NERSC NEWS
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER
Berkeley Lab
Office of Science
U.S. DEPARTMENT OF ENERGY
DECEMBER 2008

Humanities and High Performance Computers Connect at NERSC

High performance computing and the humanities are finally connecting — with a little matchmaking help from the Department of Energy (DOE) and the National Endowment for the Humanities (NEH). Both organizations have teamed up to create the Humanities High Performance Computing Program, a one-of-a-kind initiative that gives humanities researchers access to some of the world's most powerful supercomputers.

As part of this special collaboration, NERSC will dedicate a total of one million compute hours on its supercomputers and technical training to humanities experts to bring them up to speed on HPC. Meanwhile, the program's participants were selected through a highly competitive peer review process led by the NEH's Office of Digital Humanities.

"A connection between the humanities and high performance computing communities had never been formally established until this collaboration between DOE and NEH. The partnership allows us to realize the full potential of supercomputers to help

continued on page 3

Cultural Analytics: This image is an example of an Interface design for the Cultural Analytics research environment, which is being developed at UCSD's Software Studies Initiative. The project has received compute time at NERSC as part of an innovative collaboration between the Department of Energy and the National Endowment for the Humanities.

NERSC to Provide Resources to INCITE Projects Studying Combustion, Fusion Energy, Materials and Accelerator Design

Researchers tackling some of the most challenging scientific problems, from improving energy efficiency in combustion devices to developing new particle accelerators for scientific discovery to studying properties of new materials, have been awarded access to supercomputing resources at NERSC.

The awards, announced Dec. 17 by DOE's Office of Science, are made under the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) Program. Seven projects were awarded a total of 17,460,000 processor-hours at NERSC after a competitive review. Launched in 2003, INCITE selects projects that not only require large-scale and intensive use of supercomputers but also promise to deliver a significant advance in science and engineering.

"As the original home of the INCITE program, NERSC staff are working to provide the necessary support for advancing these high-impact science projects while maintaining our commitment to all other users, whose work has broad impacts across all scientific disciplines," said NERSC Division Director Katherine Yelick.

continued on page 5

Prize-Winning Algorithm Perfected at NERSC

A computer method created to understand the energy-harnessing potential of nanostructures for solar cell design was awarded a special Gordon Bell Prize for algorithm innovation by the Association for Computing Machinery (ACM).

Scientists say the award-winning Linearly Scaling Three Dimensional Fragment (LS3DF) method's initial runs on NERSC's Cray XT4 system, nicknamed Franklin, provided key insights about the application that eventually paved the way for its ACM Gordon Bell Prize win.

"By incorporating the correct chemical formulas into efficient computer programs, scientists can learn a lot about the structures and properties of molecules and solids," says Lin-Wang Wang, a computational material scientist who led the development of LS3DF. "I like to think of computers as chemistry's third pillar. In most cases, computer simulations complement information obtained by chemical experiments, but in some cases they can also predict unobserved phenomena."

According to Wang, nanostructures — tiny materials 100,000 times finer than a human hair — may hold the key to energy independence. He believes that a fundamental understanding of nanostructure behaviors and properties could provide solutions for curbing our dependence on petroleum, coal and other fossil fuels.

The LS3DF computer algorithms use a novel "divide-and-conquer" technique to efficiently gain insights into how nanostructures function in systems with 10,000 or more atoms, giving scientists an idea of the structure's energy harnessing potential. More traditional methods calculate the entire structure as a whole system and are much more time consuming and resource intensive.

"We still don't quite understand how the electron moves around in a nanostructure, and how such properties depend on the size, geometry, composition and surface passivations," said Wang. "Understanding this dependence will allow us to design nanostructures for desired applications. Using our improved LS3DF method will help us to understand and predict these properties."

continued on page 2

Prize-Winning Algorithm *continued from page 1*

The method's first successful science run occurred on 17,280 cores of the original dual-core Franklin computer system at NERSC. In one hour, team members calculated the electronic structure of a 3,500-atom ZnTeO alloy. This run verified that the code could be used to compute properties of the ZnTeO alloy that previously had been experimentally observed. This calculation reached 36 teraflop/s, or 36 trillion calculations per second, at 40 percent efficiency. Wang says this run paved the way for the LS3DF paper to be one of six Gordon Bell finalists.

Innovative Leader: Lin-Wang Wang helps LS3DF team secure a special Gordon Bell Prize for Algorithm Innovation.

According to Wang, LS3DF clinched the Gordon Bell Prize for algorithm innovation on another Franklin test-run. This year, Franklin was upgraded from a dual-core system to a quad-core system, and now contains 38,640 processors. After the upgrade, the LS3DF code reached 136 teraflop/s with 40 percent efficiency on the machine. He notes that this result essentially secured the win.

"Getting codes to run with such high efficiencies on massively parallel machines like Franklin is not a trivial task. We are very grateful to NERSC for providing computer time for us to do full machine runs on Franklin. Without the help from NERSC, this project would not be possible. We are also thankful for the support from NERSC staff," says Wang.

He notes that Zhengji Zhao of NERSC's User Services Group played a unique role in the project. In addition to supporting LS3DF team members with technical issues on Franklin, she is one of the original developers of the LS3DF method.

"Zhengji carried out many tests to make sure the code ran at the optimum parameters and at the large concurrencies. Such expertise from a NERSC staff member is critical in carrying out our mission," says Wang.

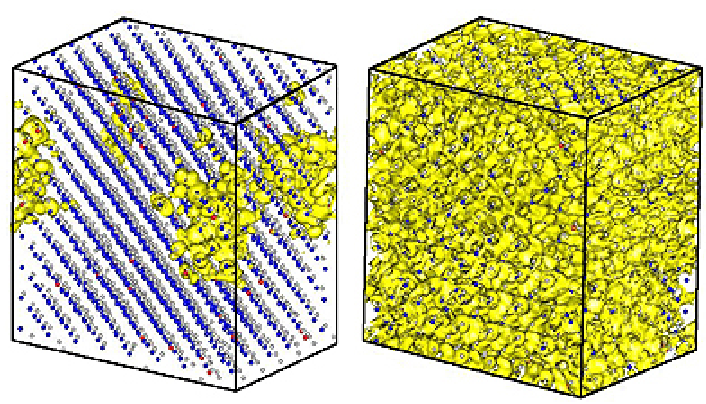
"It is a great experience to participate in such an exciting project," says Zhao. "NERSC has a long tradition of encouraging its staff to participate in scientific projects, and I am glad to see the LS3DF code, which was my postdoc project three years ago, run so well. Now that I am in a career position at NERSC, I look forward to taking the lessons learned in this project and using them to help the scalability of other applications."

Other members of the LS3DF team — David Bailey, Hongzhang Shan and Erich Strohmaier of the SciDAC Performance Engineering Research Institute — also worked hand-in-hand with Wang and his colleagues to analyze the performance of LS3DF and to identify potential performance improvements. Responding to this analysis, Berkeley Lab researchers assisted with a major revision of the code, which led to the prize-winning submission.

In addition to Wang, Zhao, Shan, Bailey, and Strohmaier, other authors on the prize-winning LS3DF paper include Byounghak Lee and Juan Meza, also from Berkeley Lab.

Once the LS3DF code had been optimized, it was a matter of days before it was running at each of the DOE supercomputing facilities. At the Argonne Leadership Computing Facility, the code ran on 163,840 cores of the IBM BlueGene/P system, reaching 224 teraflop/s or 40.5 percent of the system's peak performance capability. Oak Ridge National Laboratory (ORNL) then invited Wang and other Gordon Bell finalists to carry out runs on ORNL's petaflop supercomputer, Jaguar, where the LS3DF application ultimately achieved a speed of 442 teraflop/s on the Cray XT5 system with 147,146 cores. In Wang's case, the winning simulation was achieved after only two runs over a two-day period, demonstrating the ease of porting and running high-performance applications on the Cray XT architecture. The project had previously been awarded time on Jaguar under DOE's Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program.

"This kind of multidisciplinary project that combines the most powerful algorithms and machines to produce scientific breakthroughs highlights the value of DOE's research and facilities program in scientific computing," says Kathy Yelick, NERSC Division Director.



A test run of LS3DF, which took one hour on 17,000 processors of Franklin, performed electronic structure calculations for a 3500-atom ZnTeO alloy. Isosurface plots (yellow) show the electron wavefunction squares for the bottom of the conduction band (left) and the top of the oxygen-induced band (right). The small grey dots are Zn atoms, the blue dots are Te atoms, and the red dots are oxygen atoms.

Humanities and High Performance Computers *continued from page 1*

us gain a better understanding of our world and history," says Katherine Yelick, NERSC Division Director.

"Supercomputers have been a vital tool for science, contributing to numerous breakthroughs and discoveries. I look forward to seeing how the same technology can further disciplines within the humanities," says Brett Bobley of NEH's Office of Digital Humanities.

Three projects have been selected to participate in the program's inaugural run.

The Perseus Digital Library Project, led by Gregory Crane of Tufts University in Medford, Mass., will use NERSC systems to measure how the meanings of words in Latin and Greek have changed over their lifetimes, and compare classic Greek and Latin texts with literary works written in the past 2,000 years. Team members say the work will be similar to methods currently used to detect plagiarism. The technology will analyze the linguistic structure of classical texts and reveal modern pieces of literature, written or translated into English, which may have been influenced by the classics.

"High performance computing really allows us to ask questions on a scale that we haven't been able to ask before. We'll be able to track changes in Greek from the time of Homer to the Middle Ages. We'll be able to compare the 17th century works of John Milton to those of Vergil, which were written around the turn of the millennium, and try to automatically find those places where *Paradise Lost* is alluding to the *Aeneid*, even though one is written in English and the other in Latin," says David Bamman, a senior researcher in computational linguistics with the Perseus Project.

According to Bamman, the basic methods for creating such a literary analysis tool have existed for some time, but the capability for analyzing such a huge collection of texts couldn't be fully developed due to a lack of compute power. He notes that the collaboration with DOE and NERSC eliminates that roadblock.

In addition to tracking changes in ancient literature, NERSC computers will also be reconstructing ancient artifacts and architecture with the *High Performance Computing for Processing and Analysis of Digitized 3-D Models of Cultural Heritage* project, led by David Koller, Assistant Director of the University of Virginia's Institute for Advanced Technology in the Humanities (IATH) in Charlottesville, Va.

Over the past decade, Koller has traveled to numerous museums and cultural heritage sites around the world, taking three-dimensional scans of historical buildings and objects — recording details down to a quarter of a millimeter.

According to Koller, a three-dimensional scan of the Renaissance statue David carved by Michelangelo contains billions of raw data points. To convert this raw data into a finished 3D model is extremely time consuming, and nearly impossible on a desktop computer. Limited compute power has also limited Koller's ability to efficiently recreate large historical sites, like Roman ruins in Italy or Colonial Williamsburg in Virginia. He hopes to use the NERSC resources to digitally restore these sites in three-dimensional images for analysis.

"The collaboration with NERSC opens a wealth of resources that is unprecedented in the humanities," says Koller.

Over the years, Koller has also digitally scanned thousands of fragments that chipped off ancient works of art, some dating back to the ancient Greek and Roman empires. Koller will to use NERSC computers to put these broken works back together again like a digital 3D jigsaw puzzle.

"The collaboration with NERSC opens a wealth of resources that is unprecedented in the humanities," says Koller. "For years, science reaped the benefits of using supercomputers to visualize complex concepts like combustion. Humanists, on the other hand, didn't realize that supercomputers could potentially meet their needs too, until NEH and DOE proposed collaboration last year.… I am really excited to see what comes out of this partnership."

In contrast to the other Humanities High Performance Computing projects that will be done at NERSC, the *Visualizing Patterns in Databases of Cultural Images and Video* project, led by Lev Manovich, Director of the Software Studies Initiative at the University of California, San Diego, is not focused on working with a single data set. Instead, this project aims to investigate the full potential of cultural analytics using different types of data, including millions of images, paintings, professional photography, graphic design, and user-generated photos; as well as tens of thousands of videos, feature films, animation, manchinema, Anime Music Videos, and user-generated videos.

"Digitization of media collections, the development of Web 2.0 and the rapid growth of social media have created unique opportunities to studying social and cultural processes in new ways. For the first time in human history, we have access to unprecedented amounts of data about people's cultural behavior and preferences as well as cultural assets in digital form," says Manovich.

For approximately three years, Manovich has been developing a broad framework for this research that he calls Cultural Analytics. The framework uses interactive visualization, data mining, and statistical data analysis for research, teaching and presentation of cultural artifacts, processes, and flows. Manovich's lab is focusing on analysis and visualization of large sets of visual and spatial media: art, photography, video, cinema, computer games, space design, architecture, graphic and web design,and product design. Another focus is on using the wealth of cultural information available on the web to construct detailed interactive spatio-temporal maps of contemporary global cultural patterns.

"I am very exited about his award to use NERSC resources. This opportunity allows us to undertake quantitative analysis of massive amounts of visual data," says Manovich. "We plan to process all images and video selected for our study using a number of algorithms to extract image features and structure; then we will use a variety of statistical techniques — including multivariate statistics methods such as factor analysis, cluster analysis, and multidimensional scaling — to analyze this new metadata; finally, we will use the results of our statistical analysis and the original data sets to produce a number of highly detailed visualizations to reveal the new patterns in our data."

DOE Announces 2009 NERSC Allocations

DOE has allocated 146.5 million Cray XT4-based hours to 334 NERSC projects, and 14 million hours remain in reserves. The five largest computational time awards were made to:

- Warren Washington of the National Center for Atmospheric Research: 12 million hours for the project “Climate Change Simulations using the Community Climate System Model”
- Doug Toussaint of the University of Arizona: 10.09 million hours for the project “Quantum Chromodynamics with Three Flavors of Dynamical Quarks”
- Donald Sinclair of the Argonne National Laboratory: 6.55 mil;lion hours for the project “Lattice Gauge Theory Simulations”
- Martin Savage of the University of Washington: 5.4 million hours for the project “Hadron-Hadron Interactions with Lattice Quantum Chromodynamics”
- Paul Kent of Oak Ridge National Laboratory: 4 million hours for the project “Computational Resources for the Nanomaterials Theory Institute at the Center for Nanophase Materials Sciences”

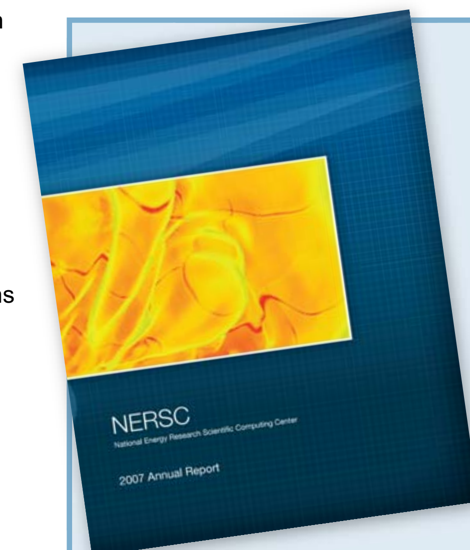
On the mass storage side, 40 million Storage Resource Units have been allocated to 359 NERSC projects, and 10 million remain in reserve. The three largest storage awards were made to:

- Graznya Odyniec of the Lawrence Berkeley National Laboratory: 5 million Storage Resource Units for the project "STAR Detector Simulations and Data Analysis"
- Stuart Freedman of the Lawrence Berkeley National Laboratory: 4 million Storage Resource Units for the project "Data Analysis and Simulations for KamLAND"
- John Bell of the Lawrence Berkeley National Laboratory: 3.5 million Storage Resource Units for the project "Applied Partial Differential Equations Center"

Five projects each received the fourth largest storage award, for 2 million Storage Resource Units:

- Curtis Covey of the Lawrence Livermore National laboratory for the project "Program for Climate Model Diagnosis and Intercomparison"
- Peter Jacobs of the Lawrence Berkeley National Laboratory for the project "Data analysis and Simulations for the ALICE Experiment at the Large Hadron Collider"
- Keh-Fei Liu of the University of Kentucky for the project "Lattice Quantum Chromodynamics Monte Carlo Calculation of Hadron Structure and Spectroscopy"
- Edward Rubin of the Joint Genome Institute for the project "Optimizing Genomic Data Storage for Wide Accessibility"
- Warren Washington of the National Center for Atmospheric Research for the project "Climate Change Simulations using the Community Climate System Model"

For more information about allocations, visit: http://www.nersc.gov/nusers/accounts/allocations/awards/



Photosynthesis, climate, combustion, supernovas, and much more! Read all about NERSC's contribution to these scientific disciplines in the 2007 NERSC Annual Report, which can be downloaded at http://www.nersc.gov/news/annual_reports/annrep07/annrep07.pdf.

Howard Walter Is Appointed NERSC Division Deputy

Second in Command: NERSC's new division deputy Howard Walter

Howard Walter, who joined NERSC in 1999 and has served as head of the Systems Department since 2005, has been named NERSC Division Deputy by Kathy Yelick.

“With his extensive experience in NERSC facilities and systems, Howard will be a great asset to me as the division deputy,” Yelick said. “Within NERSC, Howard will help coordinate across multiple groups and departments within the division. He will also be responsible for some LBNL-wide projects and for ongoing operational issues such as safety in the workplace.”

As head of the Systems Department since November 2005, Howard has overseen the Computational Systems Group, the Computer Operations and ESnet Support Group, the Mass Storage Group, the Data Systems Group, and the Networking, Security and Servers Group.

Walter, who came to LBNL from NASA Ames, was initially responsible for shepherding through the planning and development of the new machine room at the Oakland Scientific Facility.

Bill Kramer Moves to NCSA

Bill Kramer

After almost 13 years at NERSC, Bill Kramer will be leaving his post as general manager to undertake a new position as Deputy Project Director for Blue Waters Project at the National Center for Supercomputing Applications (NCSA) in Urbana, Ill.

"LBNL and NERSC are very special. I have been at both longer by far than any place I have worked because of the mission to impact diverse science, the fantastic staff and commitment to the highest quality systems and services," says Kramer. "What I will miss most are the NERSC people. People make NERSC work, and I was fortunate to work with innovative and highly dedicated people."

During his tenure at the Berkeley Lab, Kramer saw NERSC through many major transitions, including a move from the Lawrence Livermore National Laboratory to Berkeley; a migration of the entire user community from vector supercomputers to highly parallel computing; and the design and implementation of both the NERSC system architecture and the NERSC service architecture. This past year, Kramer played an integral role in managing the hardware upgrade of NERSC's Cray XT4 system, called Franklin, to quad-core processors, and setting up the procurement process for the NERSC-6 system, the next major supercomputer acquisition to support the Department of Energy Office of Science's computational challenges.

"In his time at NERSC, Bill has successfully stood up some of the world's fastest machines and established the standard by which production computing centers are run," says Kathy Yelick, NERSC Division Director. "As I transitioned into the role of NERSC Director this past year, Bill's wealth of knowledge and experience was invaluable to me."

INCITE Projects

continued from page 1

Here are descriptions of the seven INCITE projects awarded computing time at NERSC:

- John Bell from Berkeley Lab was awarded 3 million processor-hours to continue his research into combustion chemistry. In particular, Bell's project focuses on the interaction of turbulence and chemistry in lean premixed laboratory flames. This research can lead to more efficient and cleaner burning combustion systems, such as those in power plants.
- Warren Mori from the University of California at Los Angeles was awarded 4.6 million processor-hours to develop simulations to answer questions about plasma-based particle accelerators that currently cannot be answered through experiments. New acceleration techniques using lasers and plasmas could lead to ultra-compact accelerators for applications in science, industry and medicine.
- Chuang Ren from the University of Rochester received 1 million processor-hours at NERSC and 1.5 million hours at Argonne National Laboratory to carry out large-scale particle-in-cell simulations of the ignition phase in fast ignition, one of the most promising new methodss for improving the viability of inertial confinement fusion as a practical energy source. The project will help make fusion energy an environmentally friendly and safe option.
- Ji Qiang from Berkeley Lab received 800,000 processor-hours to optimize the design and improvement of beam delivery systems for the next-generation X-ray free electron lasers (FELs), which have excellent applications in physics, material science, chemical science and bioscience.
- Leeor Kronik from the Weizmann Institute of Science in Isarel was awarded 810,000 processor-hours to study the strucures of novel electronic materials. The results from the research will help clarify pressing issues in figuring out the electronic structure of organic/inorganic interfaces with applications in areas such as semiconductors.
- Paul Bonoli of the MIT Plasma Science and Fusion Center received 5 million processor hours to simulate how particles behave as they are driven by electromagnetic waves in fusion reactors. Understanding the movement of particles in fusion reactors, in which plasmas will be heated to 100 million degrees Celsius, will be critical for designing working fusion reactors as future energy sources.
- James Freericks of Georgetown University was awarded 2,250,000 processor-hours for simulating the behavior of materials using a new method known as pump-probe time-resolved photoemission. By bombarding materials with intense pulses of light, the material can reach a non-equilibrium state and allow researchers to gain new insight into the nature of properties.

WHAT IS NERSC NEWS?

NERSC News publishes every other month and highlights the cutting-edge research performed using the National Energy Research Scientific Computing Center, the flagship supercomputer facility for DOE's Office of Science. NERSC News editor Linda Vu can be reached at 510 945-2402 or LVu@lbl.gov. Find previous NERSC News articles at http://www.nersc.gov/news/nerscnews.