NERSC NEWS

NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

MARCH 2008

Nano-Behavior

SCIENTISTS USE SIMULATIONS TO STUDY DNA INTERACTIONS WITH PROTEINS, POLYMER

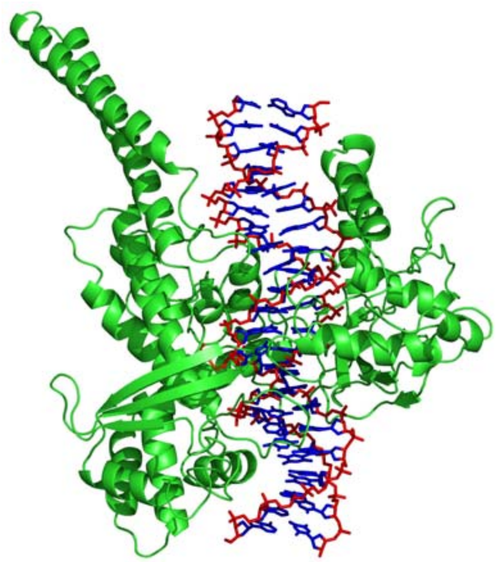
A team from the University of California at Irvine is exploring the complex interactions of human DNA with proteins and nanoparticles in a research that tackles key problems in designing and delivering anti-cancer drugs and gene therapy.

The research team, headed by Ioan Andricioaei, is using the computer power and services at NERSC to better understand how certain proteins affect DNA's behavior during cell replication. The research also looks at the dynamics between DNA and polyamidoamine (PAMAM) dendrimers, nano-particles that could be used to transport drugs and genetic materials into the body for treatment.

The research builds on previous molecular dynamic simulations by the team. The 2008 allocation of 7.4 million MPP hours represents the first major undertaking at NERSC by the Irvine research team.

"Our simulations require thousands of processors to simulate the vast conformational changes we are attempting to study," says Jeff Wereszczynski of the Andricioaei team. "With the power of Franklin (Cray XT4), what used to take us

continued on page 2



The image shows the relationship between Topoisomerase and DNA. Topoisomerase (in green) is an enzyme that unknots the DNA, allowing proteins to transcribe and form a new set of genetic materials during cell division.

Green Computing

BERKELEY LAB RESEARCHERS EXPLORE ENERGY EFFICIENT COMPUTING, FROM SYSTEM DESIGNS TO APPLICATIONS

NERSC and other Berkeley Lab researchers are taking on energy efficiency research that aims to influence the computing industry in designing and building computer and storage technologies that will benefit the scientists, consumers and the environment.

Thanks to the Laboratory Directed Research and Development Program (LDRD) at Berkeley Lab, researchers are exploring subjects in computer architectures, algorithms and mass storage system designs to improve energy efficiency of scientific computations. The LDRD program provides special funding for promising research projects, and this set of related LDRD projects includes researchers from NERSC and the Computational Research Division (CRD) at Berkeley Lab.

Pursuing energy efficient computing research at NERSC makes sense, both to save energy and to shift NERSC resources away from energy costs and towards systems and services that directly benefit the

continued on page 3

Sharing Expertise

NERSC, SWISS NATIONAL COMPUTING CENTRE SET UP STAFF EXCHANGE PROGRAM

NERSC and the Swiss National Computing Centre (CSCS) have 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

continued on page 7

Making Impact on Science

NERSC'S 2008 ALLOCATIONS REFLECT ITS ABILITY TO SERVE A BROAD RANGE OF SCIENCE

With its mission to support both high-impact and broad-impact science, NERSC is providing DOE 678.51 million MPP hours on its four supercomputers in 2008 for research in disciplines that include climate, material science, astrophysics, life sciences, computer science and combustion.

NERSC can offer more computing resources this year thanks to the new Cray XT4, named Franklin after Benjamin Franklin. With three other supercomputers available to serve its 3,000 users, NERSC continues to demonstrate its ability to support a wide range of projects in scopes and sizes.

Researchers have the flexibility to use any of the four supercomputers. Instead of allocating raw CPU hours, NERSC allocates MPP (massively parallel processing) hours that take into account the different computing power among the systems.

The charts and graphs below show the breadth of sciences being done at NERSC. The INCITE (Innovative and Novel Computational Impact on Theory and Experiment) allocations support computationally intensive projects that are expected to produce high-impact sci-

continued on page 5

DNA

continued from page 1

several months we can now do in weeks."

In the first project, the scientists want to know how an enzyme called topoisomerase I, which loosens the tightly packed DNA during replication, would react to topotecan. Topotecan is a chemical compound known to disrupt the enzyme's work, and is currently used for treating ovarian and lung cancer. But exactly how topoisomerase I and DNA interact with topotecan isn't clear. Understanding their dynamics could lead to better-designed cancer treatments.

The research conducted by Wereszczynski looks specifically at the energy use and movement of DNA as supercoils are relaxed, a vital step to allow cell division, and at how topotecan affects this function. Previous simulation work by the team already showed a difference in the mechanisms at which twisted DNA strands relax without interference from topotecan, so that proteins can access genetic information and jumpstart the replication process (Sari and Andricioaei, Nucleic Acids Research, 2005).

In the second project, Andricioaei and Mills aim to figure out how DNA behaves when it's attached to a nano-particle such as PAMAM dendrimer. Does the pairing alter the structure or properties of both objects, for example? Does it matter what part of the PAMAM dendrimer surface is used?

PAMAM dendrimers are tiny, and a variety of molecules can be easily attached to their surface. For this reason, PAMAM dendrimers are considered promising vehicles for delivering healthy DNA to fix faulty genes.

Based on their previous work, the scientists knew that both the DNA and the dendrimer would become deformed if the widest edge of a PAMAM dendrimer is parallel to the axis of a DNA. The DNA would wrap around the dendrimer while the dendrimer would flatten against the DNA. Understanding the energetics and dynamics behind this interaction is key in advancing the nanotechnology uinderlying gene therapies further.

In the third project, the researchers seek to understand the role of portal protein in the life of a bacteriophage, which infects a bacterium by pushing its genetic material into the host's cells. Since the early 1900s, scientists have discovered that some viruses can act as parasites and kill their bacterial hosts. Since antibiotics are easy to manufacture and store, the idea of using bacteriophages to treat illnesses didn't catch on, except in the former Soviet Union and Eastern Europe.

Understanding how bacteriophages work can lead to better medicines, including countering the effects of bioweapons.

What happens after a bacteriophage injects its DNA into a bacterial cell is the focus of the simulation performed by Jeremiah Nummela in the Andricioaei group. A bacteriophage consists of a protein shell enclosing its genetic material. When the viral DNA hijacks the protein synthesis mechanisms of the bacterium and begins to replicate itself, the process also involves the forming of capsids. Capsids are protein shells that would house these new DNA strands and become new bacteriophages.

The DNA strands enter the new capsids through the portal proteins. Importing the genetic materials through the portal proteins and packing them tightly requires a lot of force. Through simulations, the researchers intend to calculate the interaction between the DNA and portal protein. Andricioaei and Nummela will use a new method they have developed to carry out the calculations (Nummela and Andricioaei, Biophysical Journal, 2007).

Learn more about the research at Andricioaei's web site.

SPOTLIGHT

SCIENTIFIC COMPUTING ARTICLE

David Skinner

An article on the role and successes of the SciDAC Outreach Center, which is run by NERSC researcher David Skinner, appeared in a recent issue of Scientific Computing magazine.

SciDAC (Scientific Discovery through Advanced Computing) is a DOE Office of Science program that supports the development of software tools for tackling research using terascale computers. The Outreach Center, launched last year, serves as a clearinghouse for software tools and information for scientists who use or conduct research on high-performance systems. It also facilitates exchanges between various SciDAC research centers and the rest of the scientific community.

For example, the Outreach Center has helped the Applied Partial Differential Equation Center (APDEC) to promote and distribute its software. The tools developed by APDEC are being used in areas such as fusion energy, combustion, astrophysics and oceanography.

In another instance, the Outreach Center worked with a database equipment company that wanted to know more about how their technology could benefit the research community. The Outreach Center connected the company with scientists who might be interested in using the technology.

"Whether it is bringing new talent to HPC (high performance computing), delivering HPC resources to new audiences, or incorporating new technologies into DOE's HPC portfolio, the SciDAC Outreach Center is there as a broker between resources and needs to help get computing done," wrote Skinner in the article, titled "Reaching Out to the Next Generation of HPC Users." Skinner also heads the Open Software and Programming Group at NERSC.

Read the article to learn more about the Outreach Center.

Energy Efficiency *continued from page 1*

scientific community. With over 3,000 users and an insatiable demand for NERSC computing and storage facilities, results from the LDRD projects could indirectly benefit scientists across scientific disciplines that rely on computing. This includes cosmology, climate, life sciences, accelerator physics, fusion, computer science and material science.

Kathy Yelick, the NERSC Director, explains the importance of this work. "Power is the most important problem in computing today, not just at the high end, but from hand-held devices and laptops to data centers and computing centers like NERSC. Power density within chips has forced the entire processor industry to put multiple cores on a chip, and within centers the total system power is a major component of cost and availability."

She describes this set of LDRDs as a "multi-faceted attack" on the problem, starting with a blank slate on the architecture end, and rethinking algorithms, applications and software to make use of energy-efficient hardware. The first goal is to do more science with less energy, and the second is to enable the next generation of exascale computing systems, which require technological breakthroughs to address the power issues at such extreme scales.

One of the projects takes a vertical slice through the problem space, looking at a single application domain and considering alternative algorithms as well as architectures for solving the problem. Climate modeling is the target application, selected because of its significance to science and the general public, and because it requires millions of CPU computing hours to explore various climate scenarios and the possible impacts of changes in policy or alternate fuel sources. The other LDRD projects take a broader look at specific aspects of the problem, including energy-efficient computing components based on multicore technology, energy efficient storage systems, and application characterization that explores the ability of various key algorithms to adapt to energy-efficient hardware.

Here are the four LDRD projects:

Climate Modeling System

The development of multicore chips has been the computer industry's solution to keep power consumption in check. But some of the current approaches to adding more complex cores per chip would eventually hit a performance plateau. John Shalf, Lenny Oliker and Michael Wehner are investigating an alternative to using conventional microprocessors in designing energy-efficient supercomputers that employ more aggressive use of parallelism and design techniques from the consumer electronics industry to more closely tailor the chip design to the needs of scientific applications. They anticipate this approach could achieve 100 times or more improvement in power efficiency and effective performance over business as usual.

Their Climate Simulator Project will build a prototype system using embedded processors — low-power chips commonly found in consumer electronics devices such as cell phones and portable music players — and tailor its performance to provide optimal power efficiency for climate modeling problems. The group has adopted battery-powered designs in consumer electronics that are very sensitive to power consumption and cost. These embedded chips are less powerful and less power hungry than conventional microprocessors in supercomputers these days, and can be more easily customized to run specific applications. A computer built with thousands of these embedded chips could extract the most energy-efficient performance that also is capable of tackling complex scientific problems. The approach not only promises to be more power-efficient than the conventional path forward in HPC — it also promises to be more cost-effective.

The Climate Simulator team is working with Tensilica, an embedded processor design firm, as well as David Randall, a professor in the Atmospheric Sciences Department at Colorado State University and a NERSC user. Randall's climate modeling code, developed under DOE's SciDAC program, is a new breed of climate modeling code that is capable of expressing enough parallelism to run kilometer-scale simulations 1000 times faster than real time on machines envisioned by the Climate Simulator research team. In order to enable dramatic changes in power efficiency, codes such as Randall's must expose orders of magnitude more parallelism than the current climate models. Employing simpler processors that are designed for parallel throughput rather than serial performance will enable substantial power efficiency gains.

"We want to find compelling solutions to scientific problems that need peta-scale machines," Shalf said. "The use of these power-efficient cores will help us achieve those goals."

Manycore Chips

The development of multicore chips represents the most significant shift in microprocessor engineering in several decades, and it opens up opportunities for exploring innovative designs for high-performance computers.

Jonathan Carter, head of the User Services Group at NERSC, is leading the 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.

Future supercomputers will likely be built with chips containing an increasing number of cores. Multicore chip designs vary, however. They include heterogeneous designs, such as the Cell processor, developed by IBM, Sony and Toshiba; graphics processing units (GPUs); and processors for the embedded market. There are also homogeneous designs, such as microprocessors by Intel and Advanced Micro Devices, the world's two largest chip makers. In many cases, multicore technologies offer higher absolute performance and more energy-efficient computation.

"This LDRD project provides a breadth of architecture coverage to our whole ultra-efficient research thrust. We want to identify candidate algorithms that map well to multicore technologies, and document the steps needed to re-engineer programs to take advantage of these architectures," Carter said. "In addition, perhaps there are design elements in multicore chips that we can influence to help design a better high-performance system."

continued on page 4

Energy Efficiency *continued from page 3*

Mass Storage

Led by CRD scientists Ekow Otoo and Doron Rotem, the "Energy Smart Disk-Based Mass Storage System" project sets out to investigate energy-efficient disk storage configurations that also provide quick access to massive amounts of data.

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. As the need for reliable long-term storage of data grows, so will the associated energy costs.

Otoo and Rotem have set out to explore new configurations that divide the disks into active and passive groups. The active group contains continuously spinning disks and acts as a cache for most frequently accessed data. The disks in the passive group would power down after a period of inactivity. Besides looking at optimal disk configurations and file placement algorithms, the researchers will also develop simulation models for analyzing energy use.

Benchmarking for Dwarfs

A project led by Erich Strohmaier proposes to develop a test bed 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 conducted with domain experts from CRD, NERSC, and UC Berkeley.

From desktop PCs to supercomputers, systems built with hundreds of cores or more are likely to hit the market as early as the beginning of the next decade. The scientific community and the computer industry need to figure out how to make efficient use of these more powerful machines. This will be especially important for high end users, as they will face systems with large differences in interconnect properties at different levels of the system architecture hierarchy.

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.

SPOTLIGHT

STORAGE KEYNOTE

Bill Kramer

NERSC General Manager Bill Kramer will be a keynote speaker at the Storage Networking World Conference next month in Florida, presenting the technology and services provided by NERSC for archiving scientific data. Computerworld, a technology magazine and website, is hosting the conference.

In his talk, titled "NERSC — Extreme Storage and Computation for Science," Kramer will discuss the storage, networking and computational requirements for the computer center. He will describe the current and future deployments of the NERSC Global Filesystem, a key component in the center's strategy to provide an integrated set of systems and services for handling petabytes of data in a highly parallel environment.

The conference features a diverse set of speakers, including Jim Swartz, Chief Information Officer at Sybase, and Laura Campbell, Associate Librarian for Strategic Initiatives at the Library of Congress.

The conference will take place in Orlando from April 7–10. Kramer will speak on April 9, and you can learn more by checking out the conference agenda.

SCIDAC OUTREACH WORKSHOP ON GENOMICS

The SciDAC Outreach Center, headed by NERSC's David Skinner, presented talks about high-performance computing tools in a workshop sponsored by the DOE's Office of Biological and Environmental Research (BER) and the Interagency Working Group on Metabolic Engineering.

The four-day workshop in February was called "Joint Genomics: GTL Awardee Workshop VI and Metabolic Engineering 2008 and USDA-DOE Plant Feedstock Genomics for Bioenergy Awardee Workshop 2008." DOE's Genomics:GTL research program aims to understand the abilities of plants and microbes that enable clean energy generation, toxic waste cleanup and other environmental benefits.

The SciDAC Outreach Center organized a half-day workshop for identifying computational resources that would benefit researchers in the new bioenergy research centers, announced last year by DOE Secretary of Energy Sameul Bodman and administered by BER. Berkeley Lab is one of the six organizations that will run the DOE Joint BioEnergy Institute (JBEI), which is one of the three bioenergy research centers and is expected to receive $125 million in funding over five years.

Skinner, head of NERSC's Open Software and Programming Group, spoke about "Grid Based Management and Parallel Molecular Dynamics Simulations." Kathy Yelick, director of NERSC, discussed "NERSC Systems and Services." Wes Bethel, head of the NERSC's Analytics Team and the SciDAC Visualization and Analytic Center for Enabling Technologies, presented "Analysis of 3D Gene Expression Data."

In addition, Nagiza Samatova from the SciDAC Scientific Data Management Center talked about "Scientific Data Management: Technologies and Applications." Ruth Pordes from the Open Science Grid (OSG) gave a talk about the OSG.

2008 Allocations *continued from page 1*

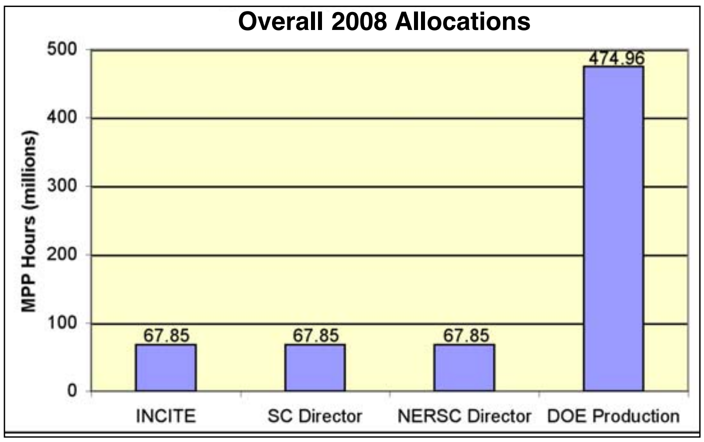
entific discoveries.

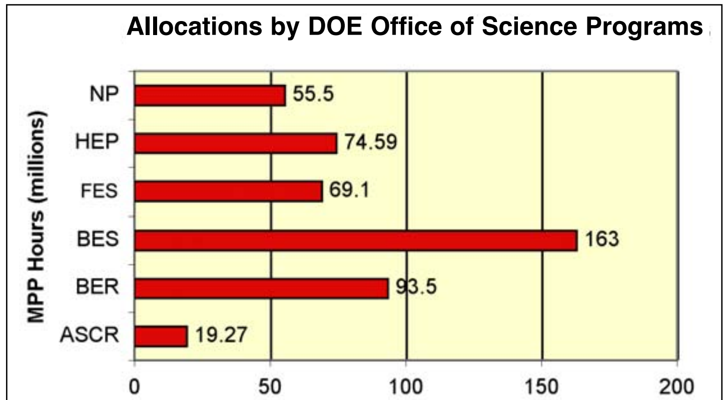
The term "DOE Production" refers to allocations made by each of the six program offices within the Office of Science. These allocations include resources for projects from the Scientific Discovery through Advanced Computing (SciDAC) program, which aims to develop software tools needed to advance discoveries on terascale computers.

Data storage is an integral component of NERSC's resources. NERSC staff perform regular assessments and upgrades of the High Performance Storage System (HPSS). Just last year, NERSC purchased new tape drives from Sun Microsystems. The new drives provide 2.5 times more capacity and four times the performance of the previous drives. Storage allocations are measured in SRUs (storage resource units).

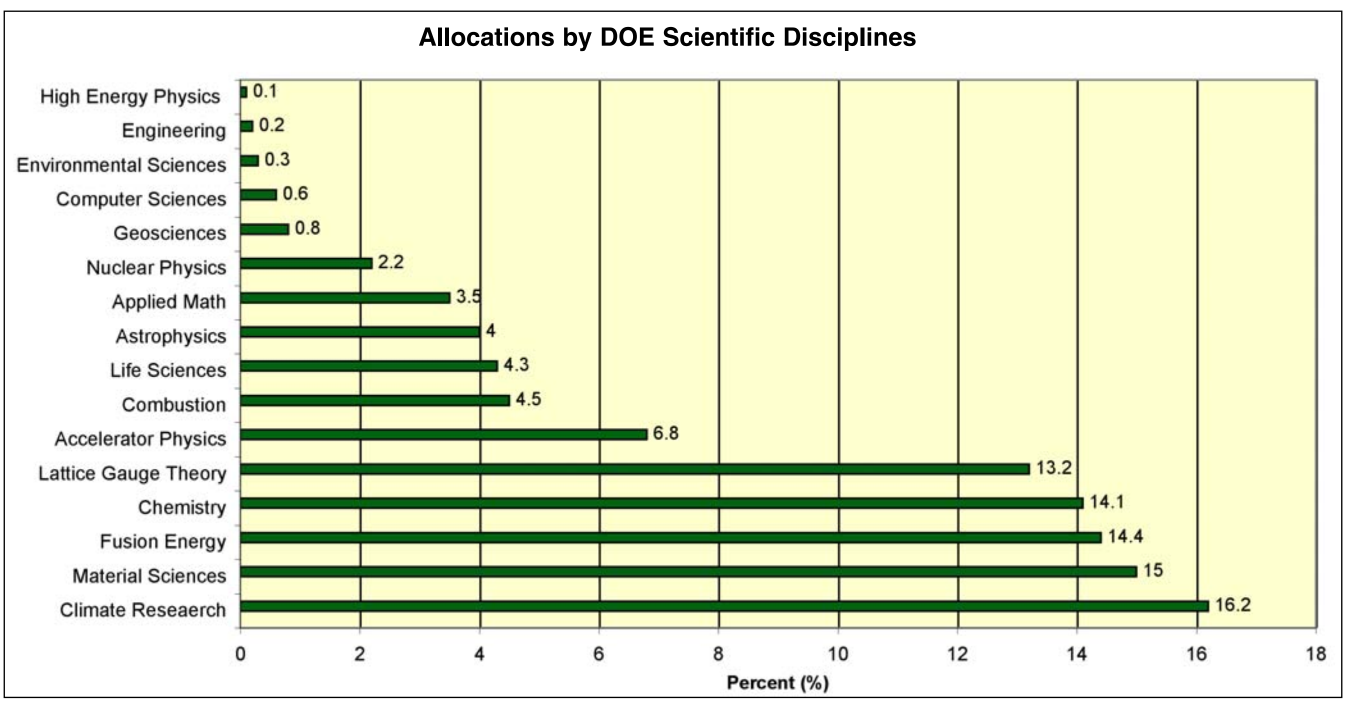
continued on page 6

Computing Hours Allocations (MPP)



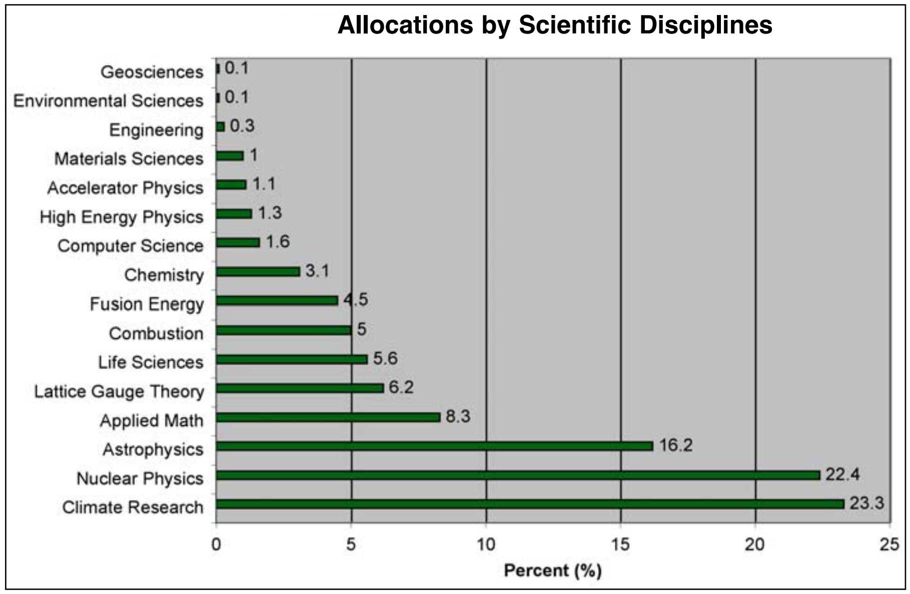


ASCR: Advanced Scientific Computing Research
BER: Biological and Environmental Research
BES: Basic Energy Sciences
FES: Fusion Energy Sciences
HEP: High Energy Physics
NP: Nuclear Physics

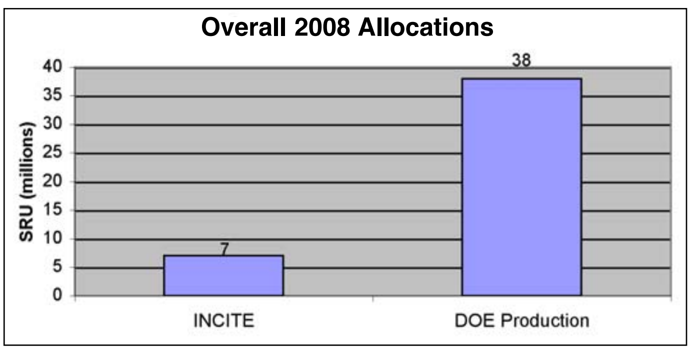


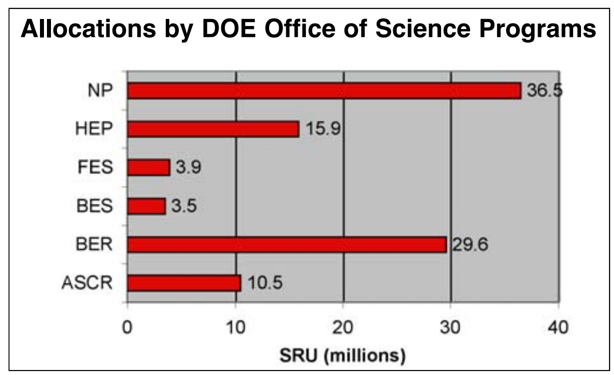
2008 Allocations

continued from page 5



Mass Storage Allocations (SRU)





Largest Allocations by Program Offices

Office of Science Program Offices	Principal Investigator	MPP Hours	Storage Resource Units	Project
ASCR	John Bell, Berkeley Lab	12,000,000	3,000,000	Applied Partial Differential Equations Center
BER (climate)	Warren Washington, National Center for Atmospheric Research	30,000,000	4,000,000	Climate Change Simulations with CCSM3: Moderate and High Resolution Studies
BER (SciDAC climate)	Paola Cessi, Scripss Institution of Oceanography	14,600,000	5,000	The Role of Eddies in the Meridional Overturning Circulation
BER (life sciences)	Valerie Daggett, University of Washington	10,000,000	50,000	Molecular Dynameomics
BES (chemistry)	Jackie Chen, Sandia Lab	13,000,000	400,000	Direct Numerical Simulations of Clean and Efficient Combustion with Alternative Fuels
BES (material sciences)	John Wilkins, Ohio State University	13,000,000	1,000	Modeling Dynamically and Spatially Complex Materials
FES	Wei-li Lee, Princeton Plasma Physics Lab	6,000,000	40,000	Turbulent Transport and Multicale Gyrokinetic Simulation
HEP (QCD)	Doug Toussaint, University of Arizona	30,000,000	50,000	Quantum Chromodynamics with Three Flavors of Dynamical Quarks
HEP (accelerator physics)	Cameron Geddes, Berkeley Lab	10,000,000	100,000	Particle Simulation of Laser Wakefield Particle Acceleration
NP (QCD)	Keh-Fei Liu, University of Kentucky	15,000,000	1,500,000	Lattice QCD Monte Carlo Calculation of Hadron Structure and Spectroscopy

Staff Exchange

continued from page 1

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.

Last year, a group from CSCS visited NERSC for a series of discussions about systems and facilities. Howard Walter, who oversees NERSC's computational systems, paid a return visit in January this year, sharing NERSC's expertise in designing and building energy-efficient computing facilities.

"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 DOE flagship facility for computational science, serving 3,000 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 Chief Operation Officer 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, which was signed in late January, the goal is "sharing and furthering the scientific and technical know-how of both institutions."

Cray Workshop

NERSC hosted a recent Cray Technical Workshop, which included presentations from NERSC staff and scientists who have used Cray supercomputers for their research.

The workshop, which took place in San Francisco last month, featured Bill Kramer, Zhengji Zhao and Katie Antypas from NERSC. Kramer, NERSC's General Manager, talked about the deployment of Franklin, the new Cray XT4 bought last year. Zhao, a consultant, discussed the performance of VAST on Franklin. VAST (Vienna Ab initio Simulation Package) is a plane wave code for quantum mechanical dynamics. Antypas, another consultant, spoke about I/O benchmarks.

The workshop also presented the perspectives of Cray users. Julian Borrill from Berkeley Lab's Computational Research Division spoke about his work on cosmic microwave background. Terry Ligocki, also from CRD, presented his work on adaptive mesh refinement scaling. Ji Qiang, from Berkeley Lab's Accelerator and Fusion Research Division, talked about his work on large-scale accelerator modeling.

Two other NERSC users presented their research at the workshop. Zhihong Lin, a professor at UC Irvine, discussed his fusion research. Nick Wright from the San Diego Supercomputer Center talked about performance modeling's role in ensuring the success of petascale computing.

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 Ucilia Wang can be reached at 510 495-2402 or Uwang@lbl.gov. Find previous NERSC News articles at http://www.nersc.gov/news/nerscnews.