laboratory
Giorgio Gratta, Stanford University
Stephen Gray, Argonne National Laboratory
Edward Seidel, Louisiana State University

U.S. Department of Energy Ernest Orlando Lawrence Award

Paul Alivisatos, University of California, Berkeley and Lawrence Berkeley National Laboratory

IEEE Computer Society Sidney Fernbach Award

David Keyes, Columbia University

Welch Award in Chemistry

William Miller, University of California, Berkeley and Lawrence Berkeley National Laboratory

American Meteorological Society Charles Franklin Brooks Award

Warren Washington, National Center for Atmospheric Research

American Physical Society Nicholas Metropolis Award for Outstanding Doctoral Thesis Work in Computational Physics

Chengkun Huang, University of California, Los Angeles

The NERSC Center

Kathy Yelick Is Named NERSC Director

Kathy Yelick, a professor of computer science at the University of California at Berkeley and an internationally recognized expert in developing methods to advance the use of supercomputers, was named director of NERSC in October 2007 and assumed her new duties in January 2008.

Yelick has received a number of research and teaching awards and is the author or co-author of two books and more than 75 refereed technical papers. She earned her Ph.D. in computer science from MIT and has been a professor at UC Berkeley since 1991 with a joint research appointment at Berkeley Lab since 1996.

“We are truly delighted to have Kathy serve as the next director of NERSC, and only the fifth director since the center was established in 1974,” said Berkeley Lab Director Steven Chu. “Her experience and expertise in advancing the state of high performance computing make her the perfect choice to maintain NERSC’s leadership position among the world’s supercomputing centers.”

Yelick, who has been head of the Future Technologies Group at Berkeley Lab since 2005, succeeds Horst Simon as head of NERSC. Simon, who has led NERSC since 1996, will continue to serve as Berkeley Lab’s Associate Director for Computing Sciences and Director of the Computational Research Division.

“When Horst Simon announced that he wanted to relinquish the leadership of NERSC, we knew he would be a tough act to follow,” said Michael Strayer, head of DOE’s Office of Advanced Scientific Computing Research, which funds NERSC. “But with the selection of Kathy Yelick as the next director, I believe that NERSC will continue to build upon its success in advancing scientific discovery through computation. We are extremely happy to have her take on this role.”

In 2006, Yelick was named one of 16 “People to Watch in 2006” by the newsletter HPCwire. The editors noted that “Her multi-faceted research goal is to develop techniques for obtaining high performance on a wide range of computational platforms, all while easing the programming effort required to achieve high performance. Her current work has shown that global address space languages like UPC and Titanium offer serious opportunities in both productivity and performance, and that these languages can be ubiquitous on parallel machines without excessive investments in compiler technology."

In addition to high performance languages, Yelick has worked on parallel algorithms, numerical libraries, computer architecture, communication libraries, and I/O systems. Her work on numerical libraries includes self-tuning libraries which automatically adapt the code to machine properties. She is also a consumer of parallel systems, having worked directly with interdisciplinary teams on application scaling, and her own applications work includes parallelization of a computational fluid dynamics model for blood flow in the heart. She is involved in a National Research Council study investigating the impact of the multicore revolution across computing domains, and was a co-author of a Berkeley study on this subject known as the "Berkeley View."[1] Yelick speaks extensively on her research, with over 15 invited talks and keynote speeches over the past three years.

"After working on projects aimed at making HPC systems easier to use, I'm looking forward to helping NERSC's scientific users make the most effective use of our resources," Yelick said. "NERSC has a strong track record in providing critical computing support for a number of scientific breakthroughs and building on those successes makes this an exciting opportunity."

NERSC Gets High Marks in Operational Assessment Review

NERSC General Manager Bill Kramer

NERSC received praises from its first ever Operational Assessment Review by a committee of scientists, national research facility leaders, and managers at the DOE Office of Science.

The review, conducted on August 28, 2007, fulfilled a requirement by the Office of Management and Budget to evaluate national research facilities for capital planning purposes. NERSC, Oak Ridge National Laboratory, and the Molecular Science Computing Facility, part of the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory, underwent reviews in 2007. Argonne National Laboratory will be reviewed in 2008.

"The committee feels that we are doing a good job and that our operations are mature and efficient," said Bill Kramer, NERSC's General Manager. "We exceeded the metrics for the review."

The committee evaluated each supercomputing center's business plan, financial management, innovation, scientific achievements, and customer satisfaction. The inaugural review also serves as a baseline for future assessments.

Reviewers emphasized the importance for each center to seek feedback from its users. Each center conducts its own user survey; at NERSC the survey participation hovers around 10 percent of all authorized users. The survey yields a valuable assessment of the software, hardware, and service offerings at NERSC.

In fact, researchers gave NERSC high marks in the 2006 survey, in which the respondents gave an av-

[1] K. Asanovic et al., "The Landscape of Parallel Computing Research: A View from Berkeley," University of California at Berkeley Technical Report No. UCB/EECS-2006-183, http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.html.

erage score of 6.3 — on a scale of 1 to 7 with 7 being "very satisfied" — for the question about their overall satisfaction with using NERSC resources.

The hard work by NERSC staff paid off when the committee gave the center kudos for its services and management.

"I would like to send along my thanks and congratulations to you and your team for such a fine job that you did on the Operational Assessment," wrote Vince Dattoria, a review committee member from the Advanced Scientific Computing Research Program Office within the Office of Science, in an email to Kramer.

"The professionalism and organizational maturity was evident in the presentations and responses to the impromptu questions from the reviewers. This is especially noteworthy since this was a baseline review with no precedents to draw from for guidance. Well done," Dattoria added.

Franklin Passes Rigorous Acceptance Test

On November 1, 2007, NERSC and Cray Inc. announced the successful completion of the acceptance test of one of the world's largest supercomputers. The powerful Cray XT4 system contains nearly 20,000 processor cores and has a top processing speed of more than 100 teraflops.

This supercomputer is being used to advance a broad range of scientific research. Named "Franklin" in honor of the first internationally recognized American scientist, Benjamin Franklin, the Cray XT4 system enables researchers to tackle the most challenging problems in science by conducting more frequent and increasingly detailed simulations and analyses of massive sets of data.

"With Franklin, we are increasing the computational power available to our 2,900 NERSC users by a factor of six, providing them with access to one of the world's fastest supercomputers dedicated to open scientific research," said Michael Strayer, associate director of DOE's Office of Advanced Scientific Computing Research, which funds NERSC. "We have high expectations that NERSC's proven track record of scientific productivity will provide many new discoveries and understandings."

The highly scalable Cray XT4 system is capable of running applications across a wide range of scientific disciplines, including astrophysics, fusion, climate change prediction, combustion, energy and biology. Franklin will enable researchers at Berkeley Lab to address such problems as developing better models of the Earth's climate and using it to predict the impact of carbon dioxide emissions and global warming. The powerful system will also allow researchers to explore clean energy technologies and validate theories that attempt to uncover evidence that explains the origin of the universe.

"NERSC's new Cray XT4 system has demonstrated that it can deliver a high sustained performance on a demanding scientific workload in a rigorous production environment while at the same time permitting users to explore scaling to 20,000 cores," said Horst Simon, Berkeley Lab's Associate Laboratory Director for Computing Sciences. "NERSC is proud to accept one of the largest 'broad impact science' supercomputers in the world for its

Franklin was the first production XT4 to run the Cray Linux Environment (CLE), an ultra-lightweight version of the standard Linux operating system. CLE is now standard on the XT4.

demanding user community."

Franklin has a theoretical peak speed in excess of 100 teraflops (100 trillion floating point operations per second). The system contains 9,672 AMD dual-core Opteron 2.6 GHz processors with 39 terabytes of memory. In assessing proposed systems, the Cray XT4 scalable architecture promised to deliver high sustained performance, which is critical to NERSC's 24x7 operation to meet users' supercomputing demands.

"We are very excited to see one of the largest supercomputers in the world opened up to the expansive user community at NERSC," said Peter Ungaro, president and CEO of Cray. "The Cray XT4 system will provide the computational power to enable researchers who compute at NERSC to efficiently tackle some of the most important problems we face today. With high sustained performance, scalability and upgradeability to petaflops capacity as its key attributes, the Cray XT4 supercomputer will help enable major advances in a number of scientific fields now and in the future."

NERSC General Manager Bill Kramer said, "I want to sincerely thank all the staff — NERSC and Cray — who worked so hard to make Franklin the first production Cray Linux Environment system, and the world's largest Cray XT4."

During negotiations to procure the system, NERSC and Cray mapped out a plan to install the Cray Linux Environment (CLE) on each of Franklin's 9,672 nodes. As a result of this partnership, NERSC became the first center with a production XT4 running CLE, an ultra-lightweight version of the standard Linux operating system. CLE makes the system easier to use, allowing users to more easily port their scientific applications from other architectures. During extensive testing, about 300 different features and functions were tested and validated, making CLE more reliable with the same or better performance than previous XT operating systems. CLE is now standard on the XT4.

As part of an extensive testing program, a number of NERSC users were given early access to Franklin to ensure that the system could handle the most demanding scientific applications.

"I am extremely impressed with Franklin," said Robert Harkness, an astrophysicist at the San Diego Supercomputer Center working on a project aimed at precisely measuring the cosmological parameters that describe the shape, matter-energy contents, and expansion history of the universe. Running an application called ENZO, the project seeks to increase our understanding of the dark energy and dark matter thought to make up more than nine-tenths of the universe. "We have run the largest instances of ENZO ever, anywhere, and found that the performance and scaling on Franklin are both strong and better than other computer platforms we have used at other computing centers."

Another project, led by Julian Borrill of Berkeley Lab, leveraged Franklin's computing power to prepare for analyzing the massive amounts of data to be sent back to Earth by the Planck satellite, set for launch in 2008. A joint U.S.-European project, Planck will use 74 detectors to measure the cosmic microwave background, the residual radiation from the actual Big Bang. Last scattered some 400,000 years after the Big Bang, it provides the earliest possible image of the universe, including encoded signatures of the fundamental parameters of all matter.

"I am delighted to report that we have just successfully created a map of the entire Planck Full Focal Plane one-year simulation," Borrill said. "This is the first time that so many data samples — three terabytes of data in 50,000 files, representing all the information collected by Planck during one mission year — have been analyzed simultaneously, a primary goal of our group's early Franklin efforts." Running his code on 16,384 processor cores, Borrill was able to complete the run in just 45 minutes.

Franklin would not have passed the acceptance test without the hard work put in by the NERSC staff to troubleshoot hardware and software problems that are typical of getting a supercomputer to work properly, especially for demanding, large-scale scientific research.

Working closely with early users and using their feedback to resolve any issues was key to prepare Franklin for the test. "One of the biggest issues was to get Franklin's reliability up to the levels that NERSC users expect," said Jonathan Carter, head of the User Services Group at NERSC. "Part of this was handled by soliciting early user feedback and reporting problems to Cray quickly. Another part was keeping the composite metrics like job failure rates, SSP, and ESP high on the list of things Cray needed to pay attention to."

Helen He, a member of the User Services Group, added: "The early users got a lot of useful work done. Many were able to run high-concurrency jobs to tackle much larger problems and model resolutions that were impossible before."

Early Results Demonstrate Franklin's Capabilities

Two teams of researchers pushing the limits of climate and weather simulations achieved noteworthy results running their codes on Franklin. The climate scientists successfully ran experimental simulations with resolutions much higher than in widely used codes. And another team of researchers set a speed performance record for a U.S. weather model, running on 12,090 of Franklin's processors. The accomplishments of these two projects are described below.

Climate Models Produce Finest Details

A team of climate researchers who obtained early access to Franklin said the powerful system produced simulations that offered high-resolution details of oceanic and atmospheric phenomena, results that were difficult to obtain from other supercomputers before.

At the DOE's behest, scientists from the National Oceanic and Atmospheric Administration's Geophysical Fluid Dynamics Laboratory (GFDL) proposed a set of experiments using climate models with resolutions many times higher than those in the standard models, such as those used by the Intergovernmental Panel on Climate Change (IPCC). The high-resolution models offer not only a closer look at physical elements of the climate, such as tropical storms, but they also enable researchers to conduct a more in-depth analysis of climate change as higher-resolution phenomena in the ocean and atmosphere are resolved.

For years, scientists worldwide have relied on simulations with resolutions in the 100-kilometer range for studying forces that shape the oceans and the atmosphere. But the resolution is not high enough to model details such as ocean vortices and clouds, phenomena that are critical for understanding regional climate variations. Developing a climate model is a computationally intensive task, and getting enough time on powerful supercomputers has always been a challenge.

GFDL scientists, located in Princeton, New Jersey, had developed models capable of modeling the global atmosphere at resolutions down to 5 km, and the ocean at resolutions between 10 km and 20 km. They also have designed experiments which generated 1 to 4 TB of data for every year of simulation. NERSC provided GFDL with the computational resources for this challenge by setting aside over 800,000 CPU hours on Franklin.

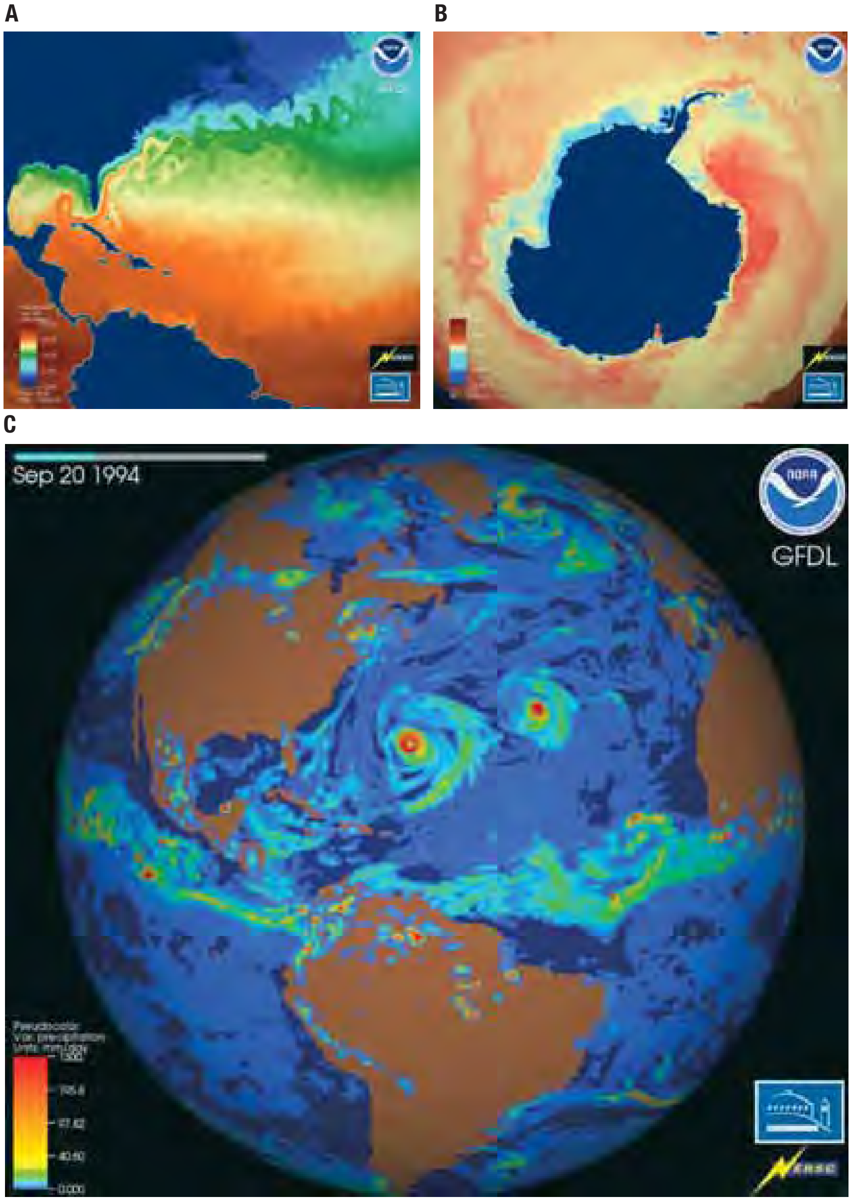
In addition to carrying out successful runs on Franklin, GFDL also received strong support from NERSC's Analytics Team in using animations to illustrate the results. Prabhat from the Analytics Team created a series of visual renderings of data that included sea surface temperatures, salinity, and clouds and precipitation in different parts of the world.

Prabhat

Richard Gerber

"We are able to increase our models' resolutions because of our access to the NERSC machine," said Venkatramani Balaji, head of the Modeling Services Group at GFDL. "One of the results is we can see category 4 or 5 hurricanes in a 20-km model, and they are what we would expect to see in the real world."

Senior software developer Christopher Kerr at GFDL and other members of Balaji's team were responsible for enabling the software



Visualizations of high-resolution climate datasets: (a) sea surface temperatures for the North Atlantic Gulf Stream; (b) salinity of the Southern Ocean; (c) hurricanes forming in the Atlantic Ocean. (Datasets provided by Chris Kerr, NOAA/GFDL; visualizations by Prabhat, NERSC/LBNL.)

infrastructure to perform these scientific experiments. Richard Gerber, a NERSC consultant, resolved system-related issues so that the experiments could be performed on Franklin.

"The results of the visualization collaboration have been outstanding," Balaji added.

The NERSC Analytics Team used the VisIt visualization and analysis package to create images and movies for climate scientists at GFDL. The team accelerated time to discovery by developing extensions to the software that eliminated costly data format conversion barriers, namely extra computation, extra data storage, and more manual processing steps. As a result, it was possible to load the simulation output generated on Franklin directly into VisIt for visual data exploration and analysis.

Developing the visualizations was a collaborative effort. The Analytics Team had extensive interactions with GFDL staff and scientists in order to create visualizations that focus on the most interesting and significant phenomena (e.g., the formation of tropical storms and ocean eddies). The visualizations present the simulation data in an accessible format using conventions familiar to the climate modeling community. The Analytics Team has also installed VisIt at GFDL to allow scientists to use the new visualization capability on a day-to-day basis, creating large-scale visualization of climate data that were impossible with conventional visualization tools.

The climate modeling project, called the Coupled High-Resolution Modeling of the Earth System (CHiMES), began as a collaboration between NOAA/GFDL and DOE. The research uses comprehensive Earth system models and historical data to examine how climate has changed over time and what external forces will likely influence the climate in the future. The models are based on the Flexible Modeling System (FMS) developed by GFDL. FMS is a powerful computational infrastructure for constructing coupled climate models on high-end scalable computer architecture.

The CHiMES project seeks to understand how the overall climate responds to high-resolution phenomena such as ocean eddies, as well as how fine-scale events such as tropical storms respond to climate change. To answer these questions, the project has been divided into two parts: one is to study the climate's predictability over decades or longer using high-resolution coupled models; the second part is to study the correlations between tropical storms and climate change, a hot topic in the research world. Work by GFDL researchers on this subject appeared in over 35 publications in scientific peer-reviewed journals last year.

For the hurricane research, CHiMES uses an atmospheric model based on the cubed sphere grid developed by lead scientist S. J. Lin at GFDL. This projection of a grid over the surface of the Earth is a more scalable basis than latitudes and longitudes for solving the equations of computational fluid dynamics on a sphere. The researchers have done test runs using the cubed sphere and found that the highly scalable methodology would enable them to carry out simulations with a 5-km resolution.

"We can go a lot further on this model," Balaji said. "If the coupled model simulations done at NERSC represent today's leading edge, this model is already showing what will be possible when the next generation of hardware becomes available."

Weather Forecast Model Sets Speed Record

A team of researchers set a performance record by running the Weather Research and Forecast (WRF) code on 12,090 of Franklin's processors at a speed of 8.8 teraflop/s — the fastest performance of a weather or climate-related application on a U.S. supercomputer. WRF is widely used for continuous weather forecasting by government, military, and commercial forecasters as well as for highly detailed weather and climate research in hundreds of universities and institutions worldwide.

Scientists from the National Center for Atmospheric Research (NCAR), the San Diego Supercomputer Center (SDSC) at UC San Diego, Lawrence Livermore National Laboratory (LLNL), and IBM Watson Research Center made up the research team that carried out the weather simulations on several high performance computers, setting national records not only in performance but also in size and fidelity of computer weather simulations.

The team's efforts open the way for simulations of greatly enhanced resolution and size, which will serve as a key benchmark for improving both operational forecasts and basic understanding of weather and climate prediction.

The scientific value of the research goes hand-in-hand with the

A new speed record for weather and climate codes — 8.8 teraflop/s — was set when the Weather Research and Forecast (WRF) model was run on 12,090 processors of Franklin. WRF is widely used for routine weather forecasting as well as research.

computational achievements. The "non-hydrostatic" WRF weather code is designed for greater realism by including more of the physics of weather and capturing much finer detail than simpler models traditionally used for global scale weather prediction. Running this realistic model using an unprecedented number of computer processors and simulation size enabled researchers to capture key features of the atmosphere never before represented in simulations covering such a large part of the Earth's atmosphere. This is an important step towards understanding weather predictability at high resolution.

"The scientific challenge we're addressing is the question in numerical weather prediction of how to take advantage of coming petascale computing power," said weather scientist Josh Hacker of NCAR. "There are surprisingly complex questions about how to harness the higher resolution offered by petascale systems to best improve the final quality of weather predictions." Petascale computing refers to next generation supercomputers able to compute at a petaflop/s (10^{15} calculations per second), equivalent to around 200,000 typical laptops.

The team also set a record for parallelism, running on 15,360 processors of the 103 peak teraflop/s IBM Blue Gene/L supercomputer at Brookhaven National Laboratory, jointly operated by Brookhaven and Stony Brook University.

"We ran this important weather model at unprecedented computational scale," added Hacker. "By collaborating with SDSC computer scientists to introduce efficiencies into the code, we were able to scale the model to run in parallel on more than 15,000 processors, which hasn't been done with this size problem before, achieving a sustained 3.4 teraflop/s."

Added John Michalakes, lead architect of the WRF code, "To solve a problem of this size, we also had to work through issues of parallel input and output of the enormous amount of data required to produce a scientifically meaningful result. The input data to initialize the run was more than 200 gigabytes, and the code generates 40 gigabytes each time it writes output data."

With this power the researchers were able to create "virtual weather" on a detailed 5-kilometer horizontal grid covering one hemisphere of the globe, with 100 vertical levels, for a total of some two billion cells — 32 times larger and requiring 80 times more computational power than previous simulation models using the WRF code.

"The calculation, which is limited by memory bandwidth and interprocessor communication, is representative of many other scientific computations," said Allan Snavely, director of the Performance Modeling and Characterization (PMaC) lab at SDSC, whose group helped tune the model to run at these unprecedented scales. "This means that what we learn in these large simulations will not only improve weather forecasts, but help a number of other applications as they enter the petascale realm."

The work was presented at SC07, the international conference for high performance computing, networking, storage, and analysis, where it was a finalist in the prestigious Gordon Bell Prize competition in high performance computing.

In preparing for the groundbreaking runs on the Stony Brook-Brookhaven and NERSC systems, the extensive problem solving required to achieve these results was made possible by running the WRF code on the Blue Gene system at DOE's Lawrence Livermore National Laboratory, the fastest supercomputer on the Top500 list, and the large Blue Gene system at the IBM Watson Research Center.

Tuning and testing were also carried out at the National Center for Computational Sciences at Oak Ridge National laboratory and on SDSC's Blue Gene system, a resource in the National Science Foundation-supported TeraGrid, an open scientific discovery infrastructure combining leadership class resources at nine partner sites. In these ongoing collaborations, the team anticipates further record-setting results.

Team members include John Michalakes, Josh Hacker, and Rich Loft of NCAR; Michael McCracken, Allan Snavely, and Nick Wright of SDSC; Tom Spelce and Brent Gorda of Lawrence Livermore; and Robert Walkup of IBM.

Story courtesy of SDSC.

Large Scale Reimbursement Program Improves Codes' Performance

Before Franklin's acceptance testing was completed, NERSC's large scale reimbursement program provided around nine million computing hours on Seaborg, NERSC's IBM SP RS/6000 system, to 21 projects that took advantage of the opportunity to scale their runs and improve their codes in preparation for using Franklin when it entered production.

The incentive attracted projects from a variety of scientific disciplines, including astrophysics, life sciences, fusion, chemistry, and climate research. Scientists said the program enabled them to pinpoint and resolve issues before running jobs on Franklin.

"We have participated in the scaling reimbursement program in order to tackle the problem of testing the predictive power of new empirical force fields for biomolecular simulation," said Nicolas Lux Fawzi, a UC Berkeley scientist on a research team led by Teresa Head-Gordon at Berkeley Lab. "Our system of interest is the Aβ peptide and various subpeptides which are associated in the formation of amyloid plaques in Alzheimer's disease. We have used the CPU time in the program to demonstrate that we can run large parallel simulations on 1024 processors using the replica exchange technique to generate the complete equilibrated ensemble for our system at a range of temperatures.

"We have very much enjoyed the chance to work with NERSC as part of the scaling program," Fawzi added. "We've received some excellent input from the NERSC staff on how to evaluate and improve the scaling of the code."

The reimbursement program was open to all NERSC users and required researchers to run 1024- to 1500-processor jobs on Seaborg, depending on whether they have participated in the program in the past. Each project could get a maximum reimbursement of 500,000 hours. Scientists also had to use the Integrated Performance Monitoring (IPM) software to gather performance information about each run.

"The quantum Monte Carlo methods developed in the Lester group are naturally amenable to parallel computing," said Brian Austin, a researcher in a group led by William A. Lester, Jr., a UC Berkeley chemistry professor and Berkeley Lab researcher. "Historically, our production jobs have run on several hundred processors with near perfect parallel efficiency. The advent of near-petascale computers such as Franklin will bring jobs with thousands of processors into the norm. In this regime, subtle changes to our mode of parallel communication have dramatic effects that were unnoticeable at previous scales: communication time increased from 2 percent to almost 50 percent as the number of processors increased from 512 to 2048. The reimbursement program has been essential to our exploration and resolution of these issues."

Don Lamb, a University of Chicago scientist who leads a supernovae research project (see page 28, also was an active participant in the reimbursement program. Lamb's research team members said the IPM was not easy to use initially, but they overcame those issues with the support of the NERSC staff.

"We had to pay more attention to exit codes issued by FLASH than we had previously, since non-zero exit codes force IPM to throw away all output. But the NERSC staff was helpful and understanding of IPM issues, never letting missing IPM output interfere with reimbursement. Where available, the IPM output supplied interesting profiling information," said Carlo Graziani, a researcher on Lamb's team.

Other scientists whose projects were among the top ten recipients of reimbursed hours were George Vahala (fusion plasmas), Doug Toussaint (quantum chromodynamics), Stephen Gray (nanoscale electrodynamics), Cameron Geddes (accelerator physics), Wei-li Lee (fusion plasmas), and Paola Cessi (climate research).

Seaborg Is Retired after Seven Years of Service

With Franklin up and running, Seaborg, the IBM supercomputer that has tackled some of the most challenging problems in astrophysics, climate research, fusion energy, chemistry and other scientific areas, was retired in January 2008 after seven years of serving NERSC users.

Since it was opened for production use in August 2001, Seaborg has provided a little over 250 million CPU hours to some 3,000 scientific users. Included in these numbers are 26.5 million CPU hours for 22 projects from the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program, which was created by the DOE Office of Science to support large-scale, high-impact projects.

COMPUTING ON SEABORG

	Total CPU Hours[2]	CPU Hours Used by Scientists	CPU Hours Used by INCITE Projects	No. of INCITE Projects	No. of Total Active Users[3]	No. of Active Science Users
2007[1]	32,652,500	32,537,200	4,283,200	7	1,048	980
2006	51,904,400	51,554,100	3,992,500	3	1,991	1,868
2005	50,180,000	49,809,100	7,281,800	3	1,903	1,749
2004	56,821,900	56,086,500	6,670,800	3	1,458	1,354
2003	39,913,300	38,088,300	2,502,400	3	1,224	1,115
2002	22,649,400	22,279,200	1,785,700	3	897	831
Total	256,696,400	252,787,500	26,516,300	22	3,165	2,996

[1] Data reflect usage through mid-September, 2007

[2] Total CPU hours includes time used by researchers, NERSC staff and HPC vendors.

[3] Active users include researchers, NERSC staff and HPC vendors.

The number of scientific papers resulting from computations on Seaborg is estimated to be over 7000.

At the time of its purchase, the IBM SP RS/6000 represented the largest single procurement in Berkeley Lab's history, costing $33 million for the system and a five-year contract. Named after the Nobel prize winner and Berkeley Lab chemist Glenn Seaborg, the supercomputer was installed in January 2001 and was ranked No. 2 on the June 2001 TOP500 list, a semi-annual ranking of the world's most powerful supercomputers. Seaborg ended its career at No. 331 on the November 2007 list.

Seaborg underwent a significant upgrade in 2003, increasing the number of its processors from the initial 2,528 to 6,656. The upgrade, costing $30 million, readied the supercomputer for large-scale projects from the SciDAC and INCITE programs.

"From the day it went online, Seaborg has been the scientific workhorse for many of the most important computational science projects undertaken within the Office of Science," said Michael Strayer, Associate Director of Advanced Scientific Computing Research within the Office of Science. "Now the system is ready to retire at a ripe old age of

In seven years of service, Seaborg provided over 250 million CPU hours to 3,000 scientific users. Its TOP500 ranking went from No. 2 in June 2001 to No. 331 in November 2007.

seven, many researchers will remember Seaborg with appreciation."

Seaborg was shut down for the last time on January 11, 2008, and was disassembled and removed for recycling within a week.

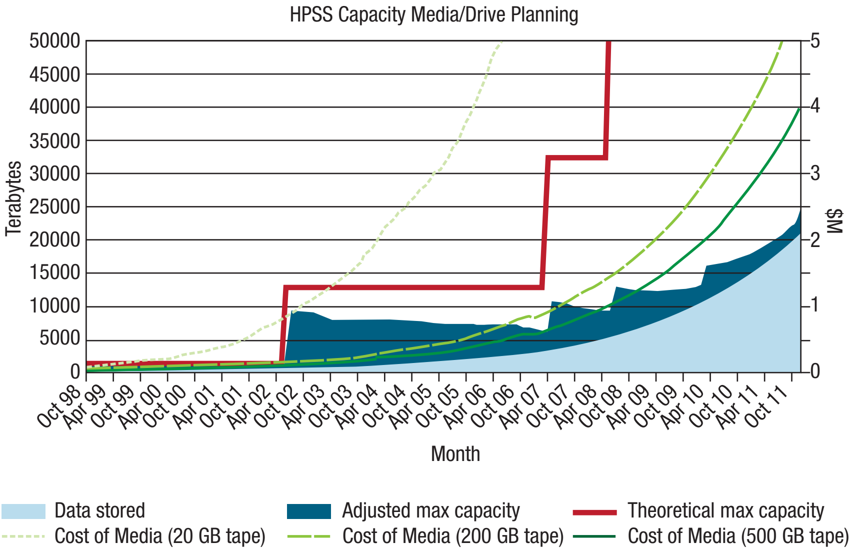
NERSC's Mass Storage Stays Ahead of the Curve

Advances in storage technology may not be as dramatic as advances in high performance computers — there is no TOP500 list for data centers — but they are just as essential to the success of a scientific computing center. NERSC's Mass Storage Group stays ahead of the curve by focusing on reliability, scalability, availability, performance, and security:

- *Reliability* is the most important goal, and NERSC has a track record of successful data preservation going back decades.
- *Scalability* involves careful planning and continuous upgrading of storage systems to anticipate the output of high-end computers and data-intensive experiments. NERSC's cumulative data storage has grown exponentially and crossed into the petascale in 2005.
- *Availability* and *performance* are crucial metrics for user productivity. NERSC's mass storage systems achieved an overall availability rate of 98.4% for FY 2007. Single-client I/O can top 300 MB/sec, and aggregate I/O reaches 750 MB/sec.
- *Security* of the data archive is maintained by NERSC's access restrictions as well as ongoing monitoring, logging, and auditing.

Two things that make storage at NERSC unique are applied storage research and operational efficiency.

NERSC's Mass Storage Group actively researches storage reliability and performance. They collect and analyze a variety of metrics, such as hardware and software failures and tape drive utilization, and apply the lessons learned to improve the production storage systems in ways that directly affect end-user experience. In addition, Bill Kramer, Jason Hick, and Akbar Mokhtarani are collaborators in the SciDAC Petascale Data Storage Institute (PDSI). NERSC's focus in PDSI is on gathering and reporting data on storage and file system reliability, and working on



This planning tool demonstrates actual and future storage growth, cost, and planned technology upgrades. The light blue curve represents data retained in the storage system; the red line represents theoretical maximum capacity of the storage system; the dark blue curve represents theoretical capacity adjusted by the media density of tapes already in the storage system. The green curves represent the cost of placing all the current data, represented by the light blue curve, on a given capacity of media; for example, it would cost about $1 million to store 5000 TB of data in April 2008 entirely on T10000A media (500 GB tape). The green curves demonstrate that continual upgrades in tape media capacity are essential to keeping costs of the storage system under control in the face of continual increased storage demand. Another important consideration is the number of tape silos required to retain a given amount of data. Introducing new silos into an existing storage system is non-trivial. The primary indicator of needing a new silo is the number of tape media slots required. As the adjusted max capacity, the dark blue curve, approaches the amount of data stored, the light blue curve, the more likely a new silo is required to handle storage growth with existing numbers and types of media in the system.

Storage Technology and Capacity Upgrades

Technology	Year Deployed	Data Rate	Cartridge Capacity	Total Storage Capacity (uncompressed)	Total Storage Capacity (compressed)
T9840A	1999	10 MB/s	20 GB	0.88 PB	1.3 PB
T9940A	2002	30 MB/s	60 GB	2.60 PB	3.9 PB
T9940B	2003	30 MB/s	200 GB	8.80 PB	13.2 PB
T10000A	2007	120 MB/s	500 GB	22.00 PB	33.0 PB

I/O benchmarking and characterization of selected petascale applications. And as a founding member of the High Performance Storage System (HPSS) collaboration, NERSC continues to contribute to ongoing HPSS development.

The operational efficiency of mass storage at NERSC is evident from the perspective of the past decade. Total data grew from a few terabytes in 1998 to several petabytes in 2007 using multiple storage technologies. Single transfer bandwidth improved from a peak of 1 MB/sec in 1998 to nearly 500 MB/sec in 2007. This growth was achieved using the same physical footprint in the machine room and with an essentially flat storage budget. This high efficiency requires careful planning and nearly continuous phased technology upgrades.

The major technology upgrade for 2007 was the installation of new Titanium 10000A tape drives, which provide two and a half times more capacity and four times the performance of the previous tape drives. Each cartridge can hold 500 GB of uncompressed data or 750 GB of compressed data. As a result, NERSC's 44,000-tape cartridge capacity is now 22 PB uncompressed and 33 PB with compression. (The table above shows storage capacity improvements over the past decade.) This year NERSC also implemented a 25% bandwidth improvement in the disk array.

Improving Access for Users

Providing users with easy access to large-scale data, computing, and storage resources is one of NERSC's key goals. Accomplishments in 2007 included the establishment of a high-bandwidth network connection between the KamLAND experiment in Japan and NERSC's PDSF and HPSS systems, and making most of NERSC's resources available on the Open Science Grid.

Overcoming Obstacles to Connect with KamLAND

KamLAND (Kamioka Liquid Scintillator Anti-Neutrino Detector) is a 1000-ton detector located in a former zinc mine near Kamioka in western Japan. With almost 1400 citations, the first report on neutrino oscillations from KamLAND is now one of the top-cited publications in nuclear and particle physics.[2] This study ruled out all but one solution to the "solar neutrino problem," while advancing the art of neutrino detection from surveys to precision measurements. KamLAND is also the only experiment in the world that can detect geologically produced neutrinos, which provide valuable information about heat generated by radioactive decay in the Earth's interior.

The experiment is now entering a new phase: the real-time detection of ^{7}Be solar neutrinos, which are emitted in the solar fusion processes. "The ^{7}Be neutrino flux is currently the least constrained quantity in the Standard Solar Model, the model that describes the processes in the Sun," said Stuart Freedman, PI of the "Data Analysis and Simulations for KamLAND" project at NERSC.

"A 5% ^{7}Be neutrino measurement is essential in order to better understand the fusion processes in the Sun and to see if the Sun is in a steady state," Freedman explained. "The neutrinos emerging from the Sun were produced just seconds earlier in the Sun's interior, while the photons that we see reflect the state of the Sun around 40,000 years ago. The eventual comparison of the neutrino-inferred luminosity to the presently visible photon luminosity will allow us to determine

[2] K. Eguchi et al. (KamLAND Collaboration), "First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance," Phys. Rev. Lett. **90,** 021802 (2003).

KamLAND is the largest low-energy anti-neutrino detector ever built.

Damian Hazen

Harvard Holmes

Wayne Hurlbert

how steady the energy production mechanisms in the Sun are."

The KamLAND project is the second-biggest user of storage resources at NERSC, with 420 TB of data accumulated over the past five years, and is one of the largest users of NERSC's PDSF cluster. NERSC is the sole U.S. site for computing in the KamLAND collaboration. The KamLAND detector is currently recording experimental data at a rate of 250 GB per day, 365 days per year (~80 TB a year), but by mid-2008 that will increase to 400 GB per day to accommodate the ^{7}Be neutrino measurements.

Until recently, data transfer from KamLAND to NERSC was accomplished by shipping tapes — hardly the most efficient method, but the only one possible at the time, given KamLAND's limited network bandwidth. But when the experimental site was upgraded from a 1 Mbps to a 1 Gbps network connection, Damian Hazen, Harvard Holmes, and Wayne Hurlbert of the NERSC Mass Storage Group undertook the formidable task of overcoming the remaining barriers: time zone challenges, language barriers (the network providers only speak Japanese), security barriers, and performance tuning. The result is that regular, reliable, and efficient data transfers are now occurring between Kamioka and NERSC.

"The high bandwidth network connection between KamLAND and NERSC has allowed us to streamline a number of data monitoring tasks and has been very beneficial for remote access to the experiment," Freedman said. "We are currently also using this connection to copy all the experimental data to the mass storage system at

Shreyas Cholia

Jeff Porter

NERSC. We can achieve a maximum performance of about 8 MB/s, which is more than adequate."

Syncing Up with the Open Science Grid

More NERSC users can now launch and manage their work at multiple computing sites by going through the Open Science Grid (OSG). NERSC's Bassi, Jacquard, DaVinci, PDSF, and HPSS systems are all available on the OSG, and Franklin will be joining them soon.

The Open Science Grid, funded by the DOE Office of Science and the National Science Foundation, is a distributed computing infrastructure for scientific research. The OSG software stack provides distributed mechanisms for moving data and submitting large parallel jobs to computing resources on the OSG grid at universities, national labs, and computing centers in North and South America, Europe, and Asia. Researchers from many fields, including astrophysics, bioinformatics, computer science, medical imaging, nanotechnology, and physics use the OSG infrastructure to advance their research.

The OSG will save valuable time and reduce headaches for scientists who carry out their research at several computing facilities. Instead of dealing with different authentication processes and software at each site, the scientists can go through the OSG to manage their computing jobs, data, and workflow across various sites using a uniform interface.

Making NERSC supercomputers available over the OSG has been a priority over the past year for Shreyas Cholia and Jeff Porter of the Open Software and Programming Group. Bill Kramer, NERSC's General Manager, chairs the OSG Council.

Connecting NERSC systems to the OSG requires bridging different pieces of local and distributed software and performing validation tests to make sure everything runs without any glitches. As new tools are added, NERSC solicits feedback from selected users who run the tools over a testbed. NERSC has set aside computing time specifically for projects carried out over the OSG, as part of an effort to attract new research and users.

Software Innovations for Science

NERSC's focus on user productivity sometimes leads to the development of new scientific software tools. Two recent examples are Sunfall, a collaborative visual analytics system that eliminated 90% of the human labor from a supernova search and follow-up workflow; and Integrated Performance Monitoring (IPM), which has been used to analyze application performance at NERSC for several years and has just been funded for deployment at all NSF supercomputing centers.

Sunfall: A Collaborative Visual Analytics System

Computational and experimental sciences are producing and collecting ever larger and more complex datasets, often in multi-institution projects. Managing, navigating, and gaining scientific insight from such data can be challenging, particularly when using software tools that were not specifically designed for collaborative scientific applications. To address this problem for observational astrophysics, Cecilia Aragon of the NERSC Analytics Team and Berkeley Lab Visualization Group led a group of computer scientists and astrophysicists to develop Sunfall, a collaborative visual analytics system for supernova discovery and data exploration.

Astrophysics lends itself to a visual analytics approach because much astronomical data, including images and spectra, is inherently visual. Sunfall (Supernova Factory

Cecilia Aragon

Assembly Line) was developed for the Nearby Supernova Factory (SNfactory),[3] an international astrophysics experiment and the largest data volume supernova search currently in operation. Sunfall is the first visual analytics system in production use at a major astrophysics project, and it won the Best Poster Award at the 2007 IEEE Symposium on Visual Analytics Science and Technology (VAST).[4]

Sunfall utilizes novel interactive visualization and analysis techniques to facilitate deeper scientific insight into complex, noisy, high-dimensional, high-volume, time-critical data. The system combines novel image processing algorithms, statistical analysis, and machine learning with highly interactive visual interfaces to enable collaborative, user-driven scientific exploration of supernova image and spectral data. The development of Sunfall led to a 90% labor savings in areas of the SNfactory supernova search and follow-up workflow; and project scientists now have new data exploration and analysis capabilities that had previously been too time-consuming to attempt.

The SNfactory is an international collaboration among Berkeley Lab, Yale University, and three research centers in France: Centre de Recherche Astrophysique de Lyon, Institut de Physique Nucléaire de Lyon, and Laboratoire de Physique Nucléaire et de Hautes Energies. SNfactory's mission is to create a large database of Type Ia supernovae, which are known for their extraordinary and uniform brightness and their role in breakthrough research showing that the universe's expansion is accelerating. The data will help researchers to understand the mysterious dark energy that propels the expansion.

The first goal in designing Sunfall was to tackle the growing amount of wide-field image data from the Palomar Observatory in San Diego. SNfactory receives 50 to 80 GB of image data each night, and its researchers must process and examine these data within 12 to 24 hours in order to get the most data from these rare stellar explosions. These supernovae only occur a few times per millennium in a typical galaxy and remain bright enough for detection only for a few weeks.

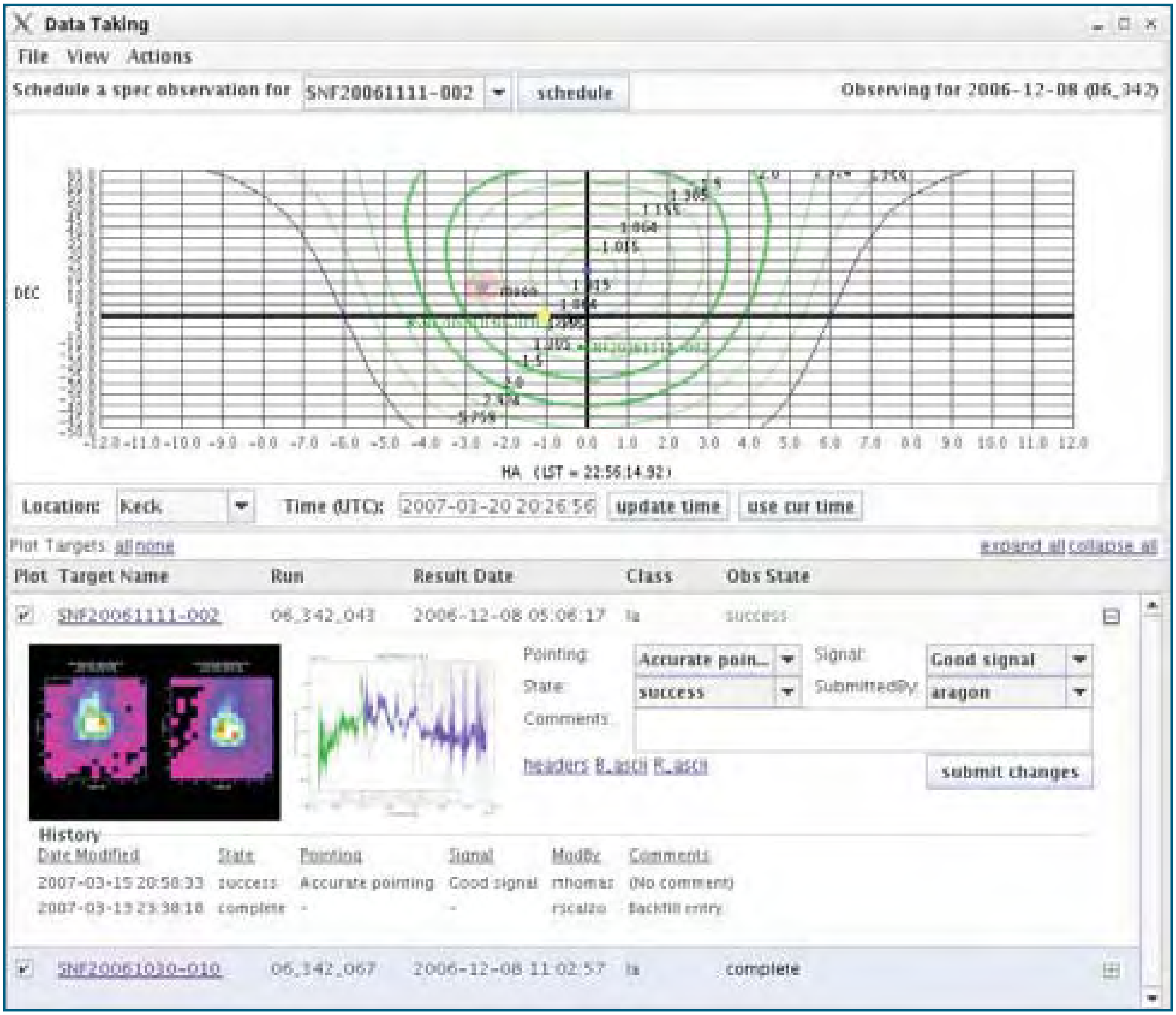
"To maximize the scientific return, this has to be a very accurate, efficient, and traceable process," said Greg Aldering, leader of the SNfactory. "The data come in seven days a week for a period of seven months, making the operations very dynamic, and there is no way to take a break to catch up. Before the Sunfall software was developed to integrate this process, it could be an overwhelming job simply to perform the necessary work, but in addition it also was difficult to keep track of whether what had been done was sufficient, much less optimal."

Processing and analyzing these data became a great challenge for the researchers, whose responsibilities require them to sift through vast amounts of image data to search for Type Ia supernovae, and then follow up with spectral and photometric observations of each supernova. Aragon and fellow Sunfall researchers created the system by modifying existing software and developing new algorithms.

Sunfall provides the software tools for extracting the Type Ia supernovae from the raw images, using statistical algorithms that reduced the number of false-positive supernovae candidates by a factor of 10. Instead of reviewing 1000 selected images each day, the researchers now only have to examine 100 images.

The software provides a visual display of three-dimensional astronomical data in an easy-to-read, two-dimensional format. Another visual display offers the signal strength and other information about each spectrum, making it easy for researchers to analyze the

[3] http://snfactory.lbl.gov/
[4] C. Aragon, S. Bailey, S. Poon, K. Runge, and R. Thomas, "Sunfall: A Collaborative Visual Analytics System for Astrophysics," IEEE Visual Analytics Science and Technology Conference, Sacramento, CA, Oct. 30–Nov. 1, 2007; http://vis.lbl.gov/Publications/2007/Sunfall_VAST07.pdf (abstract), http://vis.lbl.gov/Publications/2007/Sunfall_VAST07_poster.pdf (poster).



Schedule a spec observation for SNF20061111-002

Observing for 2006-12-08 (06_342)

Location: Keck

Time (UTC): 2007-03-20 20:26:56

Plot Targets: all none

Plot	Target Name	Run	Result Date	Class	Obs State
☑	SNF20061111-002	06_342_043	2006-12-08 05:06:17	Ia	success
☑	SNF20061030-010	06_342_067	2006-12-08 11:02:57	Ia	complete

Pointing: Accurate poin...

State: success

Comments:

Signal: Good signal

SubmittedBy: aragon

History

Date Modified	State	Pointing	Signal	ModBy	Comments
2007-03-15 20:58:33	success	Accurate pointing	Good signal	rthomas	(No comment)
2007-03-13 23:38:18	complete	-	-	rscalzo	Backfill entry

Sunfall's SNwarehouse Data Taking window. The observer can follow the targets on the sky visualization; take notes on the success or failure of each observation, telescope status, and weather conditions; and reschedule targets if necessary.

information. Sunfall also closely monitors and detects any problems that crop up while NERSC supercomputers process the data — there are visual displays of job queues, completion times, and other information.

A Sunfall package called Data Forklift automates data transfers among different types of systems, databases, and formats. Having a reliable and secure data transfer mechanism is critical given that SNfactory researchers are located in different countries and time zones. By using Sunfall's Supernova Warehouse (SNwarehouse) tool, scientists can easily access, modify, annotate, and schedule follow-up observations of the data.

"The new capabilities of SNwarehouse produced an immediate transformation, allowing us to shift our focus onto the science analysis of our follow-up spectroscopy," Aldering said.

Developing Sunfall was no easy task. Aragon and other members of the interdisciplinary team met frequently, often daily, to discuss various proposed designs and implementations, report progress, and ask for feedback on the developing system. Ideas for technical solutions often came during regular weekly science meetings in which SNfactory researchers discussed their work.

In addition to Aragon, other Sunfall project members included Stephen J. Bailey, formerly in Berkeley Lab's Physics Division, Sarah Poon and Karl Runge of the Space Sciences Lab at UC Berkeley, and Rollin Thomas, who worked for SNfactory and recently joined Berkeley Lab's Computational Cosmology Center.

"This project showed that an interdisciplinary team can have success solving high-data-volume science needs," Aragon said. "Now we have experience in solving a practical problem that can be applied in other scientific domains."

IPM to Be Deployed at NSF Supercomputer Centers

The National Science Foundation (NSF) in 2007 approved a proposal that will deploy a nimble performance evaluation tool, developed by NERSC's David Skinner, on all major NSF supercomputers.

The software, Integrated Performance Monitoring (IPM),[5] analyzes the performance of HPC applications and identifies load balance and communication problems that prevent them from running smoothly and achieving high performance. IPM is easy to deploy and use in systems with thousands or tens of thousands of processors, making it a good tool for petascale computing.

Skinner, leader of NERSC's Open Software and Programming Group, developed IPM in 2005. Since then, the software has won fans beyond NERSC, including the San Diego Supercomputer Center, the Center for Computation and Technology at Louisiana State University, the Swiss National Supercomputing Center, and the Army Research Laboratory.

IPM overcomes shortcomings exhibited by other performance analysis software, Skinner said. For example, IPM has low overhead and requires no source code modifications, making it easy for researchers to use. Its fixed memory footprint also ensures that running the software will not negatively impact the applications being profiled.

"An understandable application performance profile is something that all researchers using parallel computing resources should expect in situ via a simple flip of a switch. It should not require additional effort of changing their code," Skinner said.

The NSF proposal reviewed real-life cases of DOE and NSF supercomputer centers using various performance monitoring tools. The principal investigators for the project are San Diego Supercomputer Center staff members Allan Snavely and Nick Wright and NERSC Director Kathy Yelick.

David Skinner

"Some means of doing performance analyses are quite invasive and disturb the application one is trying to study; others are more lightweight but don't provide adequate information to researchers to improve their codes. Some require all users of a system to actively participate in the profiling activities; others are more passive, operating in the background. Some scale to thousands of tasks and some do not," said Skinner.

The comparisons enabled the researchers to convince NSF to deploy IPM in all of its supercomputer centers. NSF has awarded $1.58 million for the project, which is scheduled to take place over three years.

Part of the project will focus on expanding IPM's capabilities, such as broadening the scope of what is profiled and improving data analysis. The software is available under an open source software license and can run on major supercomputer architectures today: IBM, Linux clusters, Altix, Cray X1, Cray XT4, NEC SX6, and the Earth Simulator.

International Leadership and Partnerships

Every year, visitors from scientific computing centers around the world visit NERSC to exchange ideas and see firsthand how NERSC

[5] 5 http://ipm-hpc.sourceforge.net/

Horst Simon (left), head of Computing Sciences at Berkeley Lab, gave a tour of the NERSC facility to a delegation from RIKEN, Japan's premier science and technology research institution, in August 2007.

NERSC General Manager Bill Kramer (left) led a tour of NERSC's computer room for visitors from the Swiss National Supercomputing Centre, including COO Dominik Ulmer (center).

operates. For example, Thomas Lippert, director of the Central Institute for Applied Mathematics at Research Center Jülich in Germany and head of the John von Neumann Institute for Computing, spent a day with NERSC staff on March 6, 2007 to discuss performance and benchmarking of scientific computing systems. After touring NERSC's machine room at the end of his visit, he said, "NERSC is a model of how high-performance computing should be done."

Other visitors during the year came from Germany, Japan, Korea, Saudi Arabia, Sweden, and Switzerland. One of the most frequent topics of conversion was energy-efficient computer architectures and facilities. And one exchange of visits resulted in an ongoing staff exchange program.

In May 2007, NERSC hosted four visitors from CSCS (Swiss National Supercomputing Centre) for a series of discussions about systems and facilities. Howard Walter, head of NERSC's Systems Department, paid a return visit to CSCS in Manno, Switzerland in January 2008, sharing NERSC's expertise in designing and building energy-efficient computing facilities. And on that occasion CSCS and NERSC signed a memorandum of understanding for a staff exchange program between the two centers.

The agreement gives more formal structure to already existing ties between the two centers. Berkeley Lab Associate Director for Computing Sciences Horst Simon is a member of the CSCS advisory board. Both centers also share a common technological focus, having selected Cray XT supercomputers as their primary systems after thorough reviews of various systems.

"While many of us at NERSC are in frequent contact with our colleagues at other supercomputing centers in the U.S., we see this agreement as a means to broaden our outreach and perspectives," said NERSC Director Kathy Yelick. "Our informal discussions have already yielded valuable insights. With a more formalized structure, we expect these exchanges to be even more productive."

The two centers also play similar roles in their national research communities: CSCS is the largest supercomputing center in Switzerland and is managed by the Swiss

Federal Institute of Technology in Zurich. NERSC is the U.S. Department of Energy's flagship facility for computational science, serving 2,900 users at national laboratories and universities around the country.

"Not only do our two centers share organizational and operational similarities, but we both have the same primary goal of advancing the scientific research of our users," said CSCS COO Dominik Ulmer. "We believe each center has a lot of expertise to share, and we are looking forward to working together on new HPC technologies that will allow us to further enhance the support and services we offer our users."

Under the agreement, staff exchanges will be arranged based on specific projects of mutual interest. Each center will continue to pay the salary and expenses of staff participating in the exchanges. According to the agreement, the goal is "sharing and furthering the scientific and technical know-how of both institutions."

Envisioning the Exascale

Spatial (x*y*z*) resolution
1 km
10000
0.2°
22 km
1000
(AMR)
Ensembl
Simulation complexity
ESM + global cloud resolving model
Earth system model (15 components)
Climate model (5 components)
1.4°
160 km
0.2°
22 km
10000
1000
400
100
70
1000 yr/
20 min
2005
Terascale
2010
Petascale
2015
Exascale
Ensemble size

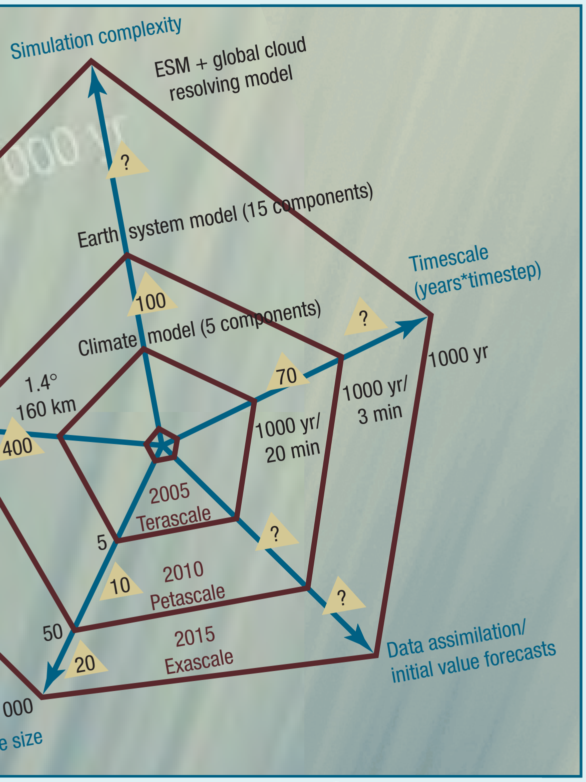

Simulation complexity
ESM + global cloud resolving model
?
Earth system model (15 components)
100
Climate model (5 components)
?
70
Timescale (years*timestep)
1000 yr
1000 yr/ 3 min
1000 yr/ 20 min
1.4°
160 km
400
2005 Terascale
5
?
10
2010 Petascale
?
50
2015 Exascale
20
Data assimilation/ initial value forecasts
000
e size

Figure 1. Investment of exascale and petascale computational resources in several aspects of a simulation: spatial resolution, simulation complexity, ensemble size, etc. Each red pentagon represents a balanced investment at a compute scale. (Image from E3 Report.)

In April, May, and June 2007, three town hall meetings were held at Lawrence Berkeley, Oak Ridge, and Argonne national laboratories to collect community input on the prospects of a proposed new DOE initiative entitled Simulation and Modeling at the Exascale for Energy and the Environment, or E3 for short. About 450 researchers from universities, national laboratories, and U.S. companies discussed the potential benefits of advanced computing at the exascale (10^{18} operations per second) on global challenge problems in the areas of energy, the environment, and basic science. The findings of the meetings were summarized in a document that quickly became known as the E3 Report.[1]

The E3 Report stated that exascale computer systems are expected to be technologically feasible within the next 15 years, but that they face significant challenges. One of the challenges that is receiving a great deal of attention throughout the HPC community is power efficiency. An exaflops system that requires less than 20 MW sustained power consumption is "perhaps achievable," according to the E3 findings, but based on straightforward scaling of existing technology, estimates are roughly an order of magnitude higher.[2] When the cost of running and cooling a supercomputer grows to exceed its procurement cost (which is already happening at major data centers), the economic viability of the project may come into question.

A task force of staff members from NERSC and Berkeley Lab's Computational Research Division held a series of internal meetings in 2007 to develop a strategic vision for NERSC's transition to the exascale.[3] Envisioning NERSC as the Keystone Facility for the DOE Office of Science, the task force addressed three broad topics: computing, power, and data.

The discussions about NERSC at the exascale were informed both by the E3 Report and by a series of meetings that had taken place in Berkeley over the previous two years, in which a multidisciplinary group of researchers (including Kathy Yelick, John Shalf, Parry Husbands, Bill Kramer, and Lenny Oliker) collaborated with faculty and students on campus to explore the implications of the recent switch to multicore processors throughout the computing industry. This collaborative exploration led to the publication of "The Land-

[1] H. D. Simon, T. Zacharia, R. Stevens, et al., "Modeling and Simulation at the Exascale for Energy and the Environment" ("E3 Report"), Department of Energy Technical Report (2007); http://www.sc.doe.gov/ascr/ProgramDocuments/TownHall.pdf.

[2] E3 Report, p. viii.

[3] Participants included Deborah Agarwal (CRD), Michael Banda, E. Wes Bethel, John Hules, William Kramer, Juan Meza (CRD), Leonid Oliker, John Shalf, Horst Simon, David Skinner, Francesca Verdier, Howard Walter, Michael Wehner (CRD), and Katherine Yelick.

scape of Parallel Computing Research: A View from Berkeley"[4] and to the establishment of UC Berkeley's Parallel Computing Laboratory (Par Lab),[5] which received major funding from Intel, Microsoft and California's UC Discovery program. The ParLab project involves Berkeley Lab faculty scientists Krste Asanovic, Jim Demmel, John Wawrzynek, Kathy Yelick, and the principal investigator of the project, David Patterson.

The "Berkeley View" report received widespread attention.[6] It asserted that an evolutionary approach to parallel hardware and software would face diminishing returns at 16 cores and beyond, when the increased difficulty of parallel programming would not be rewarded by a commensurate improvement in performance. "We concluded that sneaking up on the problem of parallelism via multicore solutions was likely to fail and we desperately need a new solution for parallel hardware and software," the authors stated.[7] Taking aim at thousands of processors per chip ("manycore"), the report proposed seven critical questions for parallel computing research and suggested directions in which the solutions might be found. One of the more provocative suggestions was that embedded computing — the small, low-power processors in consumer electronics products such as cell phones, PDAs, and MP3 players — and high performance computing would have more in common in the future than they did in the past.

The anticipation of a paradigm shift in high performance computing set the context for the discussions of NERSC's transition to the exascale. The path forward envisioned in those discussions is summarized below.

NERSC Computing

The need for an increase in computational resources is well documented in the E3 Report, the DOE Greenbook,[8] and the SCaLeS Report.[9] The scientific requirements go beyond the traditional Office of Science work that NERSC has supported for the past three decades, adding untouched areas of life science, energy resources (nuclear, biofuels, and renewable), energy efficiency, climate management, nanotechnology, and knowledge discovery. Simulations will grow in complexity, spatial resolution, timescales, ensemble sizes, and data assimilation (Figure 1). The computational needs are far beyond what can be supplied today by NERSC alone, NERSC combined with the other DOE centers, or even an augmented NERSC with other DOE centers.

NERSC and the Office of Science's Leadership Computing Facilities will be linked by ESnet in a fully integrated computational science environment. NERSC's role as the Keystone Facility in this environment will be to provide exceptional quality of service and

- high-impact computing (defined below)
- broad-impact computing for the diversity of DOE mission science
- efficient and transparent access to and management of simulation and experimental data
- integrated data analysis tools and platforms
- integrated support for SciDAC and other science community-developed tools
- outreach to new HPC user communities.

NERSC will continue to support both *high-impact* and *broad-impact* science workloads. High-impact work is defined as ultrascale workflows or applications that require 20–100% of the largest resources at any given time. Broad-impact work is science that runs at scale, with high throughput, using 1–20% of

[4] Krste Asanovíc, Rastislav Bodik, Bryan Catanzaro, Joseph Gebis, Parry Husbands, Kurt Keutzer, David Patterson, William Plishker, John Shalf, Samuel Williams, and Katherine Yelick, "The Landscape of Parallel Computing Research: A View from Berkeley," University of California at Berkeley Technical Report No. UCB/EECS-2006-183, http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.html.

[5] http://parlab.eecs.berkeley.edu/

[6] For example, John Shalf's talk "Overturning the Conventional Wisdom for the Multicore Era: Everything You Know Is Wrong" was voted one of the most popular at the 2007 International Supercomputing Conference in Germany.

[7] "Berkeley View," p. 3.

[8] S. C. Jardin, ed., "DOE Greenbook: Needs and Directions in High Performance Computing for the Office of Science. A Report from the NERSC User Group." PPPL-4090/LBNL-58927, June 2005; http://www.nersc.gov/news/greenbook/2005greenbook.pdf.

[9] David E. Keyes, Phillip Colella, Thom H. Dunning, Jr., and William D. Gropp, eds., *A Science-Based Case for Large-Scale Simulation ("The SCaLeS Report"),* Washington, D.C.: DOE Office of Science, Vol. 1, July 30, 2003; Vol. 2, September 19, 2004; http://www.pnl.gov/scales/.

Table 1: Characteristics of scientific discipline codes

	Multi-physics, multi-scale	Dense linear algebra (DLA)	Sparse linear algebra (SLA)	Spectral methods (FFTs) (SM-FFT)	N-body methods (N-Body) (PIC)	Structured grids (S-Grids)	Unstructured grids (U-Grids)	Data Intensive (Map Reduce)
Nanoscience	X	X	X	X	X	X		
Chemistry	X	X	X	X	X			
Fusion	X	X	X			X	X	X
Combustion	X		X			X	X	X
Astrophysics	X	X	X	X	X	X	X	X
Biology	X	X					X	X
Nuclear		X	X		X			X
System balance implications	General-purpose balanced system	High-speed CPU, high flop/s rate	High performance random access memory	High interconnect bisection bandwidth	High-speed CPU, fast random access memory	High performance unit and strided memory access	Irregular data and control flow, high performance random access memory	High storage and network bandwidth/ low latency

the resources at a given time. NERSC expects to support between 10 and 20 high-impact science projects and from 200 to 250 broad-impact projects. In addition to broad-impact and high-impact science, NERSC will be posed to support three to five INCITE-like “breakthrough science” projects a year. These projects receive preferred processing and services from NERSC in order to meet their milestones.

The system architecture will focus on the computational needs of the user base by using multiple HPC systems, with at least two large-scale systems in place at a time. While it is possible to create specialized system solutions for a single algorithmic approach, Table 1 shows it is not feasible to segregate system balance by discipline, since different workflow steps and/or approaches within a discipline require different system balance. Often these different algorithms exist in the same codes. Likewise, a system balanced for just one computational approach cannot fully serve a discipline area alone.

Since NERSC will support a diverse workload as indicated in Table 1, resources can be provisioned in one of two ways. First, provide general-purpose systems that are optimized to do well with the entire or a large segment of the workload. Second, provide a small number of specialized systems that are very efficient at particular algorithms and in aggregate support the entire diverse workload. NERSC is open to either approach or one that balances a portfolio of general and special systems that overall provides DOE with a very efficient facility for its entire problem space. In either case, a key requirement is easing the burden on scientists to move data between systems. In all likelihood, NERSC will have some large general-purpose systems and a few specialized systems for breakthrough science areas or specific algorithmic needs.

NERSC’s general large-scale systems will provide an appropriate relationship between sustained performance, usable memory, and usable disk space. Regardless of whether the general-purpose or the specialized system approach is taken, a key system architecture component that will make scientists succeed will be having a high performance, parallel, facility-wide file system tightly integrated with NERSC large-scale systems. Balanced system architectures also include:

- high performance local-area network
- wide-area network interfaces matching or exceeding ESnet backbone speeds
- archival storage with
  - large-scale near line data repository
  - online data cache
- data focused systems
  - community data services: “Google for Science” (described below, page 66)
  - visualization and analysis

- servers and specialized systems
    - Web services, NERSC Information Management System (NIM), etc.
- Infrastructure special-arrangement systems
    - PDSF, Planck cluster
- cyber security
- advanced concept systems.

NERSC will remain open to and will encourage deploying systems from multiple vendors, with major systems arriving at three-year intervals. At least two systems will be on the floor at any given time, one providing a stable platform while the next-generation system is brought into production.

NERSC Power

The electrical power demands of ultrascale computers threaten to limit the future growth of computational science. For decades, the notion of computer performance has been synonymous with raw speed as measured in flop/s. This isolated focus has led to supercomputers that consume egregious amounts of electrical power. Other performance metrics have been largely ignored, e.g., power efficiency, space efficiency, reliability, availability, and usability. As a consequence, the total cost of ownership of a supercomputer has increased extraordinarily. With the cost of power possibly exceeding the procurement costs of exaflop systems, the current approach to building supercomputers is not sustainable without dramatic increases in funds to operate the systems.

The new design constraint for processing elements is electrical power. While Moore's Law is alive and well, smaller transistors are no longer resulting in faster chips that consume less energy. Traditional methods for extracting more performance per processor have been well mined. The only way to improve performance now is to put more cores on a chip. In fact, it is now the number of cores per chip that is doubling every 18 months instead of clock frequency doubling as in the past.

Consequently, the path towards realizing exascale computing depends on riding a wave of exponentially increasing system concurrency. This is leading to reconsideration of interconnect design, memory balance, and I/O system design. The entire software infrastructure is built on assumptions that are no longer true. The shift to multicore and manycore processors will have dramatic consequences for the design of future HPC applications and algorithms.

To reach exascale computing cost-effectively, NERSC proposes to radically change the relationship between machines and applications by developing a tightly coupled hardware/software co-design process. We will directly engage the Office of Science applications community in this cooperative process of developing models to achieve an aggressive goal of *100 times the computational efficiency and 100 times the capability* of the mainstream HPC approach to hardware/software design. We propose to use global cloud system resolving models for climate change simulation as one of the key driver applications to develop the hardware/software co-design methodology. This hardware/software co-design process can dramatically accelerate the development cycle for exascale systems while decreasing the power requirements.

Reducing Waste in Computing

The low-power embedded computing market — including consumer electronics products such as cell phones, PDAs, and MP3 players — has been the driver for CPU innovation in recent years. The processors in these products are optimized for low power (to lengthen battery life), low cost, and high computational efficiency.

According to Mark Horowitz, Professor of Electrical Engineering and Computer Science at Stanford University and co-founder of Rambus Inc., "Years of research in low-power embedded computing have shown only one design technique to reduce power: reduce waste." The sources of waste in current HPC systems include wasted transistors (surface area), wasted computation (useless work, speculation, stalls), wasted bandwidth (data movement), and optimizing chip design for serial performance, which increases the complexity (and power waste) of the design.

Efficient designs must be specific to application and/or algorithm classes, as suggested by a study that examined the dual-core AMD processor on Cray XT3 and XT4 systems to assess the current state of system balance and to determine when to invest more resources to

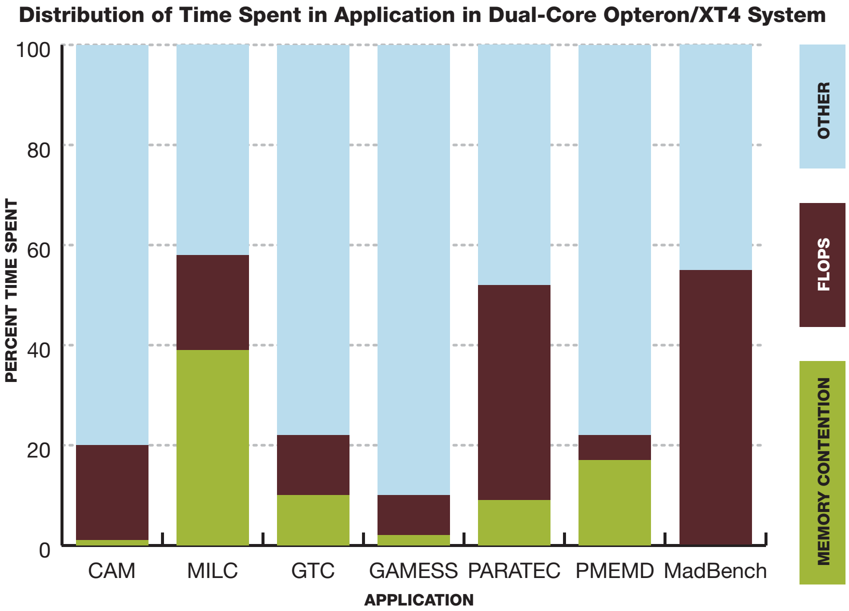


Figure 2. A breakdown of where time was spent in a subset of the NERSC SSP application codes suggests that different applications have different requirements for computational efficiency.

improve memory bandwidth.[10] The study used the NERSC SSP benchmark, which is a diverse array of full-scale applications that represent a significant fraction of the NERSC workload. A breakdown of time spent in various components of the codes (Figure 2) shows that surprisingly little time could be attributed to memory contention corresponding to basic memory bandwidth limitations. The largest fraction of time (the "other" category) is attributed to either latency stalls or integer/address arithmetic. Theoretically, these applications should all be memory-bandwidth bound, but instead the study shows that most are constrained by other microarchitectural bottlenecks in existing CPUs, and that different applications have different balance requirements.

A core designed to a specific set of application resource requirements can get 10 to 100 times better performance per watt, as shown by studies from Stanford University[11] and from Tensilica, Inc.[12] Figure 3 illustrates this potential by showing the area and performance differences between general purpose, embedded (used in BlueGene/P), and application-tailored cores. The figure shows how much area and power desktop processors waste because they are optimized for serial code. The DOE applications, because they are already highly parallel, are an excellent driver for understanding how processors can be designed to optimize for efficient parallel execution rather than serial execution.

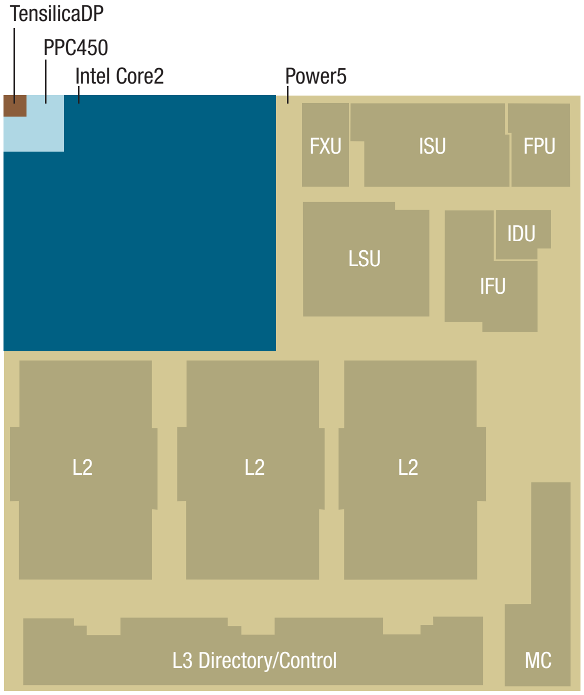
The "Berkeley View" report concludes that parallelism is an energy-efficient way to achieve performance. A system with many simple cores offers higher performance per unit area for parallel codes than a comparable design employing smaller numbers of complex cores. Lower complexity makes a chip more economical to design and produce, and smaller processing elements provide an economical way to improve defect tolerance by providing many redundant cores that can be turned off if there are defects.

Figure 3 shows that moving to a simpler core design results in modestly lower clock frequencies, but has enormous benefits in chip surface area and power consumption. Even if it is assumed that the simpler core will offer only one-third the computational efficiency of the more complex out-of-order cores, a manycore design (hundreds to thousands of simple cores) could still provide an order of magnitude more power efficiency for an equivalent sustained performance. As the figure illustrates, even with the smaller cores operating at one-third to one-tenth the efficiency of the largest chip, 100 times more cores can still be packed onto a chip and consume one-twentieth the power.

[10] J. Carter, H. He, J. Shalf, E. Strohmaier, H. Shan, and H. Wasserman, "The Performance Effect of Multi-Core on Scientific Applications," Cray User Group (CUG2007), Seattle, Washington, May 7–10, 2007.

[11] M. Horowitz, E. Alon, D. Patil, S. Naffziger, R. Kumar, and K. Bernstein, "Scaling, Power, and the Future of CMOS," IEEE International Electron Devices Meeting, December 2005.

[12] Chris Rowen, "Application-Specific Supercomputing: New Building Blocks Enable New Systems Efficiency," SIAM Conference on Computational Science and Engineering, February 19, 2007, Costa Mesa, California.



Power5 (server)
- 389 mm^2
- 120 W @ 1900 MHz

Intel Core2 sc (laptop)
- 130 mm^2
- 15 W @ 1000 MHz

PowerPC450 (BlueGene/P)
- 8 mm^2
- 3 W @ 850 MHz

Tensilica DP (cell phones)
- 0.8 mm^2
- 0.09 W @ 650 MHz

Figure 3. Relative size and power dissipation of different CPU core architectures. Simpler processor cores require far less surface area and power with only a modest drop in clock frequency. Even if measured by sustained performance on applications, the power efficiency and performance per unit area is significantly better when using the simpler cores.

Effective performance per watt is the critical metric.

This design approach raises the challenges of creating ultrascale parallel applications.

A Tightly Coupled Hardware/ Software Co-Design Process

If the HPC community emulated the embedded computing industry, we could potentially reduce not only power requirements but also design costs and time to market. A key limiting factor in the market-driven approach to HPC procurements is the length of the feedback loop on system designs. Due to the high design investment cost, the vendor must make compromises in the system design to accommodate a wide variety of applications. The application scientists cannot provide performance feedback to the vendor until hardware is released for testing and evaluation. This multi-year cycle is a source of significant inefficiencies for scientific productivity, because it can take years for each new iteration of hardware to become available for testing and evaluation by the application scientists. A hardware/software co-design approach could dramatically accelerate this process.

For years, NERSC has engaged in a cooperative effort with hardware designers, which we call Science-Driven System Architecture, which involves engaging application scientists in the early parts of the hardware design process for future-generation supercomputing systems.[13] This approach is consistent with the recommendations of the High-End Computing Revitalization Task Force (HECRTF)[14] and the National Research Council report *Getting Up to Speed: The Future of Supercomputing,*[15] both of which recommend partnerships with vendors in the early stages of the product development process.

NERSC proposes to focus this cooperative effort toward a new design paradigm: *application-driven HPC.* This approach involves identifying high-impact exascale scientific applications, tailoring the system architecture to the application resource requirements, and co-designing algorithms and software together with the semi-custom hardware.

This co-design process would be impossible using the typical multi-year hardware lead times for complex, serial-optimized chips. However, a typical embedded processor vendor may generate up

[13] H. D. Simon, W. T. Kramer, W. Saphir, J. Shalf, D. Bailey, L. Oliker, M. J. Banda, C. W. McCurdy, J. Hules, A. Canning, M. Day, P. Colella, D. Serafini, M. F. Wehner, and P. Nugent, "Science-Driven System Architecture: A New Process for Leadership Class Computing," Journal of the Earth Simulator, Volume 2, March 2005, pp. 2–10; Lawrence Berkeley National Laboratory technical report LBNL-56545; http://repositories.cdlib.org/lbnl/LBNL-56545.

[14] Federal Plan for High-End Computing: Report of the High-End Computing Revitalization Task Force (HECRTF), Washington, D.C.: National Coordination Office for Information Technology Research and Development, May 10, 2004.

[15] National Research Council Committee on the Future of Supercomputing, *Getting Up to Speed: The Future of Supercomputing,* S. L. Graham, M. Snir, and C. A. Patterson, eds. (Washington, DC: National Academies Press, 2004).

to 200 unique designs every year for simple, specialized chips. In order to keep up with the demand for semi-customized designs, leading embedded design houses such as IBM Microelectronics, Altera, and Tensilica have evolved sophisticated toolsets to accelerate the design process through semi-automated synthesis of custom processor designs. NERSC proposes to leverage the expertise of this technology sector by collaborating with Mark Horowitz of Stanford University and Rambus, Inc., and Chris Rowen of Tensilica, Inc.

NERSC's co-design process will utilize the Berkeley Research Accelerator for Multiple Processors (RAMP),[16] an FPGA emulation platform that makes the hardware configuration available for evaluation while the actual hardware is still on the drawing board. Making use of large field programmable gate arrays (FPGAs), RAMP looks like the real hardware to software developers, who can efficiently test their target application software on varying hardware configurations. The flexibility of RAMP allows rapid changes in the details of the hardware configuration (e.g., the number of processors, number of floating point units per processor, size and speed of caches, prefetching schemes, speed of memory, etc.). Since RAMP allows these explorations at speeds 1,000 times faster than conventional software-based cycle-accurate simulators, hardware/software co-design is now feasible.

The software side of the co-design process will be supported by auto-tuning tools for code generation that are being developed by the SciDAC Center for Scalable Application Development Software,[17] led by John Mellor-Crummey of Rice University.

An Ultrascale Application: Ultra-High-Resolution Climate Change Simulation

NERSC's hardware/software co-design methodology is broadly applicable and could be applied to a number of Office of Science disciplines that could effectively utilize ultrascale resources, resulting in different hardware configurations within a consistent framework. As a pilot project that could result in breakthroughs in both the domain science and computer science, NERSC proposes to use climate change simulation to illustrate this power-efficient computing approach, resulting in a synergy of reducing the carbon footprint needed to predict climate change with unprecedented accuracy.

A major source of errors in climate models is poor cloud simulation. The deep convective processes responsible for moisture transport from near-surface to higher altitudes are inadequately represented at current resolutions. Current-generation climate models can be extended to about a 20 km horizontal resolution in the atmospheric component without major reformulation, but at finer resolutions, the treatment of cumulus cloud processes breaks down. Fortunately another alternative presents itself at the 1 km scale, where cloud systems (but not individual clouds) can be resolved. Current technologies for regional modeling at this scale are extendable to global models, but the computational platform will need to achieve around 10 petaflop/s (Pflop/s) sustained.[18] A detailed extrapolation of the resource requirements of a current-generation atmospheric model showed that it is unlikely that multicore chip technology will achieve this goal in the next two decades within practical hardware or power budgets. An energy-efficient hardware architecture capable of achieving the aggressive requirements of the kilometer-scale model could employ 20 million much simpler cores using existing 90 nm technology.

Actual development of a 1 km cloud system resolving global model is a significant multi-year effort. The SciDAC project "Design and Testing of a Global Cloud Resolving Model"[19] is leading the way with a grid-cell spacing of approximately 3 km on a highly uniform icosahedral grid. This

[16] S. Wee, J. Casper, N. Njoroge, Y. Teslyar, D. Ge, C. Kozyrakis, and K. Olukotun, "A Practical FPGA-based Framework for Novel CMP Research," Proceedings of the 15th ACM SIGDA Intl. Symposium on Field Programmable Gate Arrays, Monterey, CA, February 2007. See also the RAMP homepage, http://ramp.eecs.berkeley.edu/.

[17] http://cscads.rice.edu/

[18] Michael Wehner, Leonid Oliker, and John Shalf, "Towards Ultra-High Resolution Models of Climate and Weather," International Journal of High Performance Computing Applications (in press).

[19] http://kiwi.atmos.colostate.edu/gcrm/

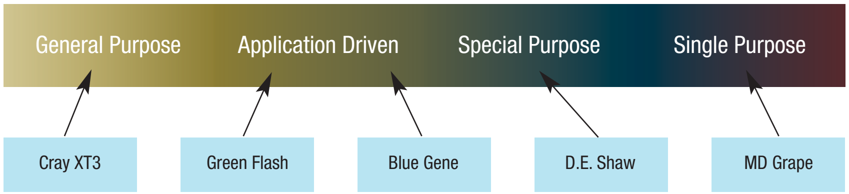


Figure 4. The customization continuum of computer architectures.

model would be capable of simulating the circulations associated with large convective clouds. Although the procedure is conceptually straightforward, there is little sense of urgency for a 1 km global model given the assumption that the computer technology is not arriving anytime soon. However, this model could run much sooner than expected on massively concurrent architectures composed of power-efficient embedded cores.

NERSC proposes a focused program to design a computing platform and the climate model in tandem. This research project has been named "Green Flash." The computer system would employ power-efficient cores specifically tailored to meet the requirements of this ultrascale climate code. The equations of motion (rather than the physics) dominate the requirements of an atmospheric model at 1 km resolution because the Courant stability condition requires smaller time steps. To be useful, the model must run at least 1,000 times faster than real time, calculating values for about 2 billion icosahedral points. At this rate, millennium-scale control runs could be completed in a year, and century-scale transient runs could be done in a month. The computational platform performance would need to reach around 10 Pflop/s sustained. This goal could be achieved with 20 million processors, modest vertical parallelization, a modest 0.5 gigaflop/s per processor, and 5 MB memory per processor.

An application-driven architecture does not necessitate a special-purpose machine, nor does it require exotic technology. As Figure 4 shows with several examples, there is a customization continuum from general-purpose to single-purpose computers, and indeed the Blue Gene line of systems was started with a very narrow application target in mind.

At the single-purpose, fully custom extreme is MD-Grape, a computer at RIKEN in Japan. MD-Grape was designed for molecular dynamics simulations (the name stands for "Molecular Dynamics — Greatly Reduced Array of Processor Elements") and has a custom ASIC chip design. It achieves 1 Pflop/s performance for its target application using 200 kilowatts of power, and cost $8.6M from concept to implementation (including labor). Although MD-Grape was custom-designed for molecular dynamics, it has proven useful for several other applications, including astrophysical N-body simulations.

An example of a semicustom design with some custom elements is the D. E. Shaw system, expected to be completed in 2008. D. E. Shaw Research, a unit of an investment firm, focuses on the development of new algorithms and specialized supercomputer architectures for ultra-fast biomolecular simulations of scientific and pharmaceutical problems. The D. E. Shaw system will use fully programmable cores with full-custom co-processors to achieve efficiency, and will simulate 100 to 1000 times longer timescales than any existing HPC system. While the programmability of the D. E. Shaw system will broaden its application reach, it will still be narrower than NERSC's proposed Green Flash.

IBM's Blue Gene is the best example to date of the kind of application-driven architecture based on an embedded processor core that NERSC envisions for Green Flash. Designed around a protein folding application, Blue Gene, over several generations, has proved to be useful for a growing list of applications, including hydrodynamics, quantum chemistry, molecular dynamics, climate modeling, and financial modeling.

Like Blue Gene, NERSC's ultra-high-resolution Green Flash would have a semicustom design. The core architecture would be highly programmable using C, C++, or Fortran. Its 100x improvement in power efficiency would be modest when compared with the demonstrated capability of more specialized approaches. *This approach would solve an exascale problem without building an exaflop/s machine.*

RAMP will be used as a testbed to design the system architecture in the context of climate model algorithms. The software implementation will be tailored to take advantage of

Michael Wehner

Lenny Oliker

John Shalf

Jonathan Carter

Erich Strohmaier

hard design limits or features of the evolving hardware implementations. NERSC's partners in the climate community will be David Randall of Colorado State University (leader of the SciDAC Global Cloud Resolving Model project), and Michael Wehner and Bill Collins of Berkeley Lab.

In addition to enabling a breakthrough in cloud-resolving climate simulation, NERSC's power-efficient, application-driven design methodology will have an impact on the broader DOE scientific workload. Our hardware/software co-design approach is geared for a *class* of codes, not just for a single code instantiation. This methodology is broadly applicable and could be extended to other scientific disciplines. Blue Gene was originally targeted at chemistry and bioinformatics applications, resulting in a power-efficient architecture, but its application has been broader than the original target. NERSC expects a similar result from the Green Flash.

Since the Green Flash concept was unveiled at the 2007 SIAM Conference on Computational Science and Engineering, it has attracted international attention.[20,21,22] Berkeley Lab is currently funding development of a prototype, with Michael Wehner, Lenny Oliker, and John Shalf leading the effort.

Related Research

Berkeley Lab is also funding two other projects on energy-efficient computing led by NERSC researchers:

- Jonathan Carter, head of the User Services Group, is leading a project to explore a wide range of multicore computer architectures and how efficiently those systems can perform on challenging scientific codes. The project, "Enhancing the Effectiveness of Manycore Chip Technologies for High-End Computing," also includes collaborators Lenny Oliker and John Shalf. They will identify candidate algorithms that map well to multicore technologies, and document the steps needed to re-engineer programs to take advantage of these architectures. They will also try to identify design elements in multicore chips

[20] Ashlee Vance, "Geeks fight the smelter with embedded processor-based box," The Register, February 2, 2008, http://www.theregister.co.uk/2008/02/02/horst_simon_cloud_computer/.
[21] Michael Feldman, "A Modest Proposal for Petascale Computing," HPCwire, February 8, 2008, http://www.hpcwire.com/hpc/2112632.html.
[22] Economist.com, "Cool it!" March 4, 2008, http://www.economist.com/displaystory.cfm?story_id=10795585.

that would contribute to a better high performance system.

- A project led by Erich Strohmaier proposes to develop a testbed for benchmarking of key algorithms that will be crucial for designing software and computers that use processors with many cores on each chip. This project is being conducted with domain experts from Berkeley Lab's Computational Research Division (CRD), NERSC, and UC Berkeley. The project, "Reference Benchmarks for the Dwarfs," will devise ways to use a set of algorithms to gauge the performance of systems from personal computers to high performance systems. The algorithms are known as *dwarfs*; each dwarf represents a class of algorithms with similar properties and behavior. The 13 dwarfs chosen for the research include algorithms important for the scientific community. Strohmaier is head of the Future Technologies Group in CRD and is a member of the NERSC Science-Driven System Architecture team.

NERSC Data

Virtually all branches of science base hypothesis testing on some form of data analysis. Scientific disciplines vary in how they produce data (via observation or simulation), in how they manage data (storage, retrieval, archiving, indexing, summaries, sharing across the science team), and in how they analyze data and communicate results. But it is widely agreed that one of the primary bottlenecks in modern science is managing and discovering knowledge in light of the tsunami of data resulting from increasing computational capacity and the increasing fidelity of scientific observational instruments.[23] And as data become too large to move, we are evolving towards a model where data-intensive services are centrally located.[24] The proposed NERSC Data effort will offer a diverse set of activities to meet this demand, including but not limited to:

- community-oriented data repositories
- browsing, exploration, and analysis capabilities that operate on the centrally located community repositories
- providing and maintaining the centrally located hardware and software infrastructure that enables these capabilities.

A key element of the NERSC long-term strategy is production-quality data management and analytics with sufficient resources to meet science needs. The potential impact to DOE in long-term cost savings and scientific opportunity is profound.

For example, NERSC already serves as the data repository for two international nuclear physics collaborations, KamLAND (page 46 above) and STAR, maintaining hundred of terabytes of data on scalable storage platforms that are managed by professional systems engineers. Centralized data storage frees scientists in both projects from spending time on system administration, allowing them to focus on the science.

There are many existing or imminent projects that could benefit from this kind of data platform, including ESG, LHC, ITER, JDEM/SNAP, Planck, SciDAC Computational Astrophysics Consortium, and JGI. Providing community-oriented data repositories for a large number of projects, along with advanced analytics tools that help extract meaning from the data, is outside the scope of the current NERSC program, but would allow NERSC to more effectively fulfill its mission of enabling scientific discovery.

Easy Access to Data Accelerates Science

The value of accessing massive datasets with powerful analytic tools was illustrated in the 2005 National Institute of Standards and Technology (NIST) Open Machine Translation Evaluation, which involved academic, government, and commercial participants from all over the world. Although it was Google's first time competing, their translation system achieved the highest scores in both Arabic- and Chinese-to-English translation, outperforming sophisticated rules-based systems devel-

[23] Richard P. Mount, ed., The Office of Science Data-Management Challenge: Report from the DOE Office of Science Data-Management Workshops, March–May 2004; http://www.sc.doe.gov/ascr/ProgramDocuments/Final-report-v26.pdf.

[24] Gordon Bell, Jim Gray, and Alex Szaley, "Petascale Computational Systems," IEEE Computer 39(1), January 2006.

oped by expert linguists.[25] Google used statistical learning techniques to build its translation models, feeding the machines billions of words of text, including matching pairs of human-translated documents.[26] In this case, Google, with more data, beat others with more expertise.

Similar results can be expected from applying advanced analytics tools to massive scientific datasets. Indeed, several projects at NERSC and Berkeley Lab's Computational Research Division (CRD) illustrate the growing need for integrated, production-quality data management and analytics. One such project is the cosmic microwave background data analysis for the Planck satellite mission. The satellite is scheduled to be launched in 2008, but the data production pipeline is already in place at NERSC. Access to both raw and processed data will be provided through a web portal for a remote community of thousands of users.

Another example of production analytics is Sunfall, the collaborative visual analytics and data exploration system created with the Nearby Supernova Factory, as described on page 48. The development of Sunfall led to a 90% labor savings in areas of the SNfactory supernova search and follow-up workflow; and project scientists now have new data exploration and analysis capabilities that had previously been too time-consuming to attempt.

A web application that combines data and functionality from more than one source is the Berkeley Water Center's (BWC's) Scientific Data Server,[27] which integrates data from several hundred FLUXNET environmental observatories worldwide to improve the understanding of carbon fluxes and carbon-climate interactions. Using Microsoft SharePoint collaboration tools and an integration with MS Virtual Earth, the BWC server offers 921 site years of data (150 variables at a 30 minute data rate) available for direct download into MS Excel.

These examples are initial illustrations of how the data needs of the scientific community are changing. These changes can be summarized as follows:

- Increasing size of data sets from experimental systems (satellites, detectors, etc.) in addition to the simulation data that has always grown with machine size and speed.
- Growing size, geographic distribution, and diversity of communities that share a common data repository; each community may be made up scientists with different specializations studying different features of the shared data set.
- Increased used of and reliance on production-quality information management, workflow, and analytics software infrastructure.

NERSC Data Program Elements

To anticipate and meet the changing needs of its user communities, the new NERSC Data program will include:

- The next-generation mass storage system
- Production infrastructure for data
  - o Hardware: computational platforms
  - o Software for data management, analysis/analytics, and interfaces between integrated data components
- Development or adaptation of reusable, broad-impact tools
  - o Analogous to Google Earth or Microsoft SharePoint
  - o Hosting and adapting SciDAC tools for the science community
- Focused data projects
  - o Consulting expertise in scientific data management, analytics, visualization, workflow management, etc.

NERSC Data Storage

NERSC is a founding development partner in the High Performance Storage System (HPSS) project,[28] software that manages petabytes of data on disk and robotic tape libraries. While HPSS has proven to be an invaluable mass storage platform, it is now about 15 years old and likely will not evolve to meet fu-

[25] NIST 2005 Machine Translation Evaluation Official Results, August 1, 2005, http://www.nist.gov/speech/tests/mt/2005/doc/mt05eval_official_results_release_20050801_v3.html.
[26] Bill Softky, "How Google translates without understanding," The Register, May 15, 2007, http://www.theregister.co.uk/2007/05/15/google_translation/print.html.
[27] http://bwc.berkeley.edu/
[28] http://www.hpss-collaboration.org/hpss/index.jsp

ture science needs. The current ways of storing data in global filesystems and archival storage systems will probably not scale to exascale. Past experience has shown that commercial storage products designed for a mass market will not meet the needs of open science.

The open science community needs to initiate the collaborative development of what might be called EXA-HPSS — a next-generation mass storage system. This system must be energy efficient, scalable, closely integrated with parallel filesystems and online data, and designed for the requirements of new data profiles (e.g., the increasing importance of metadata). Archival storage needs to go beyond file-based access to support a broader set of data storage and retrieval operations and more user-friendly functionality. With decades of experience serving a large community of science users, NERSC is in the best position to lead the specification, design, and research effort for a next-generation mass storage system, and to participate in R&D of an interface to support efficient use of the system.

Berkeley Lab's Scientific Data Management Group is already initiating research into an "Energy-Smart Disk-Based Mass Storage System," envisioned as an energy-efficient, low-latency, scalable mass storage system with a three-level hierarchy (compared to HPSS's two-level hierarchy). Today's storage systems in data centers use thousands of continuously spinning disk drives. These disk drives and the necessary cooling components use a substantial fraction of the total energy consumed by the data center. This project is exploring new configurations that divide the disks into active and passive groups. The active group contains continuously spinning disks that act as a cache for the most frequently accessed data. The disks in the passive group would power down after a period of inactivity. This is a prime example of the kind of research needed to develop the next generation of storage technology.

NERSC Data Production Infrastructure

The NERSC Data production infrastructure will consist of computational platforms for high-capacity and high-throughput interactive analytics, high-capacity and energy-efficient mass storage, high performance intra- and inter-networking capability, and a robust collection of software tools for realizing production analytics solutions. Software tools will include applications and libraries for data management, analysis, visualization, and exploration, as well as applications and libraries enabling scientific community access, e.g., web portal infrastructure, a new data archive interface, etc.

A good analogy for this infrastructure is Google, where a significant investment in computational and software infrastructure enables the retrieval of data most relevant to a query from a variety of sources and presents it quickly in an easily comprehensible, usable, and navigable form. NERSC's long-term vision is to provide this type of on-demand capability ("Google for Science") to our users and stakeholders. The resulting solutions span a diverse range: community-centric data repositories and analysis, portal-based interfaces to data and computation, high performance and production-quality visual analytics pipelines/workflows and systems.

The role of the new NERSC Data program is to provide the hardware infrastructure commensurate with need, which includes both sufficient capacity (absolute number of CPUs, memory, storage, network bandwidth, etc.) as well as capability (e.g., GPUs, large memory footprint nodes, etc.).

NERSC Data Tools

"Google for Science" may become the next paradigm for scientific analytics if one considers the powerful capabilities that Google and similar search engines put in the hands of anyone with Internet access. For example, in response to a user query about a given location (address, intersection, landmark, or latitude and longitude), Google Maps and Google Earth can access a plethora of satellite, aerial, and surface photos, map images, and textual information on roads, buildings, businesses, landmarks, and geography, then present that information in the desired format — text, map tiles, pictures, or a combination. Users can zoom in or out of images, change the directional orientation, change the angle of aerial photos, or move in any direction, all through an easy-to-use web interface. Providing the ability to access, navigate, and manipulate scientific data this easily is NERSC's vision of "Google for Science."

One of the keys tools developed at Google that makes these capabilities possible is MapReduce, a programming model and an associated implementation for generating and processing large data sets.[29] MapReduce is used to regenerate Google's index of the World Wide Web as well as perform a wide variety of analytic tasks, including grep, clustering, data mining, and machine learning — more than ten thousand applications to date. The basic steps in MapReduce are:

- read a large quantity of data
- *map* the data: extract interesting items
- shuffle and sort
- *reduce*: aggregate and transform the selected data
- write the results.

MapReduce has a number of features that suggest it could be used very productively in scientific analytics. Its functions can be applied to numeric, image, or text data, e.g., simulations, telescopic images, or genomic data. Its simple, extensible interface allows for domain-specific analysis and leverages domain-independent infrastructure. It makes efficient use of wide area bandwidth by shipping functions to the raw data and returning filtered information. And it hides messy details, such as parallelization, load balancing, and machine failures, in the MapReduce runtime library, allowing programmers who have no experience with distributed or parallel systems to exploit large amounts of resources easily.

A key element of the NERSC Data program will be to develop or adapt reusable, broad-impact open tools such as MapReduce, Microsoft SharePoint, or other software that can simplify production analytics, allowing researchers to focus on scientific discovery rather than the detailed operation of analytics tools.

Hosting and adapting SciDAC-developed tools for the science community will be an essential part of this effort. DOE's investment in data management technologies focuses on infrastructure for data storage and I/O (PDSI, HPSS), indexing and searching (SDM Center), workflow management (SDM Center), and location-transparent access to distributed data (SDM Center and Open Science Grid). Such infrastructure generally consists of software comprised of standalone executables to libraries of callable methods/routines that implement focused capabilities. A gap in DOE's program, and consequently a long-term challenge, is having dedicated professional staff at production computing facilities responsible for the ultimate production deployment of such new capabilities. Given the demands of production use, there will necessarily be periods of evaluation where new technologies are subject to beta testing, including scrutiny by cybersecurity experts. This activity of the NERSC Data program — tool evaluation, testing, and feedback — will fill a gap in DOE's current computational programs.

Focused Data Projects

The NERSC Data program will provide integrated, production-quality analytics pipelines for experimental and computational science projects, working directly with science stakeholders to design and deploy production analytics capabilities for science communities. This unique capability is targeted at science projects that have a significant need for high performance, production-quality data management, processing, and analysis capabilities (e.g., ESG, JDEM/SNAP, etc.).

The scope of these production analytics capabilities is diverse and will be driven by data-intensive science needs. This approach, where the primary focus is on the needs of individual science communities, provides opportunities for major breakthroughs in both the domain science and computer science, with the additional benefit of spinning off generally applicable technologies for broader use. This service will be allocated through an INCITE-like process to take advantage of the NERSC staff's expertise in consulting, analytics, and technology evaluation, testing, integration, and hardening.

NERSC anticipates working in the following areas:

- **Data formats and models.** High performance, parallel data I/O libraries will optimize data storage, retrieval, and exchange on NERSC and other platforms. The NERSC Data team will eval-

[29] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Proc. OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004; http://labs.google.com/papers/mapreduce-osdi04.pdf.

uate these technologies, participate in R&D efforts to create an improved technology, and directly consult with science projects to deploy these technologies.

- **Community-wide data access.** Science projects need straightforward, unfettered, yet authenticated and authorized access to their community data regardless of location across multiple sites. Access to data could potentially be based on files, like the current approach familiar to users of HPSS and other typical filesystems; or based on "objects," where the result of a "data gather" operation is performed by an agent on the user's behalf and later made available to the user.
- **Data filtering and processing.** Often raw data must undergo additional processing (e.g., gap filling, filtering, etc.) before being ready for downstream use by consumers. NERSC Data staff will contribute in several areas, such as knowledge of the best algorithms for filtering/processing and their deployment on parallel machines, and assistance in deploying these algorithms in scientific workflows.
- **Data exploration.** Science users want to be able to quickly and easily explore their data, either with a traditional application (run on NERSC resources or at the user's location) that reads files and displays results, or through a web-based application that interfaces to back-end infrastructure at NERSC to access and process data, then displays results through the web interfaces. The intersection between filtering and exploration can be based on queries, which return subsets of data that meet certain criteria. In the commercial world, such systems are typically implemented using a relational database management system and SQL queries. A body of work from DOE's Scientific Data Management SciDAC program shows that commercial RDBMS systems are not adequate to meet the needs of large, data-intensive science activities.[30]
- **Data analysis.** These activities include generating statistical summaries and analysis, supervised and unsupervised classification and clustering, curve fitting, and so forth. While many contemporary applications provide integrated data analysis capabilities, some science projects will want to run standalone analysis tools on data collections offline as part of a workflow to produce derived data for later analysis.
- **Data visualization.** The role of visualization and visual data analysis in the scientific process is well established.
- **Workflow management.** The goal of workflow management is to automate specific sets of tasks that are repeated many times and thus simplify execution and avoid typical human errors that often occur when repetitive tasks are performed.
- **Interfaces and usability.** Recent production analytics workflows like Sunfall show the dramatic increase in scientific productivity that results from careful attention to the combination of highly capable analytics software and highly effective interfaces to those software tools. One primary NERSC Data objective is to increase scientific productivity for data-intensive activities through well designed and engineered interfaces.

NERSC Data staff will be a bridge between programs in production computing and data storage and complementary applied research efforts in visualization, data management, ultra-efficient platforms, networking, distributed systems, and networking middleware. This bridge will complete the cycle between research and development and production deployment in computing centers, with active participation in areas of emerging architectures and novel algorithms.

With many decades of experience serving a large community of science users, NERSC is in the best position to deliver production-quality

[30] K. Wu, W.-M. Zhang, V. Perevoztchikov, J. Laurent, and A. Shoshani, "Grid Collector: Using an Event Catalog to Speed Up User Analysis in a Distributed Environment," presented at Computing in High Energy and Nuclear Physics (CHEP) 2004, Interlaken, Switzerland, September 2004; http://www.osti.gov/bridge/servlets/purl/882078-E3rSLU/882078.PDF.

data management and knowledge discovery infrastructure to the DOE science community. The NERSC Data program will expand the scope of the NERSC mission to include capabilities that are responsive to the data needs of DOE science, needs that are inseparable from the computational requirements.

Appendix A

NERSC Policy Board

Daniel A. Reed (Chair)
Microsoft Corporation

David Dean
(ex officio, NERSC Users Group Chair)
Oak Ridge National Laboratory

Robert J. Goldston
Princeton Plasma Physics Laboratory

Tony Hey
Microsoft Corporation

Sidney Karin
University of California, San Diego

Pier Oddone
Fermi National Accelerator Laboratory

Tetsuya Sato
Earth Simulator Center/Japan Marine Science and Technology Center

Stephen L. Squires
Hewlett-Packard Laboratories

Appendix B

NERSC Client Statistics

In support of the DOE Office of Science's mission, the NERSC Center served 3,113 scientists throughout the United States in 2007. 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.

More than 1,500 scientific publications in 2007 were based entirely or in part on calculations done at NERSC; a list is available at http://www.nersc.gov/news/reports/ERCAPpubs07.php.

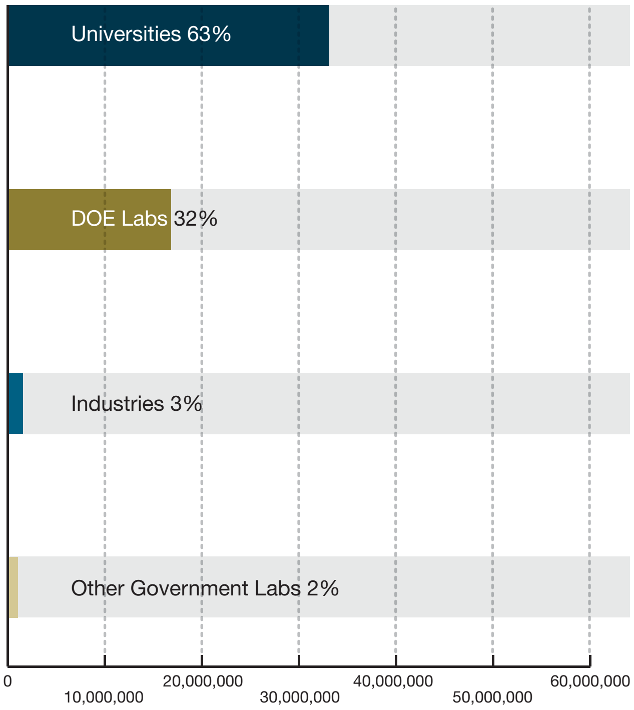


Figure 1. NERSC MPP usage by institution type, 2007.

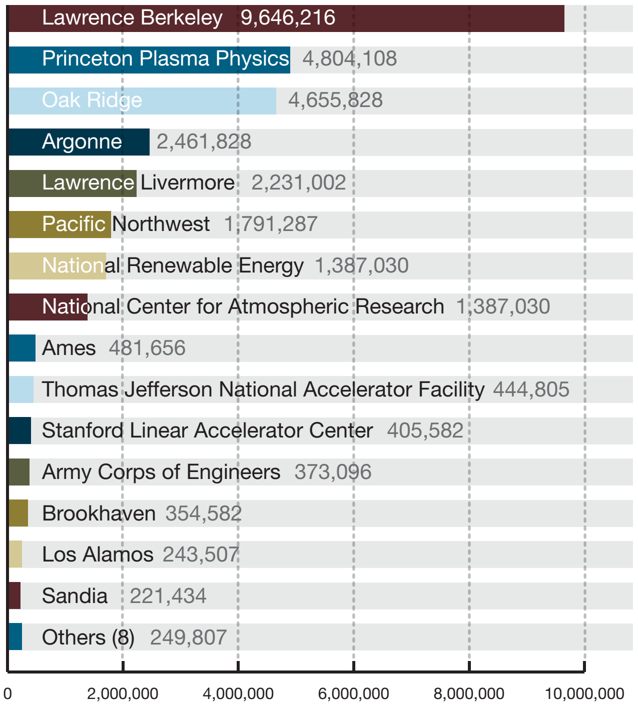


Figure 2. DOE and other Federal laboratory usage at NERSC, 2007 (MPP hours).

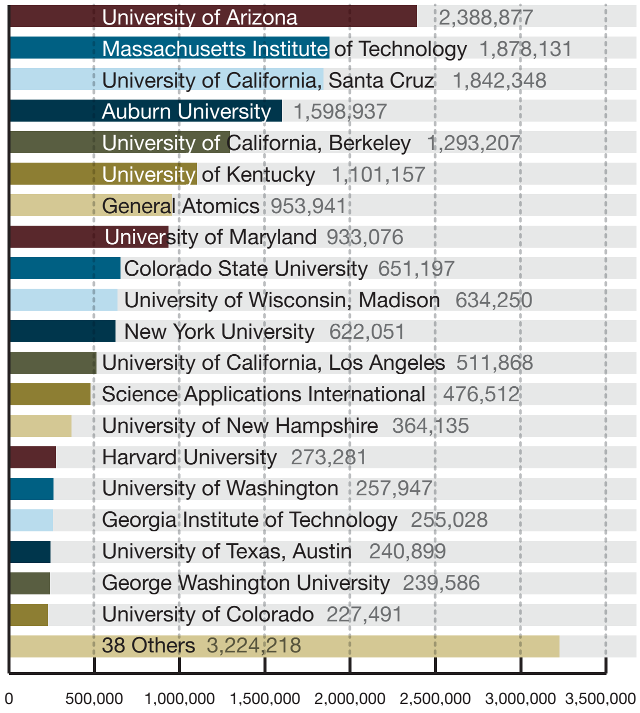


Figure 3. Academic and private laboratory usage at NERSC, 2007 (MPP hours).

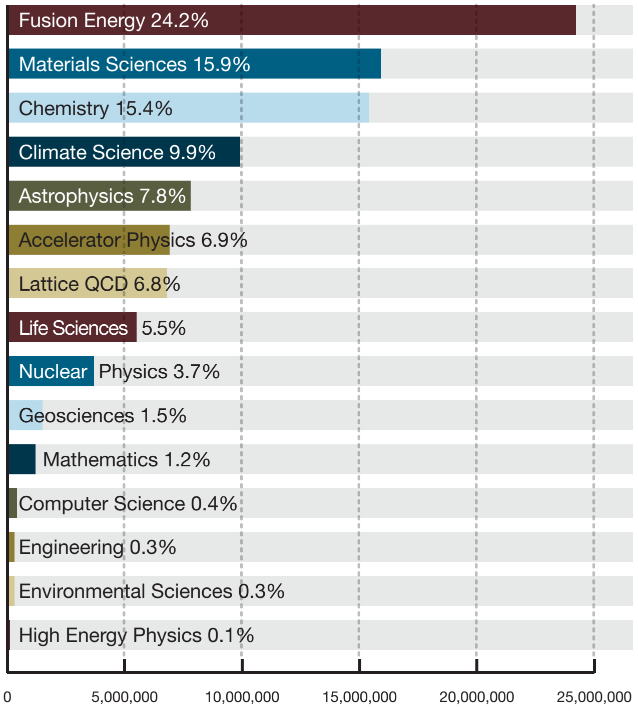


Figure 4. NERSC usage by scientific discipline, 2007 (MPP hours).

Appendix C

NERSC Users Group Executive Committee

Office of Advanced Scientific Computing Research

Kirk Cameron, Virginia Polytechnic Institute and State University
Mike Lijewski, Lawrence Berkeley National Laboratory
Ravi Samtaney, Princeton Plasma Physics Laboratory

Office of Basic Energy Sciences

Bas Braams, Emory University
Eric Bylaska, Pacific Northwest National Laboratory
Thomas Miller, University of California, Berkeley

Office of Biological and Environmental Research

David Beck, University of Washington
Brian Hingerty, Oak Ridge National Laboratory
Adrianne Middleton, National Center for Atmospheric Research

Office of Fusion Energy Sciences

Andris Dimits, Lawrence Livermore National Laboratory
Stephane Ethier (Vice Chair), Princeton Plasma Physics Laboratory
Alex Friedman,* Lawrence Livermore and Lawrence Berkeley National Laboratories
Jean-Luc Vay,** Lawrence Berkeley National Laboratory

Office of High Energy Physics

Olga Barranikova,* University of Illinois at Chicago
Julian Borrill,** Lawrence Berkeley National Laboratory
Cameron Geddes,** Lawrence Berkeley National Laboratory
Warren Mori,* University of California, Los Angeles
Frank Tsung, University of California, Los Angeles

Office of Nuclear Physics

David Bruhwiler,** Tech-X Corporation
David Dean (Chair),* Oak Ridge National Laboratory
Patrick Decowski,* Lawrence Berkeley National Laboratory
Peter Messmer,** Tech-X Corporation
James Vary, Iowa State University

Members at Large

Yuen-Dat Chan,* Lawrence Berkeley National Laboratory
Angus Macnab,** Woodruff Scientific, LLC
Ned Patton,** National Center for Atmospheric Research
Gerald Potter, Lawrence Livermore National Laboratory
Douglas Swesty,* State University of New York at Stony Brook
Xingfu Wu, Texas A&M University

** Outgoing members*
*** Incoming members*

Appendix D

Office of Advanced Scientific Computing Research

The primary mission of the Advanced Scientific Computing Research (ASCR) program is to discover, develop, and deploy the computational and networking tools that enable researchers in the scientific disciplines to analyze, model, simulate, and predict complex phenomena important to the Department of Energy. To accomplish this mission, the program fosters and supports fundamental research in advanced scientific computing—applied mathematics, computer science, and networking—and operates supercomputer, networking, and related facilities. In fulfilling this primary mission, the ASCR program supports the Office of Science Strategic Plan's goal of providing extraordinary tools for extraordinary science as well as building the foundation for the research in support of the other goals of the strategic plan. In the course of accomplishing this mission, the research programs of ASCR have played a critical role in the evolution of high performance computing and networks. Berkeley Lab thanks the program managers with direct responsibility for the NERSC program and the research projects described in this report:

Michael R. Strayer
Associate Director, ASCR
Melea Baker
Administrative Specialist
Barbara Helland
Senior Advisor
Julie Scott
Financial Management Specialist
Betsy Riley
Detailee

Facilities Division

Michael R. Strayer
Acting Division Director
Daniel Hitchcock
Senior Advisor
Sally McPherson
Program Support Specialist
Vincent Dattoria
General Engineer
Robert Lindsay
Computer Scientist
Yukiko Sekine
Computer Scientist

Computational Science Research and Partnerships (SciDAC) Division

Fred Johnson
Acting Division Director
Amy Clark
Program Support Specialist
Teresa Beachley
Program Support Assistant
Walter Polansky
Senior Scientific Advisor
Christine Chalk
Physical Scientist
Lali Chatterjee
Physical Scientist
George Seweryniak
Computer Scientist
Thomas Ndousse-Fetter
Program Manager
Steven Lee
Mathematician, Detailee
Susan Turnbull
Detailee
Bill Spotz
Mathematician, IPA
Sandy Landsberg
Mathematician
Osni Marques
Computer Scientist

Appendix E

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.

Jill P. Dahlburg, Chair
Naval Research Laboratory

Robert G. Voigt, Co-Chair
College of William and Mary

F. Ronald Bailey
NASA Ames Research Center (retired)

Gordon Bell
Microsoft Bay Area Research Center

Marsha Berger
Courant Institute of Mathematical Sciences

David J. Galas
Battelle Memorial Institute

Roscoe C. Giles
Boston University

James J. Hack
Oak Ridge National Laboratory

Thomas A. Manteuffel
University of Colorado at Boulder

Horst D. Simon
Lawrence Berkeley National Laboratory

Ellen B. Stechel
Sandia National Laboratories

Rick L. Stevens
Argonne National Laboratory

Virginia Torczon
College of William and Mary

Thomas Zacharia
Oak Ridge National Laboratory

Appendix F

Acronyms and Abbreviations

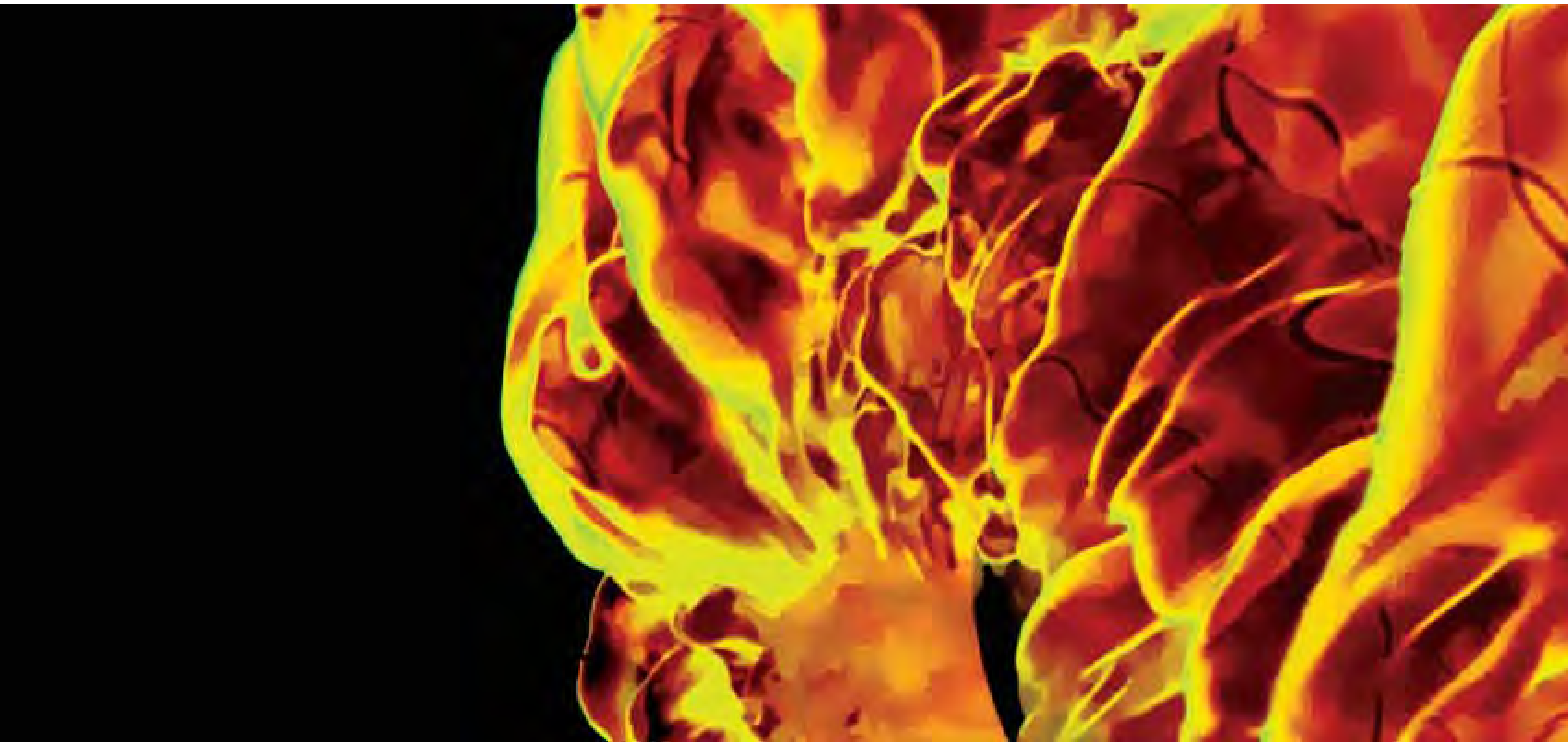
ASC	Advanced Simulation and Computing (DOE)
ASCR	Office of Advanced Scientific Computing Research (DOE)
ASIC	Application-specific integrated circuit
BER	Office of Biological and Environmental Research (DOE)
BES	Office of Basic Energy Sciences (DOE)
BWC	Berkeley Water Center
CCSM	Community Climate System Model
CHiMES	Coupled High-Resolution Modeling of the Earth System
CIRES	Cooperative Institute for Research in Environmental Sciences
CLE	Cray Linux Environment
CPU	Central processing unit
CRD	Computational Research Division, Lawrence Berkeley National Laboratory
CSCS	Swiss National Supercomputing Centre
DOE	U.S. Department of Energy
DT	Deuterium-tritium
ESG	Earth Systems Grid
FES	Office of Fusion Energy Sciences (DOE)
FMO	Fenna-Matthews-Olson (photosynthetic protein)
FPGA	Field programmable gate array
GB	Gigabyte
GFDL	Geophysical Fluid Dynamics Laboratory (NOAA)
GPU	Graphics processing unit
HECRTF	High-End Computing Revitalization Task Force
HEP	Office of High Energy Physics (DOE)
hPa	Hectopascals
HPC	High performance computing
HPSS	High Performance Storage System
ICF	Inertial confinement fusion
IEEE	Institute of Electrical and Electronics Engineers
INCITE	Innovative and Novel Computational Impact on Theory and Experiment (DOE)
I/O	Input/output
IPCC	Intergovernmental Panel on Climate Change
IPM	Integrated Performance Monitoring
ITER	A multinational tokamak experiment to be built in France (Latin for “the way”)
JDEM	Joint Dark Energy Mission
JGI	Joint Genome Institute (DOE)
KamLAND	Kamioka Liquid Scintillator Anti-Neutrino Detector
KRFG	Korea Research Foundation Grant
LBNL	Lawrence Berkeley National Laboratory
LED	Light-emitting diode
LHC	Large Hadron Collider
LHC2	Light-harvesting complex 2
LLNL	Lawrence Livermore National Laboratory
MB	Megabyte
MIBRS	Miller Institute for Basic Research in Science
MSCF/EMSL	Molecular Science Computing Facility at the Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory

MW	Megawatt
NCAR	National Center for Atmospheric Research
NCCS	National Center for Computational Sciences at Oak Ridge National Laboratory
NCEP	National Centers for Environmental Prediction
NERSC	National Energy Research Scientific Computing Center
NIF	National Ignition Facility
NIM	NERSC Information Management system
NIST	National Institute of Standards and Technology
NOAA	National Oceanographic and Atmospheric Administration
NP	Office of Nuclear Physics (DOE)
NSF	National Science Foundation
ORNL	Oak Ridge National Laboratory
OSG	Open Science Grid
PB	Petabyte
PCM	Parallel Climate Model
PDA	Personal digital assistant
PDSI	Petascale Data Storage Institute (SciDAC)
PDSF	Parallel Distributed Systems Facility (NERSC)
PEM	Polymer electrolyte membrane
Petaflops	Quadrillions of floating point operations per second
PEtot	Parallel Total Energy code
Pflops	Petaflops
PI	Principal investigator
PIC	Particle-in-cell
PNNL	Pacific Northwest National Laboratory
RAMP	Research Accelerator for Multiple Processors
RDBMS	Relational database management system
SC	Office of Science (DOE)
SciDAC	Scientific Discovery through Advanced Computing (DOE)
SDM	Scientific Data Management Center (SciDAC)
SDSC	San Diego Supercomputer Center
SIAM	Society for Industrial and Applied Mathematics
SLP	Sea level pressure
SNAP	SuperNova Acceleration Probe
Teraflops	Trillions of floating point operations per second
UC	University of California
UCLA	University of California, Los Angeles
UPC	Unified Parallel C
VASP	Vienna Ab-Initio Simulation Package
VFF	Valence force field
WRF	Weather Research and Forecast

For more information about NERSC, contact:

Jon Bashor
NERSC Communications
Berkeley Lab, MS 50B4230
1 Cyclotron Road
Berkeley, CA 94720-8148

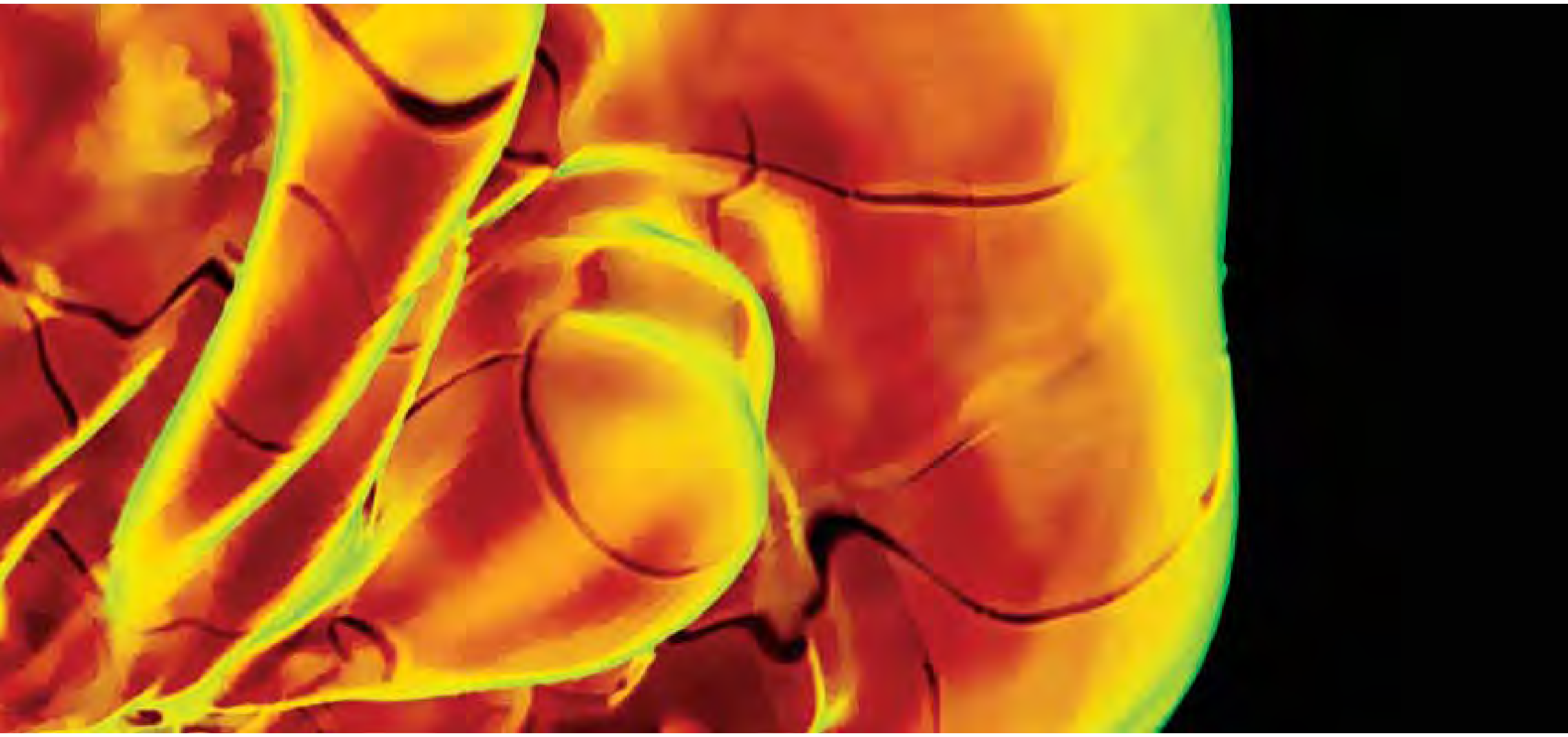
email: jbashor@lbl.gov
phone: (510) 486-5849
fax: (510) 486-4300



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

Published by the Berkeley Lab Creative Services Office in collaboration with NERSC researchers and staff.
JO15600

Editor and principal writer: John Hules.
Contributing writers: Jon Bashor, Ucilia Wang, Lynn Yarris, and Paul Preuss (Berkeley Lab); Warren Froelich and Jan Zverina (San Diego Supercomputer Center).



DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.

NERSC

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY

Office of Science
U.S. DEPARTMENT OF ENERGY