LBNL-43131

# ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

## How Are We Doing?

## A Self-Assessment of the Quality of Services and Systems at NERSC (Oct. 1, 1997–Dec. 31, 1998)

William T. Kramer

**National Energy Research Scientific Computing Division**

Compiled May 1999

LBNL-43131

# How Are We Doing?

## A Self-Assessment of the Quality of Services and Systems at NERSC (Oct. 1, 1997 — Dec. 31, 1998)

Issued May 1999 by the
High Performance Computing Department
William T. Kramer, Department Head

National Energy Research Scientific Computing Center

Ernest Orlando Lawrence Berkeley National Laboratory
Berkeley, California 94720

This work was supported by the Department of Energy's Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information and Computational Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

# CONTENTS

## INTRODUCTION

In 1999, the National Energy Research Scientific Computing Center marks its 25th anniversary as one of the nation's pioneering scientific computing facilities. Since its early days as the Controlled Thermonuclear Research Computer Center when cycles were provided on borrowed computers, NERSC has evolved to serve a widening user base with state-of-the-art high-performance computers, mass storage systems, networking capabilities and intellectual resources. We believe that this last component — the intellectual resources of our diverse staff — is what differentiates NERSC from other centers (some of which are now emulating our model) by providing more than just computing resources. Another key element of NERSC's successful operation is its ability to anticipate and meet the varying needs of clients. In order to further this strong working relationship, NERSC staff and clients meet periodically via our national user group, ERSUG (the Energy Research Supercomputer Users Group) to share views, offer training and identify problems and solutions.

The decision to move NERSC from Livermore to Berkeley Lab in 1995 was aimed at changing the way the center does business, to enhance the facility's ability to accelerate the pace of scientific discovery. With the move to Berkeley, the name was changed to the National Energy Research Scientific Computing Center. This signaled the new approach to clients — one of enabling scientific computing, not just providing supercomputer cycles.

We have re-invented NERSC with a new architecture, a design that emphasizes both our high-performance computing systems and our commitment to excellent client service. Working in parallel, these two sides of our organization aim to give clients the necessary resources for conducting their research. Our job is to give our clients the reliable tools they need — client support, software and access to computing resources — and our success is measured in large part by the quality of science produced by our clients.

To ensure that we are meeting those needs, we have established a set of 10 performance goals pertaining to our systems and service. These goals were developed in consultation with our staff, our client community and our stakeholders. Within the NERSC organization, we framed our goals through both top-down and bottom-up processes. Each group within the organization came up with their own goals, and then the entire staff met to review, revise and refine the goals of each group, and the goals of the organization as a whole. Senior managers then outlined overall management goals and reviewed the groups' goals to ensure that they supported the management goals, and merged them into a single package.

With our agreed-upon goals in place, we can now gauge just how well we're doing in meeting expectations. We have tried to ensure that the goals reflect our efforts from a client's perspective, as opposed to an internal one. For example, a measurement of system availability needs to reflect the number of hours a machine is available to our clients, not how long it takes to identify a problem and initiate corrective action on our end. The goals we have set out cover the following areas:

- Reliable and Timely Service
- Innovative Assistance
- Timely and Accurate Information
- New Technologies

- Wise Technology Integration
- Progress Measurement
- High-Performance Computing Center Leadership
- Technology Transfer
- Staff Effectiveness
- Protected Infrastructure

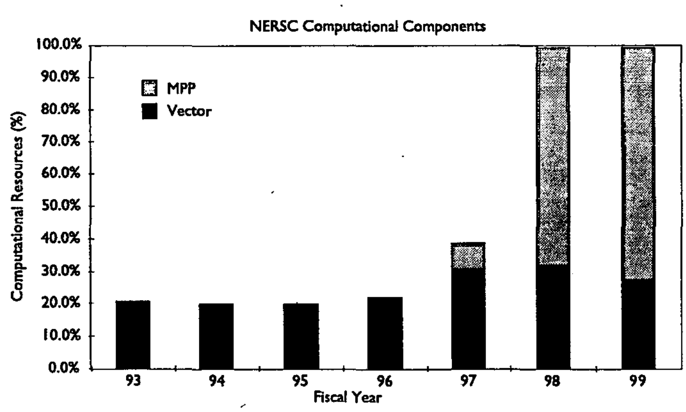
In the fall of 1998, NERSC conducted a survey of our clients to gain a better understanding of how they see us, our facilities and services. The results, based on responses by 138 clients, showed we're moving in the right direction. Overall, clients say the center provides a high level of service. On a scale of 1 (very unsatisfied) to 7 (very satisfied), scores ranged from 5.8 for consulting services, C90 uptime, and CFS reliability to 4.4 for C90 and T3E turnaround time. Many clients pointed out that NERSC is an excellent high-end production center, with an excellent client focus. More information regarding the survey can be found later in this report under the "Progress Measurement" section.

To give our clients, our sponsors and our own staff a better idea of how we're performing, we will produce annual performance reports, such as this one, which covers FY98. We believe it shows we're constantly improving our effectiveness, yet also seeking other ways to do even better. As with any organization, there is always room for improvement. Improvements, in turn, raise subsequent expectations. Given this situation, NERSC intends to constantly assess, improve and set new standards for the systems and services we provide to clients.

At the same time, we will continue to provide our clients with the greatest amount of computer time we can allocate, and we are constantly striving to improve our facilities and procedures to optimize the number of cycles we can provide. The following chart demonstrates our past and anticipated current allocation of cycles.

## A STATISTICAL SNAPSHOT OF NERSC CLIENTS

When talking about meeting the computational science needs of NERSC clients, we're covering some pretty broad territory, whether you measure it geographically, scientifically or organizationally. Here are some statistics on the NERSC user community.



### *By Type of Institution*

To start with, NERSC predominantly serves clients at DOE national laboratories, who account for 66 percent of the center's use. Universities account for 33 percent and industry use is about 1 percent. Just under 1 percent is the total usage by other federal facilities.

By calculating usage by clients at different types of institutions, the picture changes somewhat. DOE labs account for 57 percent of the total client base, and university clients account for 35 percent. Industrial clients total 7 percent, and other federal clients are 1 percent.

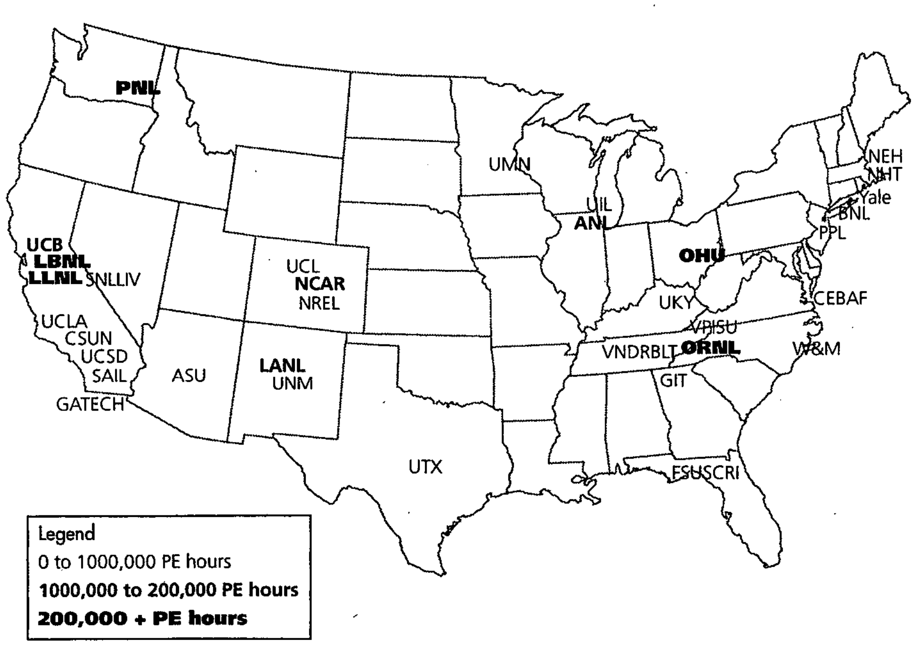
### *By National Laboratory*

NERSC provides computational resources to clients at more than 15 national laboratories. The largest number of lab clients are at Lawrence Livermore (24 percent), followed by Princeton Plasma Physics Lab (20 percent), then Oak Ridge (12 percent) and Los Alamos (10 percent) national labs. Lawrence Berkeley is the recipient of 9 percent of the accounts, and Pacific Northwest National Laboratory has 7 percent. Argonne, Brookhaven and the Stanford Linear Accelerator Center each have 3 percent of the accounts, and other DOE agencies and the Naval Research Laboratory each have 2 percent. The National Renewable Energy Lab, NASA, Thomas Jefferson Accelerator Facility, Ames Laboratory and the National Center for Atmospheric Research each have 1 percent.

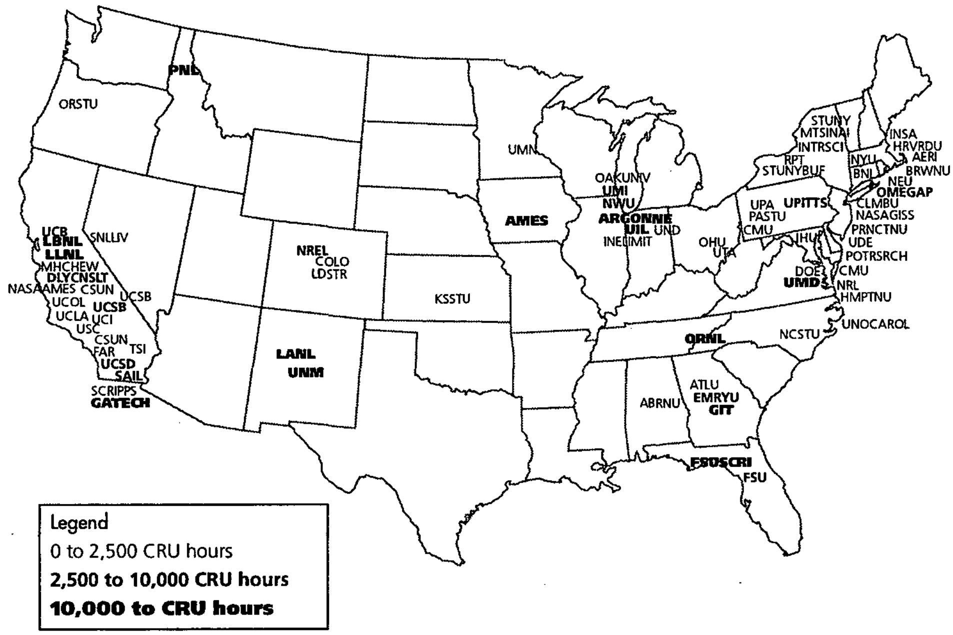
### *By Geographical Distribution*

These maps show the distribution of NERSC clients, categorized by size of allocation, across the nation. This first map illustrates allocation of time on the massively parallel Cray T3E-900, while the second map shows institutions which have allocations for using the parallel vector machines, the Cray J90SEs.

**MPP Allocations**



**PVP Allocations**




In short, NERSC is truly a national research facility, providing computing resources to clients at 242 research organizations in 32 states.

## RELIABLE AND TIMELY SERVICE

***Goal: Have all systems and support functions provide reliable and timely service to their clients.***

NERSC strives to provide reliable service to all of our clients. Our efforts address two general areas:

- how reliably our systems operate (i.e. availability to clients), and
- how responsive we are to clients when they have a problem.

To meet our goals, various components of the NERSC organization must work hand in hand. When a problem arises, we respond promptly to acknowledge, address and correct it. As described previously, NERSC provides clients with both the high-performance computing systems and the expert services for achieving research goals. To measure how well we're doing regarding hardware systems and related support, we have established metrics for system reliability and responsiveness to clients' problems.

The chart below shows various aspects of our systems' reliability. Figures in bold represent the measured time, while goals are shown in parentheses. Scheduled availability refers to the amount of time the systems are expected to be available (accounting for scheduled maintenance and upgrades), while overall availability refers to just that — the percentage of time a system is available overall. The "Mean Time to Repair" refers to the amount of time between a system failure and the point at which full service is restored to clients.

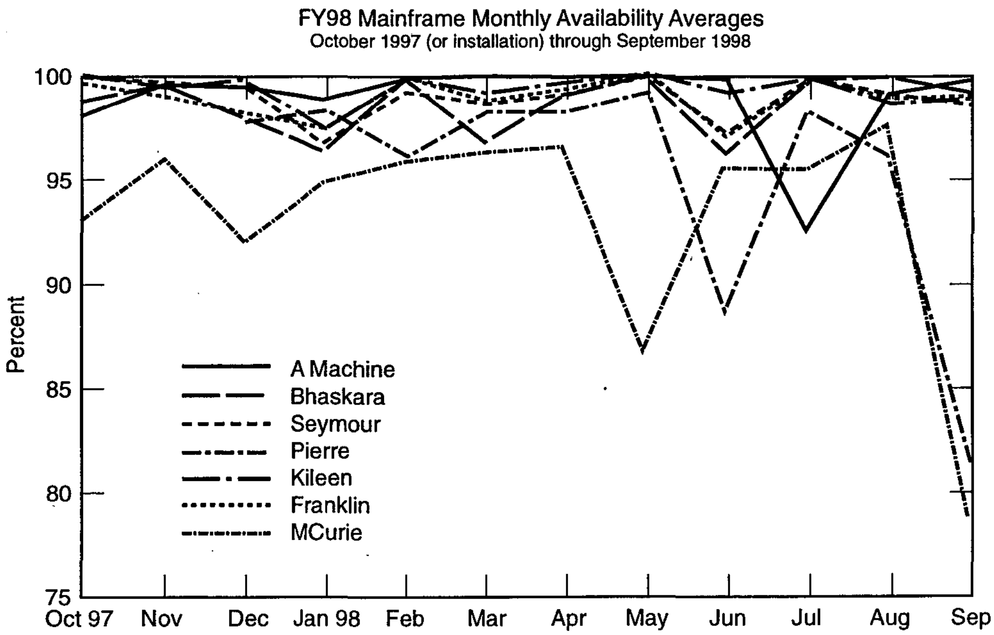
In most cases, our measured performance exceeds the goals. (Special note: The Mean Time to Repair for Servers is somewhat skewed as it represents the only outage of the year, which lasted for 11 hours.)

**SYSTEM METRICS FOR FY98**

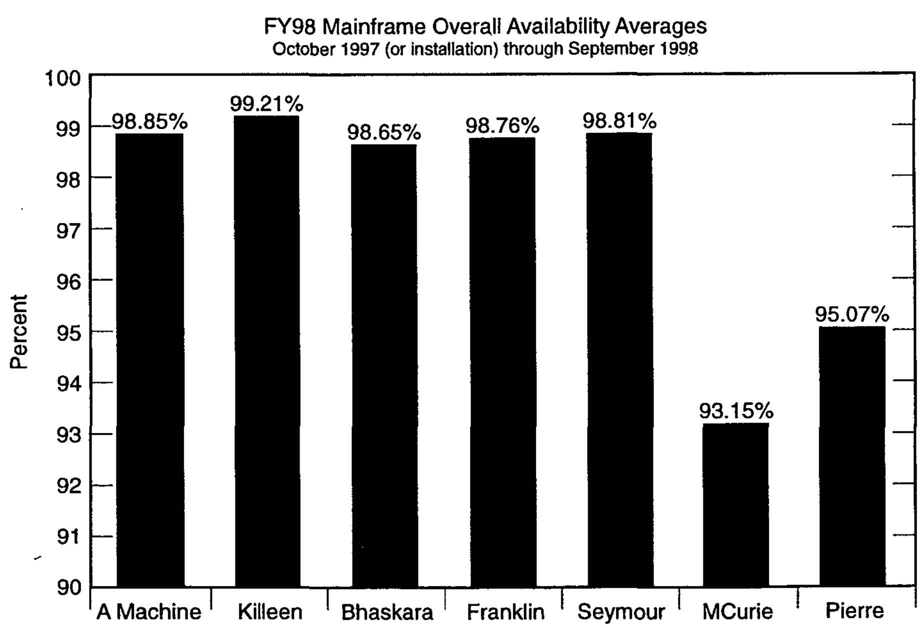
**Measured (Goal)**

| | % Availability | | Mean Time Between Interruptions | Mean Time to Repair |
|---|---|---|---|---|
| | Scheduled | Overall | (Hours) | (Hours) |
| <u>Systems</u> | | | | |
| Vector Systems | **99.52** (96) | **98.85** (95) | **308** (96) | **3.4** (4.0) |
| Storage Systems | **98.73** (96) | **97.97** (95) | **104** (96) | **2.1** (4.0) |
| Parallel Systems | **98.68** (90) | **93.82**(85) | **63** (96) | **3.9** (4.0) |
| <u>Workstations</u> | | | | |
| Servers (fs/gw) | **99.9**(96) | **100** (96) | **8,760** (340) | **11.0** (8.0) |
| Clusters | NA (96) | NA (95) | NA (40) | NA (8.0) |

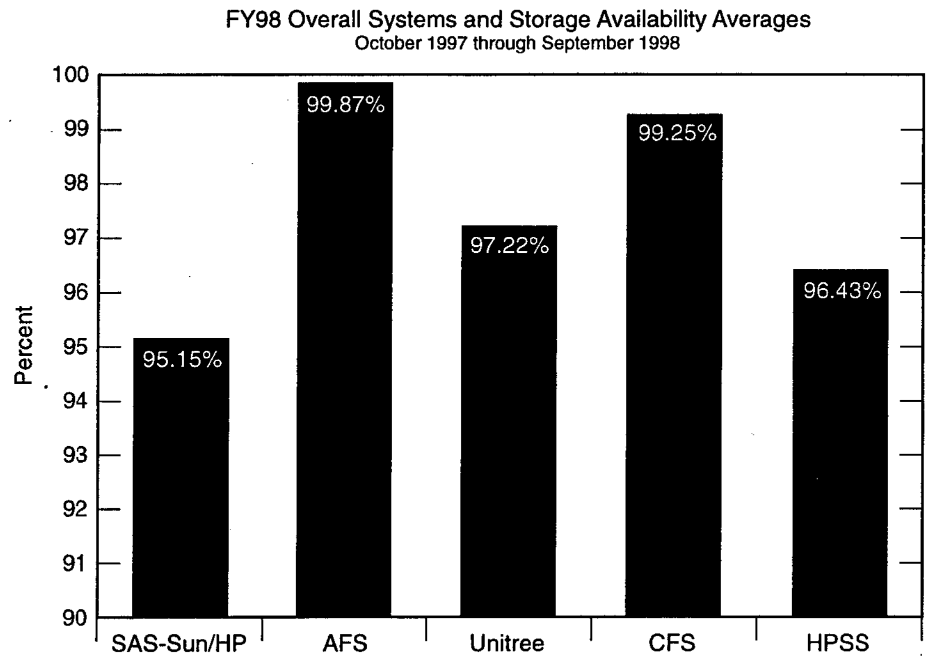
This graph shows the monthly availability of each of the NERSC mainframe computers for the year from September 1997 through September 1998. Availability of both Cray T3E computers (MCurie and Pierre) dropped in September as the two systems were reconfigured as a single 640-processor machine.




This graph shows overall average availability of NERSC mainframe computers for the year from October 1997 through September 1998.



This chart shows yearly availability of small computing systems and storage systems.




NERSC's service goals are to respond to clients' problems within four working hours and to resolve at least 90 percent of those problems within two working days. The times reported in this section refer to working days and do not include weekends or after-hours periods.

Here's how we're doing in terms of meeting those goals.

- Spot checks confirm that NERSC meets the goal of responding to problems within four hours.
- During calendar year 1998, a total of 3,138 "trouble tickets" were created. Of these, 188 took more than two days to close. This translates to a 94 percent closure within the two-day target.

Not all problems can be resolved within two days, and in some cases these problems are put "on hold" as longer-term solutions are needed. Reasons for putting a problem on hold include software requests, ongoing coding projects, "bugs" waiting for a vendor-supplied fix, and user not responding to a request for input within two days. These problems obviously take longer to resolve.

Problems not resolved (and a Not on Hold) within 72 working hours are escalated for more in-depth review. This escalation, done manually om FU 98, was automated in January 1999 and this process will help ensure that outstanding problems are addressed. NERSC staff also periodically review problems and client requests to ascertain areas needing attention with an eye toward fixing them with minimum disruptions in service.

# INNOVATIVE ASSISTANCE

***Goal: NERSC aims to provide its clients with new ideas, new techniques and new solutions to their scientific computing issues.***

NERSC is committed to helping its clients achieve better performance and results from their computational science efforts. Merely providing the computers, data storage and software is not enough to do this, so NERSC has initiated a range of activities to provide innovative assistance. From individual outreach to general tutorials, these programs demonstrate NERSC's commitment to maintaining our reputation as one of the world's leading scientific computing centers, as well as an integral part of research conducted by Department of Energy Office of Science programs. As a partner in more than half of DOE's Grand Challenges, NERSC is helping scientists blaze new trails on the frontiers of computational science. Here are some examples of our innovative service:

## Meeting the Grand Challenges

The Materials, Methods, Microstructure and Magnetism Grand Challenge team consists of scientists from NERSC, Oak Ridge and Brookhaven National Laboratories, and Ames Laboratory. This project, which has applications in fields ranging from computer data storage to power generation and utilization, aims to dramatically advance the computational study of materials, leading to a much better understanding of the relationship between the atomic-scale structure of materials and their macroscopic structural and electrical properties. Over the summer and fall of 1998, the group scaled up their code to produce increasingly useful results, with the assistance of Andrew Canning from the NERSC Scientific Computing Group.

At each step of the way, Andrew worked with both the researchers and system managers to obtain optimum performance and results. This effort represents a significant milestone because it both developed the theory to make such scalable simulations possible, and reached the scale of more than 1,000 atoms. (More information about this work can be found in the section High Performance Computing Leadership.)

The NERSC User Services Group provided workshop assistance to two other Grand Challenge projects:

Numerical Tokamak Turbulence Project Grand Challenge: More than a dozen fusion scientists from around the country held a workshop at NERSC May 5–8, 1998, to develop more powerful computational modeling tools. The National Transport Code workshop was aimed at modernizing the existing fusion modeling codes with a set of reusable software components that will give researchers greater flexibility in modeling particle transport problems. The codes were developed years ago and are greatly limited in the scope of such calculations. In addition to developing codes for running larger simulations, workshop participants also sought to add a graphical interface to their codes so they can view how well the code is running in real-time (the old codes don't allow this) and make changes as the code is running. The intended result is a more flexible and collaborative set of tools.

Computational Accelerator Physics: Advanced Modeling for Next Generation Accelerators Grand Challenge: NERSC cosponsored the International Computational Accelerator Physics Conference in Monterey on Sept. 14-18, 1998 and coordinated an afternoon tutorial on related computing subjects. Subjects included an Introduction to Parallel Computing, High Performance Computing at NERSC, Parallel Scientific Libraries and examples involving beam dynamics and electromagnetics.

**Experiments with Web-Based Lecture Presentations**

NERSC's first experimental attempt to record a talk for web-based distribution was a talk Mark Durst of NERSC User Services gave at the Mathematical Sciences Research Institute on Aug. 27, 1998. NERSC plans to build on this experience of recording (audio only, as well as audio and video), encoding, and providing such materials on the web, in a way that accounts for differing download and streaming bandwidth availabilities. NERSC will evaluate different technologies and see which best meet the needs of clients. The experience is improved by attaching HTML files, PowerPoint slides, or other web-compatible materials to the audio-visual aspects of the presentation, and figure out where the added value exists in the presentation of the technical information.

**Speeding Up Computational Science**

Jonathan Carter of the User Services Group worked with Berkeley Lab's Greg Scholes to tune and parallelize his program for determining the electron density of molecules. Although Scholes' code was an innovative method for doing this research, his program initially ran faster on an SGI Octane workstation than on the Cray C90. By fusing three loops to increase vector length and decrease loop overhead, Carter achieved a speedup from 37 Mflops to 97 Mflops.

To provide better support for climate modeling research, R. K. Owen of User Services rewrote and patched parts of *netcdf* to use T3E parallel I/O features needed by NERSC's climate research community. *netcdf* is software used for representing scientific data in a machine-independent way. User Services' Harsh Anand coordinated user requirements for *netcdf* and built test cases from clients' codes.

**Conversion from Fortran 77 to Fortran 90**

When Fortran 77 was discontinued as a programming language by SGI for J90s, NERSC announced this change months in advance and provided transition guides for NERSC clients. Members of the User Services Group individually called about 175 Fortran 77 clients to make sure they knew of the pending change — and to offer assistance in converting to Fortran 90.

**Meeting Users' Storage Needs**

The default software that comes with HPSS (High Performance Storage System) did not meet user needs in terms of flexibility and ease of use. Staff members from the Mass Storage (Nancy Meyer) and User Services (Majdi Baddourah, Jonathan Carter and Tom DeBoni) groups obtained HSI from Mike Gleicher (at the time at SDSC). HSI offers more concise commands, is more convenient and has a richer range of features. The utility was tested, documented and then made available to NERSC clients.

NERSC Division Director Horst Simon has noted that in the arena of high-performance computing, if you stand still you're really falling behind. Although this notion is usually applied to hardware systems, we at NERSC also believe it applies to our staff's services to clients. As a result, we're constantly seeking new ways to help clients achieve their scientific computing goals.

# TIMELY AND ACCURATE INFORMATION

***Goal: Provide timely and accurate information and notification of system changes to the client community so they can most effectively use the NERSC systems.***

The NERSC staff strives to provide our clients with timely and accurate information which may affect those clients' research efforts. Not only do we give adequate notice of changes and outages, we also try whenever possible to provide an explanation of the reasons behind the change and the expected impact on clients. As an example of timeliness:

- Planned system changes were announced at least seven days in advance (except in one instance); all planned system outages were announced at least 24 hours in advance.
- All system changes and planned outages were announced in advance on the NERSC "What's New" Web page. Some major changes were also announced by email to PIs. By combining web postings and email announcements, we are able to more quickly inform clients of changes, which in turn allows us to implement changes more quickly.

In 1998, the NERSC staff posted 481 notices to let clients know the latest status of NERSC systems.

## User Training

NERSC also organizes training programs to meet the needs of both new and experienced clients to help them make the most of their allocations. In addition to organizing training about changes in existing systems and new systems, NERSC staffers monitor trouble reports to identify areas in which additional training may be of benefit. Here are some examples of the various workshops, training sessions and talks given by NERSC employees and members of the NERSC user community:

Two sessions of "NERSC Training for New T3E Clients," a two-day series of lectures, were offered in February for Berkeley Lab clients and in April, in conjunction with the ERSUG meeting. Topics included an introduction to the T3E system, parallel programming, parallel I/O, and message passing interfaces.

During the ERSUG meeting on April 6–9, 1998, at Lawrence Berkeley National Laboratory, two days of training lectures were presented, including talks by NERSC staff and NERSC clients as part of the Users Helping Users (UHU) program. The focus of this event was efficient use of the NERSC T3E systems, and the talks covered advanced techniques and tools for such use.

UHU presentations at ERSUG:

- "Portable Lattice QCD Code on the T3E," Doug Toussaint, University of Arizona.
- "Global Arrays: A Portable Shared Memory Programming Environment for Massively Parallel Computers," Jarek Nieplocha, Pacific Northwest National Laboratory.
- "Parallelization of a 3D Plasma Fluid Turbulence Model on the NERSC Cray T3E at NERSC," Jean-Noel LeBoeuf, Oak Ridge National Laboratory.
- "Electron Structure Code," Jeff Tilson, Argonne National Laboratory.
- "HPF for Fortran Clients — Productivity Gains Using HPF on the T3E," Robert Ryne, Los Alamos National Laboratory.

NERSC presentations at ERSUG:

- "Application of the PVODE Solver to Parallelize the Fluid Transport Code UEDGE," Xueqiao Xu, Lawrence Livermore National Laboratory.
- "Overview of the ACTS Toolkit: A Set of Tools That Make It Easier to Write Parallel Programs," Bill Saphir, NERSC.
- "Scalapack: A Library for Parallel Dense Linear Algebra," Xiaoye "Sherry" Li, NERSC.
- "Aztec: A Library for the Parallel Solution of Sparse Linear Systems," Kesheng "John" Wu, NERSC.
- "Remote Visualization at NERSC," Stephen Lau, NERSC.
- "The TAU Portable Profiling and Tracing Package," Sameer Shende, University of Oregon.
- "Multigrid Solvers for Finite Element Matrices on Unstructured Grids with PETSc," Mark Adams, UC Berkeley.
- "Shmem and Synchronization Primitives on the Cray T3E," Adrian Wong, NERSC.
- "Parallel and Distributed Computing with PVM on the T3E," Youngbae Kim, NERSC.
- "T3E Individual Node Optimization," Mike Stewart, SGI/Cray Research.
- "T3E Multiprocessor Optimization and Debugging," Majdi Baddourah, NERSC.
- "I/O on the T3E," Richard Gerber, NERSC.
- "A Case Study in T3E Optimization," Jonathan Carter, NERSC.

# NEW TECHNOLOGIES, EQUIPMENT, SOFTWARE, AND METHODS

***Goal: Ensure that future high-performance technologies are available to Office of Science computational scientists.***

Since it was established in 1974, NERSC has provided its client community with the most up-to-date computing resources available. Since moving to Berkeley, two Cray T3E supercomputers (now merged into one system), a cluster of four Cray J90se computers and a high-performance storage system have been added to NERSC's equipment roster, giving the center one of the most powerful lineups of computing power in the country. As this report was prepared, NERSC was involved in the procurement progress for its next machine, referred to as NERSC-3. As new machines are introduced to the center, systems experts carefully analyze performance and work with manufacturers to ensure that the equipment meets the high-performance needs of NERSC clients.

However, implementing the newest technology also requires a careful evaluation in advance. The supercomputing field is littered with the names of companies promising the newest, fastest and best one year and then disappearing the next. Before NERSC adopts a new technology, the equipment must be fully evaluated and tested. (See accompanying section on Wise Technology Integration.)

To help do this, NERSC established the Advanced Systems Group in early 1998. The group's task is to investigate, and when feasible, help test and assess new technologies as part of NERSC's overall technology implementation effort. Of course, not every piece of equipment meets NERSC's needs, but by evaluating a range of production and prototype systems, as well as further developing those installed in our facility, NERSC is able to offer its clients systems which are leading-edge as well as robust.

## Building a Bigger, Better Cray T3E

To provide state-of-the-art capabilities in scalable parallel computing, NERSC has obtained two Cray T3E computers, one 512-processor machine and another with 96 processors. During the summer of 1998, Pierre, the smaller of NERSC's two Cray T3Es, received a number of hardware upgrades, making the machine a more useful tool for users. Previously, Pierre had 104 processors, each running at 300 megahertz. Those processors were all replaced with 450-megahertz processors, and another 48 of the faster processors were also added, so the machine's processors were as fast at those of MCurie, the other T3E. Sixteen additional 4.5 GB disks, available after an upgrade to MCurie, were installed on Pierre, along with a multi-purpose node (MPN) used by the machines I/O system. As part of the upgrade, users' home file space is now maintained by SGI's Data Migration Facility (DMF) and quotas were increased to 800 megabytes. The upgrade resulted in Pierre having 128 processors for running parallel applications. An accompanying increase in the size of the queues allowed the machine to run larger jobs for longer periods.

Next, the two Cray T3E supercomputers were successfully melded to create a unified, 640-processor machine (actually, 696 processors in all, but with 640 PEs for computing). The merger required a complete rewiring of the system, with a special team brought in from SGI/Cray to do the job, which took four days. Mike Welcome, the member of NERSC's Computational Systems

Group responsible for the T3E, handled the effort from the NERSC side. Also working on the project were on-site Cray employees Steve Luzmoor and Bryan Hardy.

The newly reconfigured machine is expected to reduce software costs, be less expensive to manage overall, and allow larger jobs to be run more frequently. The more powerful machine, which retained the name of the 512-processor machine, MCurie, went on line Oct. 4, 1998.

As presently configured, the T3E provides 512 processors dedicated to batch jobs, with another 128 processors dedicated to interactive parallel jobs by day and batch jobs by night. There are 15 processors for the operating system and the remainder are used for serial work.

Once it was merged, the T3E was tested using the standard LINPACK benchmark tests. In the 690-processor mode, the machine achieved 443 gigaflop/s, with a peak of 621 Gflop/s. Using the 640 computational processors, a sustained rate of 414 Gflop/s was recorded, with a peak of 576 Gflop. In January 1998, NERSC's 512-processor machine's sustained performance rate was 321 Gflop/s.

### Evaluating Advanced Systems

One of the first projects taken on by the Advanced Systems Group was evaluation of WildFire, a prototype shared-memory multiprocessor under development at Sun Microsystems. A goal of WildFire is to evaluate the effectiveness of leveraging large SMPs in the construction of even larger systems. (Note: "WildFire" is Sun's internal codename for an advanced server architecture that is under development. WildFire prototype systems have not been tested, optimized, or qualified for sale; they have been built for evaluative purposes only. Elements of the WildFire prototype may or may not be used in future Sun products. NERSC has obtained its WildFire through participation in the WildFire Beta/Collaborative Research program.)

### Evaluating the Application of Tera's Multi-Threaded Architecture

In 1998, NERSC signed an agreement with the San Diego Supercomputer Center to evaluate potential applications of a multi-threaded architecture Tera supercomputer installed at SDSC. The agreement gives NERSC 20 percent of computer time available to users of the two-processor system, which was accepted in April 1998, as well as 20 percent of the time available once the system was upgraded to four processors.

SDSC's Tera is the first machine delivered by Seattle-based Tera Computer Company. The computer uses a unique multi-threaded architecture (MTA) which is designed as a hybrid between shared-memory vector computers and distributed-memory parallel machines. The Tera design seeks to overcome the "latency" problem, which is the extra time required to send information between processors and memory on a parallel machine. This latency can prevent a parallel machine from achieving high performance.

Because of the unique design and its potential, SDSC's deployment of the Tera is being supported and evaluated by the National Science Foundation and the Defense Advanced Research Projects Agency. DOE is also interested in the architecture and is funding NERSC's contract with SDSC and NERSC's participation in the system evaluation.

**Realizing the Potential of PC Clusters for Scientific Computing**

For a number of years now, clusters of desktop computers have been heralded as the supercomputer of the future. Despite wide-ranging development programs, clusters continue to remain futuristic ideals. NERSC's Future Technologies Group is aggressively developing tools to realize the potential of small clusters, which are ideal for parallel code development, special-purpose applications, and small- to medium-sized problems. Large clusters show promise as alternatives to massively parallel computers for certain applications. The group's ultimate goal is to develop software for easy-to-use "plug-n-play" PC clusters that require little effort to set up and maintain, and which provide a full-featured environment similar to that of a traditional supercomputer. Part of this project includes participation in the Intel-supported Millennium project at UC Berkeley.

To develop and test the necessary software for cluster computing, NERSC has established a 36-node cluster. The cluster includes 32 single-processor 400 MHz Pentium II nodes and two 4-processor Pentium Pro nodes.

Development projects within the PC cluster project include:

- M-VIA, a high-performance modular implementation of VIA (Virtual Interface Architecture) for Linux. VIA is a new standard being promoted by Intel, Compaq and Microsoft that enables high-performance communication on clusters.
- Message Passing Interface on VIA. MPI is the critical VIA "app" for high performance computing.
- Other applications of VIA. We are looking at other applications that may benefit from VIA.
- Porting UC Berkeley Network of Workstations (NOW) software to Linux.
- Cluster administration and management tools.
- Upgrading the Parallel Distributed Systems Facility for High Energy and Nuclear Physics experiments.

**Making a Smooth Transition to Better Data Storage**

For the NERSC Mass Storage Group, 1998 was a year of conversions in support of the overall goal of consolidating various systems and providing new capabilities. At the start of the year, NERSC supported the following storage products: CFS (Common File System), UniTree, Networker, AFS (Andrew File System), and DFS (Distributed File System). The group's goal was to put newly acquired hardware into useful production and consolidate the capabilities provided — without any interruption of service or loss of data. The provided capabilities (in priority order) are:

- archival storage
- distributed file systems
- backup service.

Because we have focused upon HPSS (High Performance Storage System) as our long-term archival solution, NERSC's highest priority was to invest the majority of effort in learning a new and highly complex archival environment, then transferring all the data currently residing in the production UniTree and CFS environments into the new HPSS system. To achieve this, the Mass Storage Group was fortunate to have the talents of Jim Daveler, an HPSS developer with years of experience, and Nancy Meyer, who acted as conversion project team leader.

Additionally, NERSC had the good fortune and foresight to be one of the five HPSS development sites, giving us full rights to the HPSS product and the ability to deploy as many systems as needed. In order to learn the HPSS system and satisfy the demand for backup services, the first production HPSS system deployed at NERSC served as a backup server. This system was put into production in October 1997 and allowed the operations staff and consultants to became familiar with HPSS. We also fine-tuned our operational procedures and documentation while preparing to import the data from our first "production" user environment, the UniTree archival system. The first "user" production HPSS environment was deployed in November 1997. Users were given access to this system and were encouraged to become familiar with the new system before all archived files were converted to the HPSS system.

In February 1998, we converted the UniTree storage system to HPSS, accompanied by an extensive amount of planning, testing and user interaction during the transition. Along with users accessing the production UniTree storage system, Cray data migration and backups also relied upon the system. The UniTree-to-HPSS conversion process had been done before by other sites and NERSC was able to draw upon that experience. The next step, however, converting the CFS system, had never been done before. NERSC completed this project in early 1999.

The Mass Storage Group continues to provide a robust AFS environment while experimenting with DFS. Both AFS and DFS backups are now saved directly into our production HPSS environment, allowing us to phase out Networker, another step in the consolidation of systems. NERSC's backup server and production HPSS user system provide a backup capability which is being utilized by our machine room server machines, including the Cray supercomputers.

**Bolstering NERSC's Array of Cray J90s**

Two additional Cray J90 computers were added to NERSC's existing cluster of four J90s until the Cray SVI systems were delivered. The two new machines, "jwatson" and "fcrick," were installed to compensate for the decommissioning of the Cray C-90, the only machine to make the move to Berkeley from Lawrence Livermore. The two added machines are also batch machines. Once disconnected and decommissioned here, the C-90 was returned to Cray Research.

# WISE TECHNOLOGY INTEGRATION

***Goal: Insure all new technology and changes improve (or at least do not diminish) service to our clients.***

Merely installing the latest technology doesn't ensure that it will meet clients' needs. In fact, sometimes the opposite is true. To ensure that the systems it provides to clients are reliable, NERSC evaluates the technology and often tests systems off-line before deciding whether to implement the system for production, or file the results away under "lessons learned."

Once a new system is placed into production, various NERSC groups then work with clients to put the systems into production, obtain feedback and further improve the technologies. Here are some examples of this wise technology integration as practiced at NERSC during the past year.

### Making the Switch to HPSS

For the past three years, NERSC's Mass Storage Group has been working to install new hardware, consolidate storage systems and provide clients with new capabilities. A primary goal has been to convert the two archive systems — CFS and UniTree — to HPSS, or High Performance Storage System. Because NERSC is one of the five HPSS development sites, we were well positioned to make the conversion without interrupting service or losing data.

To do this, NERSC has phased in HPSS. The first step was to set up a backup server as an HPSS system which could help staff learn the system while at the same time satisfy demand for backup services. The first "user" production HPSS environment was deployed in November 1997. Clients were given access to this system and were encouraged to become familiar with the new system before all archived files were converted to the HPSS system. In February 1998, the UniTree storage system was converted to HPSS. The UniTree-to-HPSS conversion process had been done before by other sites and NERSC was able to draw upon that experience. The next step, converting the CFS system, however, had never been done before. NERSC completed this project in early 1999, drawing heavily upon the experience in gradually melding HPSS into NERSC's overall system configuration. (For more details on HPSS, see the New Technologies section in this report.)

### The Parallel Distributed Systems Facility (PDSF)

The PDSF is a networked, distributed computing environment used to meet the detector simulation and data analysis requirements of large-scale high energy and nuclear physics (HENP) investigations. The system originated from the Particle Detector Simulation Facility for the now-canceled Superconducting Supercollider. The original system consisted of various Sun, SGI and Hewlett-Packard workstations.

After coming to Berkeley Lab, PDSF was thoroughly reconfigured with new workstations featuring Pentium processors and the Linux operating system. Although the technology has been completed updated, the PDSF system will provide the same functionality as originally intended and is expected to provide a strong computing platform for numerous future HENP experiments. Features of the new cluster of workstations include:

- Significant CPU power
- Large data storage capacity

- Environment capable of implementing parallel processing schemes
- Flexibility towards expansion and upgrades as production needs grow.

The bottom line for PDSF is that the new system is less expensive to operate and maintain, is easier to run, and provides better performance.

### And Sometimes It's Wise to Say No

The Advanced Systems Group was responsible for administering a non-production computer system and assessing the potential of an SGI Origin2000. This 64-processor system on loan to NERSC was tested in regard to its software and architecture, with use limited to NERSC staff. The group concluded that the system was not appropriate for meeting NERSC's mission and will be returned to the manufacturer.

## PROGRESS MEASUREMENT

***Goal: As a national facility serving thousands of researchers, NERSC has a responsibility to our clients to measure and report on how we're doing in terms of providing service, support and facilities.***

This report is one of several documentation efforts on how we're doing in terms of meeting the goals we have set for ourselves. We also report — and get feedback — regularly at biannual ERSUG meetings and post statistics on the Web.

To track our progress in meeting these goals, we have established a set of metrics to measure our performance over a wide range of efforts. These measures range from the readily apparent, such as how long it takes to resolve a client's problem and the percentage of time our systems are available to clients, to the less obvious, such as how often our staff members transfer their technological expertise to the high-performance computing community. These metrics and resulting data are presented throughout this report.

Essential to meaningful measurement is taking the perspective of our client community. In 1998, for the first time since NERSC moved to Berkeley, we conducted a survey of all our clients. The survey, conducted by NERSC's User Services Group, is intended to provide user feedback about every aspect of NERSC's operation, help judge the quality of services, point to areas for improvement, and show how NERSC compares to similar facilities.

Clients rated NERSC in the areas of computational and file storage resources, account and allocations support, consulting services, documentation, and training. An area that wasn't included explicitly in the survey, but that clients commented on, was software support. The three areas rated most important are network access, the center overall, and consulting services. The three areas that received the highest ratings are consulting services, network access, and account support services.

Here are some comments by survey respondents:

"Gives us access to high-end, high-capacity and capability machines without too many limitations on the way codes can be run. There is a good mix of machines, serial and parallel, and a good mix of running modes, interactive and batch."

"Supplies reliable cycles to a broad user base. Excellent user support, probably the best in the world for high-end computing."

"As a service center NERSC does a great job keeping the T3Es up and providing easily accessible information relating to hardware and software. Moreover, the intellectual quality of NERSC support and research staff is excellent and they are easy to communicate with."

"Over the years NERSC has been the most reliable source of production computing for scientific research and development, by concentrating on scientific computing ahead of computer science."

You can read more about the survey and results on the web at:

http://www.nersc.gov/whatsnew/survey/

As the unparalleled leader in unclassified computing in the United States, NERSC is committed to providing our clients and DOE sponsors with the most reliable and productive scientific computing resources. An integral part of this is repeatedly taking stock of our efforts, using this as a benchmark for further improvements, and then continuing the process.

## HIGH-PERFORMANCE COMPUTING CENTER LEADERSHIP

***Goal: Improve methods of managing systems within NERSC and be the leader in large-scale computing center management.***

As the Department of Energy's largest unclassified scientific computing facility, NERSC is a leader in implementing new technologies, helping define the supercomputing center of the future and delivering the scientific, engineering and management expertise to benefit both NERSC and the high-performance computing community at large. During the period covered by this report, NERSC's T3E-900 was ranked fifth among the world's top 500 supercomputers and was the highest-ranked unclassified machine in the United States. However, leadership is more than a ranking on a list. Examples of NERSC's expertise and knowledge-based leadership in enhancing large-scale computing include:

### Taking Scientific Computing to a New Level

In 1998, NERSC was a partner in a computational science effort which achieved breakthrough results both in the type of science enabled and the level of computing performance achieved. The project by the Materials, Methods, Microstructure and Magnetism Grand Challenge team, consisting of scientists NERSC, Oak Ridge and Brookhaven National Laboratories, and Ames Laboratory, seeks to bridge the gap of understanding about the properties of certain materials. Specifically, scientists know about properties at the atomic scale, as well as at the macroscale. But in between is an unknown area, called the mesoscale, where a material's properties change as the size increases over time. Understanding the mesoscale is a primary goal of computational materials science and this project provided a key step toward that goal. This project, which has applications in fields ranging from computer data storage to power generation and utilization, aims to dramatically advance the computational study of materials, leading to a much better understanding of the relationship between the atomic-scale structure of materials and their macroscopic structural and electrical properties. One of the findings was that magnetic fields in metallic magnets "migrate" among regions of atoms as the temperature increases. This was made possible by scaling the size of the code up to model 1,012 atoms.

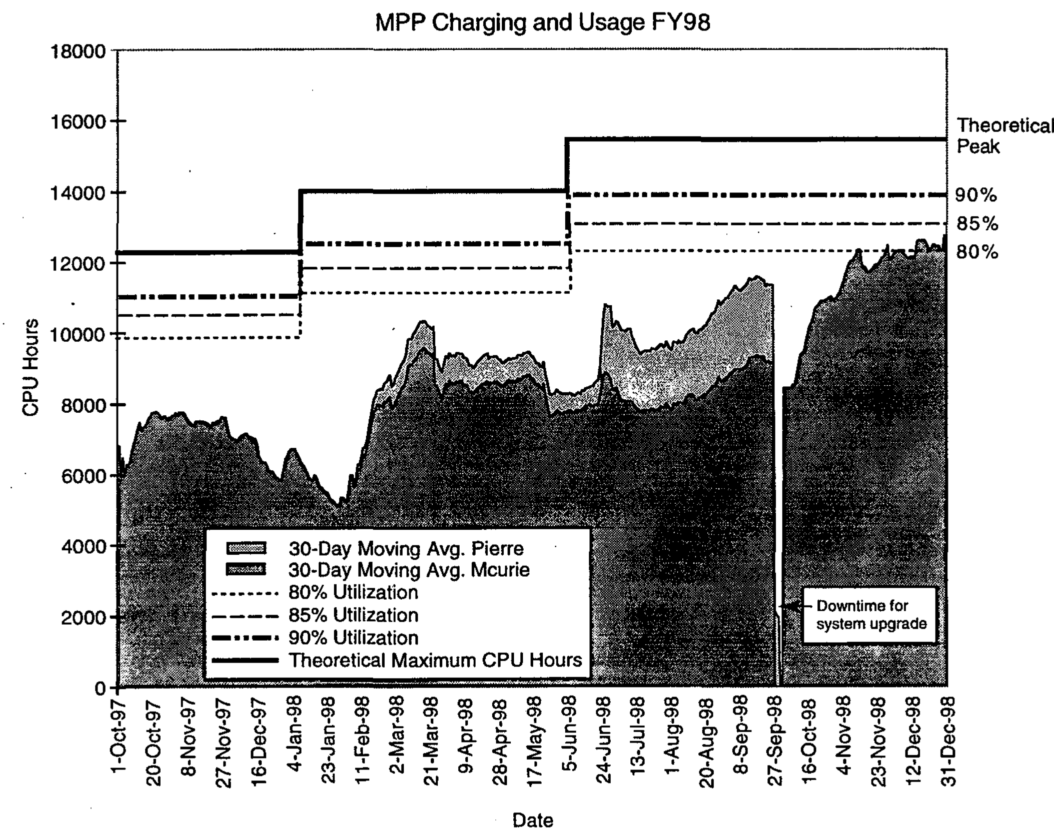
The simulation also achieved breakthrough performance in November 1998 by achieving 1.02 teraflop/s performance — making it the first real scientific application to achieve real scientific results at more than 1 teraflop/s performance. To accomplish this feat, the team was granted access to a Cray T3E supercomputer still on the assembly floor of a Cray manufacturing facility.

The code was developed and debugged on NERSC's 512-processor Cray T3E-900. The team ran their computer application on three increasingly powerful Cray T3Es, each time achieving higher performance. Over the summer of 1998, the team achieved a level of 657 gigaflop/s on a U.S. government Cray T3E-1200. This achievement resulted in the group being chosen as one of four finalists for the Gordon Bell Award, awarded annually to honor the best achievement in high-performance computing. The team was named the winner of the prize when it was announced Nov. 12, 1998, at the SC98 conference. While attending the conference, Andrew Canning of NERSC's Scientific Computing Group logged onto the Cray machine from his hotel room and maintained the connection for 14 hours to finally exceed the 1 teraflop/s level. At each step of the way, Andrew worked with both the researchers and the operators of the computers to obtain optimum performance and results.

**Achieving Greater Capability with Existing Systems**

The NERSC Systems Group, in addition to being responsible for developing and managing NERSC's computing infrastructure, also assesses and develops hardware and software, as well as oversees the installation and configuration of new systems.

One major effort during 1998 centered on NERSC's two Cray T3E supercomputers. The smaller machine, "Pierre," received a number of upgrades in preparation for merging it with the larger "MCurie" to create a single 640-processor system. At the same time, the Systems Group worked closely with analysts from SGI/Cray to develop and implement scheduling features for the T3E. With better scheduling, and the checkpoint/restart capability achieved during the previous year, NERSC will be able to start, stop and move jobs on the computer. This will result in greater utilization (the goal is 90 percent or more) and the capability for running larger jobs more often. According to a Cray representative, this capability will be unequaled among other centers using T3Es, although NERSC's experience is expected to benefit other centers.




**Providing a Million Hours in a Matter of Months**

In October 1997 NERSC passed the milestone of delivering 1,000,000 parallel computing CPU-hours to the scientific community. The records of all parallel computation performed on the prototype Cray T3E-600 from late January to early September, and on the full configuration 512-processor T3E-900 from late July to October 23 indicated that Grand Challenge users received

425,000 parallel computing hours, while the remaining 575,000 hours were used by more than 100 other research groups.

### Looking Ahead to the Next Generation of Scientific Computing Hardware

In March 1998, NERSC established a new group, called the Advanced Systems Group, to assess and implement new computing technologies. The group's primary activity in 1998 was participating in the procurement of the next generation computer for NERSC, which is expected to go on line in mid-1999. The decision as to which technology to pursue is critical to NERSC's future, and the NERSC-3 procurement is the first to be conducted since NERSC moved to Berkeley Lab.

### Enhancing DOE's Scientific Visualization

From March 30 to April 1, 1998, NERSC's Visualization Group hosted a "Workshop on Interoperability of DOE Visualization Centers," inviting representatives from all DOE national laboratories and other research organizations to discuss and report on ways to improve the interactions between scientific visualization groups. The workshop was developed by NERSC at the request of the Mathematical, Information, and Computational Sciences (MICS) Division in the Office of Energy Research. In addition to MICS, organizations attending were Berkeley Lab; Idaho National Engineering and Environmental Lab; Argonne, Brookhaven, Lawrence Livermore, Los Alamos, Sandia, Oak Ridge and Pacific Northwest national labs; NASA-Ames Research Center; Stanford Linear Accelerator Center; the National Center for Supercomputing Applications; and Georgia Tech. The workshop addressed technical issues relevant to the interchange and sharing of visualization research and software. The result was a document that identifies technically feasible solutions to the challenges posed by using tools developed in unfamiliar and sometimes incompatible environments, and ways to improve communications and overall interoperability. More information about the workshop and a link to the report can be found on the web at: http://www-vis/DOEvis.html

### Center for Bioinformatics and Computational Genomics Created

As the Human Genome Project continues to decipher the secrets of our genetic makeup, scientists around the world are gaining new insight into the biology of health and disease. Computational tools are crucial in making the discoveries possible. To streamline the retrieval of key information from the ever-growing banks of data, the Center for Bioinformatics and Computational Genomics, or CBCG, has been created in NERSC. The new center will unify bioinformatics efforts at Berkeley Lab and support the Department of Energy's Genome Program. The center is led jointly by Sylvia Spengler from Life Sciences and Manfred Zorn from Computing Sciences.

Mapping and sequencing the human genome is currently an international effort involving two dozen large centers, including the Department of Energy's Joint Genome Institute. The current rate of overall sequencing of base pairs is 150 million base pairs per year. However, to fully sequence the 3 billion base pairs in the human genome by 2005, the participating centers will need to sequence 2 million per day. The information contained in the genome data is of immeasurable value to medical research, biotechnology, pharmaceuticals and researchers in fields ranging from microorganism metabolism to structural biology. At expected data rates, the sequences generated each day for the next five years will represent hundreds of new genes and proteins.

### Critical Tool Library for HENP Calculations Ported to T3E

A NERSC-led collaboration with Brookhaven National Lab and the Pittsburgh Supercomputing Center has ported the CERNLib collection of codes to NERSC's Cray T3E supercomputer, completing the task in May. Craig Tull, leader of NERSC's High Energy and Nuclear Physics (HENP) Support Group, and Mark Durst of the User Services Group provided expertise for the work. Completion of the project paves the way for conducting detailed simulation of physics data from the STAR detector being installed at Brookhaven. The detector is scheduled to begin taking real data starting next year. Closely coupled to the work is the High Energy and Nuclear Physics Data Grand Challenge, a project to develop tools to allow physicists to analyze and manage the massive amounts of data which will be generated by next generation of experiments.

According to Craig, the HENP community has traditionally relied on single-processor UNIX platforms and there had been no incentive to port CERNLib to massively parallel processor machines. Until now. With the STAR experiment, as well as the BaBar Project at the Stanford Linear Accelerator Center, scientists will be confronted with huge quantities of data, and therefore need more powerful analytical tools. The CERNLib package provides the tools and the T3E provides the power. Born at CERN, the European Organization for Nuclear Research, CERNLib has been developed and supplemented by HENP researchers around the world. The codes are centrally maintained and codified by CERN in Geneva, Switzerland.

### Massive Data Transfer Tests System for Supporting HENP Experiments

When it comes on line in October 1999 at Brookhaven National Lab, the STAR Project (the Solenoidal Tracker at RHIC, the Relativistic Heavy Ion Collider) is expected to generate up to 1.5 terabytes of data a year from high energy and nuclear physics experiments. To make sure the computing and storage hardware and software can handle the load, massive amounts of data from simulations have been produced at NERSC and the Pittsburgh Supercomputer Center. In August and September 1998, the two centers poured data into Brookhaven at an average of 50 gigabytes a day. In all, about half a terabyte was transmitted by NERSC, with a like amount from PSC. Participants in the STAR experiments then conducted a mock data challenge to make sure data can be entered, processed and stored. If all goes as planned, data from actual experiments will start being taken in October 1999. The project was a collaboration not only between NERSC, PSC and Brookhaven, but also among three groups in Berkeley Lab's Computing Sciences organization, NERSC's HENP Support and Mass Storage groups and the DPSS project.

### NERSC Co-Hosts Symposium on Solving Irregular Problems

The Fifth International Symposium on "Solving Irregularly Structured Problems in Parallel" was held August 9-11, 1998, at Berkeley Lab. Called "IRREGULAR '98," the symposium focused on algorithmic, applicational and system aspects arising in the development of efficient parallel solutions to irregularly structured problems. A primary goal was to foster cooperation among practitioners and theoreticians in the field. The invited speaker was John Gilbert of Xerox PARC. This was the fifth in a series, with previous symposia being held in Geneva, Lyon, Santa Barbara and Paderborn. The meeting is co-sponsored by the Mathematical Sciences Research Institute and is supported by DOE's Office of Science.

### Lab Hosts Applied Math Conference for Minority Students

About 50 minority students in graduate mathematics programs from across the nation attended a three-day conference at the Lab in September to get a firsthand look at the state of applied math research and the opportunities available. The conference, which NERSC Chief Technologist David Bailey helped organize, was co-sponsored by NERSC, the Mathematical Sciences Research Institute (MSRI) and the Joint Alliance for Minorities in Mathematics (JAMM).

In summary, NERSC aims to be a leader in providing computational science resources to its client community and within the high-performance computing community as a whole. Achieving this goal demands technological expertise, a dedicated staff with a varied range of backgrounds and a willingness to share their experience, and close interactions with researchers to determine their needs and develop the tools to meet them.

# TECHNOLOGY TRANSFER

***Goal: Export knowledge, experience, and technology developed at NERSC, particularly to and within NERSC client sites.***

In assembling the staff at Berkeley, NERSC has hired employees from some of the nation's pre-eminent scientific and research facilities. Their expertise is being shared with NERSC clients, as well as other scientific communities. Here are some examples of how we're sharing our technological expertise with our clients and the high-performance community:

### Making Outstanding Contributions to High Performance Computing

Phillip Colella, a mathematician and leader of NERSC's Applied Numerical Algorithms Group, was named the recipient of the IEEE Computer Society's 1998 Sidney Fernbach Award, given each year to one person who has made "an outstanding contribution in the application of high performance computers using innovative approaches."

Presented during SC98, the annual conference on high-performance networking and computing, the award is named for one of the pioneers in the development and application of supercomputers for solving large computational problems. The award is sponsored by the Institute of Electrical and Electronics Engineers Computer Society. Colella received the award on Thursday, Nov. 12, at the conference in Orlando, Fla.

Colella, who has been a staff member of NERSC and the Computing Sciences Directorate at Berkeley Lab since 1996, was recognized "for fundamental contributions in the development of software methodologies used to solve numerical partial differential equations, and their application to substantially expand our understanding of shock physics and other fluid dynamics problems," according to a letter from Doris Carver, president of the IEEE Computer Society.

### Providing Up-to-Date User Training

Two sessions of "NERSC Training for New T3E Users," a two-day series of lectures, were offered in February for Berkeley Lab users and in April, in conjunction with the ERSUG meeting. Topics included an introduction to the T3E system, parallel programming, parallel I/O, and message passing interfaces. For more information about this training, see the earlier section on Timely and Accurate Information.

### Invited Talks:

- "First-Principles Plane-Wave Codes on Parallel Computers," Andrew Canning, DOE workshop on Mathematical Aspects of Materials Science Modeling, Gatlinburg, TN, April 1998.

- "Parallelization of Electronic Structure Codes Based on Plane Wave Expansion, " Andrew Canning, American Physical Society Annual Meeting, Los Angeles, March 1998.

- "Parallel Atmospheric Data Assimilations," Chris Ding, Second International Workshop for Software Engineering and Code Design for Parallel Meteorological and Oceanographic Applications, Scottsdale, AZ, June 1998.

- "Material Science on Parallel Computers," Andrew Canning, Cray User Group meeting, Stuttgart, Germany, June 1998.

**Conference Presentations and Proceedings**

- "Parallel Lanczos Method for Symmetric Eigenvalue Problems," Kesheng "John" Wu, poster presentation at SC97, San Jose, CA, November 1997.

- "Thick-Restart Lanczos Method for Symmetric Eigenvalue Problems," Kesheng "John" Wu, Copper Mountain Conference on Iterative Methods, Colorado, March 1998.

- "Remote Visualization Efforts at LBNL/NERSC," "LBNL/NERSC Site Report," and "Parallelization of Radiance for Real Time Lighting Visualizations," Stephen Lau, Department of Energy Computer Graphics Forum, St. Michaels, MD, April 1998.

- "Thick-Restart Lanczos Method for Symmetric Eigenvalue Problems," Kesheng "John" Wu, Scientific Computing and Computational Mathematics conference, Stanford University, CA, April 1998.

- "The NetLogger Methodology for High Performance Distributed Systems Performance Analysis," Brian Tierney, IEEE High Performance Distributed Computing ( HPDC-7 ), July 1998, Chicago.

- "Accuracies of Restarted Lanczos Method," Kesheng "John" Wu, International Workshop on Accurate Solution of Eigenvalue Problems, State College, PA, July 1998.

- "Preconditioning Techniques for Large Eigenvalue Problems," Kesheng "John" Wu, University of Science and Technology of China, Hefei, China, July 1998.

- "Parallel Lanczos Method for Symmetric Eigenvalue Problems," Kesheng "John" Wu, Anhui University, Hefei, China, July 1998.

- "Thick-Restart Lanczos Method for Symmetric Eigenvalue Problems," Kesheng "John" Wu, International Symposium on Theory and Algorithms for Large Scale Matrix Problems, Dalian, China, August 1998.

- "A new Lanczos Method for Electronic Structure Calculations," Andrew Canning, Horst Simon and Kesheng "John" Wu, Proceedings SC98, IEEE, Orlando, Fla., November 1998.

- "S-HARP: A Scalable Parallel Dynamic Partitioner for Adaptive Computations," Horst Simon and Andrew Sohn, SC98, November 1998, Orlando, Fla.

**Workshops and Tutorials**

- "What is XML?" John McCarthy, LBNL Summer Seminar Series, July 1998.
- "The Ubiquitous FFT," Peter Tang, LBNL Summer Seminar Series, July 1998.
- "Embarrassing Parallelism and Pseudo-Random Number Generation," Mark Durst, LBNL Summer Seminar Series, August 1998.
- "Introduction to NERSC Resources," Mark Durst, Mathematical Sciences Research Institute lecture series, Berkeley, September 1998.
- At the International Computational Accelerator Physics Conference held in Monterey, CA, September 1998 (afternoon tutorial session):
  - "Introduction to Parallel Programming on the NERSC T3E," Jonathan Carter.
  - "High Performance Computing at NERSC," Richard Gerber.
  - "Parallel Scientific Libraries," Peter Tang.
  - Examples provided by NERSC users Robert Ryne (LANL, beam dynamics) and Brian McCandless (SLAC, electromagnetics).
- "A Rapid and Practical Introduction to the Message Passing Interface (MPI)," William Saphir, SC97 tutorial, SC97, San Jose, CA, November 1997.
- "MPI-2," William Saphir, SC97 tutorial, SC97, San Jose, CA, November 1997.

**Publications**

- "Solving Irregularly Structured Problems in Parallel," book edited by Horst Simon (with A. Fereira, J. Rolim, and Shang-hua Teng), Proceedings Irregular 98, Springer Lecture Notes in Computer Science No. 1457, August 1998.
- "A Fast Poisson Solver for the Finite Difference Solution of the Incompressible Navier-Stokes Equations," Horst Simon (with Gene H. Golub, Lan Chieh Huang, and Wei-Pai Tang), SIAM Journal on Scientific Computing, Volume 19, Number 5, pp. 1606-1624, 1998.
- "HARP: A Dynamic Spectral Partitioner," Horst Simon, Andrew Sohn and Rupak Biswas, Journal of Parallel and Distributed Computing 50, April 1998, pp. 88-103.
- "Challenges of Future High-End Computing," David H. Bailey, in High Performance Computer Systems and Applications, Jonathan Schaeffer, editor, Kluwer Academic Press, Boston, 1998.

### Conference Organization

- SC 97 and SC98: Members of the NERSC staff played important roles in these annual conferences. From chairing technical committees to assisting with communications to organizing the exhibit floor to coordinating special events, NERSC employees devoted significant time over the course of the year to helping plan, coordinate and carry out various aspects of the conferences in San Jose and Orlando.

- "Visualization 97," sponsored by the IEEE Computer Society Technical Committee on Computer Graphics, Phoenix, AZ, October 1997. NERSC Visualization Group Leader Nancy Johnston co-chaired the six-day meeting, group member Kevin Campbell was Audio-Visual co-chair and Stephen Lau was E-mail/Networking chairman.

### Evaluation and Input

The Virtual Interface Architecture (VIA) specification released in December 1997 by Intel, Compaq and Microsoft was evaluated in draft form by NERSC computer scientists. VIA provides standardized low-latency, high-bandwidth communication mechanisms on commodity networks and is expected to benefit distributed computing in clusters, such as the Millennium collaboration between NERSC, UC Berkeley and Intel. VIA could make clusters more useful for high-performance scientific applications. NERSC is also developing a version of VIA to run on the Linux operating system.

### Sponsored Workshop

From March 30 to April 1, 1998, NERSC's Visualization Group hosted a "Workshop on Interoperability of DOE Visualization Centers," inviting representatives from all DOE national laboratories and other research organization to discuss and report on ways to improve the interactions between scientific visualization groups. The workshop was developed by NERSC at the request of the Mathematical, Information, and Computational Sciences (MICS) Division in DOE's Office of Science.

# STAFF EFFECTIVENESS

***Goal: Leverage staff expertise and capabilities to increase efficiency and effectiveness.***

An important part of the agreement to move NERSC from Livermore to Berkeley was a commitment to do more with less. The "less" side of this equation is easy to measure. NERSC is currently operating on a smaller budget than before and with fewer employees. For example, NERSC had a full-time technical staff of 79 employees in FY 1994. In FY98, we have a staffing level of 64. The annual budget has also been reduced, from a high of $40 million to the current level of $29 million, while the amount spent on equipment used by clients remains the same. On the other side of the equation, NERSC now addresses expanded expectations and responsibilities. In order to meet the "more" requirement, NERSC staff and resources must be carefully deployed for maximum effectiveness and efficiency. One of our primary objectives is to increase overall effectiveness, not only at NERSC, but within the computational science community. Here are some examples of how we're achieving that goal:

## Making Tools Widely Available

In August 1998, NERSC unveiled a dedicated web site for the ACTS (Advanced Computational Testing and Simulation) Toolkit, a set of DOE-developed software tools that make it easier for programmers to write high-performance scientific applications for parallel computers.

The purpose of ACTS is to accelerate the adoption and use of advanced computing by DOE programs for their mission-critical problems. While DOE has been motivated to develop the tools for its own programs, it also encourages their adoption and use by non-DOE computational efforts. The ACTS project is focusing on establishing a modest collection of useful parallel tools and emphasizing further development that will allow the tools to interoperate. ACTS is funded by the DOE 2000 initiative. The ACTS Toolkit web site has been created by NERSC with support from the MICS Division in DOE.

Once the site was up and the tools made available, members of NERSC's Future Technologies Group began publicizing and encouraging interested researchers to put the tools to work. NERSC staff were also ready to help interested researchers with the applications.

The tools include "traditional" numerical libraries (e.g., linear system and ODE solvers), tools that provide infrastructure to manage some of the complexity of parallel programming (such as distribution of arrays and communication of boundary information), and tools that provide run-time support ranging from performance analysis to real-time remote visualization.

For more information, visit the web site at: http://www.nersc.gov/ACTS or send email to: acts-support@nersc.gov

## Learning from Our Peers

In July 1998, members of NERSC's User Services and Operations groups visited the San Diego Supercomputer Center to exchange technical information in the areas of web support, consulting, training, operations, and HPSS. A series of meetings were held to share information between NERSC and SDSC, which is part of NPACI, the National Partnership for Advanced Computational Infrastructure. NERSC is a member of NPACI and a partner for joint evaluation of the Tera computer.

One project by SDSC which NERSC emulated was a survey of clients (see a report on results elsewhere in this report). The visit has also led to ongoing exchanges of information on specific topics.

**Funding to Enhance Climate Modeling Tools**

DOE's Biological and Environmental Research (BER) Program began funding a joint climate research project between NERSC and the Geophysical Fluid Dynamics Laboratory (GFDL) to investigate how widely used climate models can be run effectively and efficiently on massively parallel processing supercomputers. NERSC will focus work initially on the Modular Ocean Model developed at the GFDL, and investigate additional models in future years. The proposal looked at the bottlenecks that have prevented the effective and efficient use of state-of-art scalable systems for climate research, then proposed developing new computing technologies to eliminate the bottlenecks. NERSC will work closely with the model developers at GFDL to incorporate the improvements into the models.

**NERSC Gets NASA Code on the Parallel Fast Track**

Chris Ding of NERSC's Scientific Computing Group has been working to adapt a NASA-developed code for assimilating climate-related observational data from various sources to run on a parallel computer. Originally developed at NASA's Goddard Space Flight Center in 1995 to run on a vector machine, the Physical-space Statistical Analysis System creates more reliable datasets against which climate modeling tools can be compared and calibrated. The assimilation's requirements exceed the capabilities of vector computers, so Chris adapted it to run on parallel machines. In getting the code to run on NERSC's 512-processor T3E, Chris was able to produce a solution to an 80,000-observation problem in 24.6 seconds, as opposed to five hours on a single node of a Cray C90.

The code was developed by NASA's Data Assimilation Office to assimilate data gathered by satellites, flying balloons, ground stations and other sources into more accurate datasets. The problem is that various data, such as for temperature, humidity and winds, have large errors because of variability in the collection methods. However, by assimilating them to a numerically simulated forecast, through a so-called "innovation equation," an optimal dataset is produced which is more accurate and consistent. This dataset is also comprehensive, since the assimilated data is "folded back" to a regular grid, providing global coverage to compensate for the less observed regions such as Africa and South America. The ability to assimilate large amounts of data is key to the Department of Energy's proposed Accelerated Climate Prediction Initiative. Not only will these data allow for more accurate climate models, but they are also expected to increase our understanding of historical climate patterns.

**NERSC Gets Foothold in Chemical Industry Research Efforts**

NERSC joined a consortium of DOE labs, universities and industry funded by the DOE Office of Industrial Technology to investigate multiphase fluid dynamics problems arising in the chemical processing industry. The consortium, officially known as the "Multiphase Fluid Dynamics Research Consortium," is composed of two major collaborations; one oriented toward experimental methods and one toward computational methods. NERSC is involved in the latter, in collaboration with the Federal Energy Technology Center (or FETC, a DOE sponsored lab in Morgantown, W.Va.), Oak Ridge National Lab, Dow Corning Co., and Fluent, Inc. (a computational fluid dynamics software vendor). The other collaboration is led by Dow Chemical Company and Sandia Lab. NERSC is providing expertise in parallel computing. In particular, this involves parallelizing a computational fluid dynamics (CFD) code developed at FETC for simulation of multiphase flows, deploying it on the Cray T3E (for solving large-scale research problems) and on a PC cluster (for high cost-effectiveness on industrial problems).

**More Effective Ways of Doing Business**

At LLNL, NERSC's hardware consisted of three production vector supercomputers, one production MPP machine and one storage system. The current configuration at Berkeley includes five production vector supercomputers, one production MPP machine and two production storage systems which will be configured as a single system during FY99.

Previously, NERSC printed a bimonthly newsletter/magazine called "The Buffer" for clients and staff. This required two full-time employees and part-time contributions from many others. This has been replaced with a monthly web-based NERSC Research newsletter, produced by the equivalent of a quarter-time position. The electronic format saves printing and design costs, conserves resources and delivers more timely information to a potentially wider audience.

NERSC has also adopted an e-mail notification approach to keep clients informed. By electronically delivering monthly accounting reports, allocation notices and other information, NERSC is saving time, money and materials. This process also requires less staff time and provides faster distribution of information.

For FY99 allocation requests, NERSC continued to refine its web-based ERCAP request system. Drawing on lessons learned during the FY97 allocation applications process, ERCAP was improved for FY98. The improved process drew generally positive feedback. The new automated allocation request procedure is resulting in significant time savings over the more time-consuming manual procedures of the past.

Training for NERSC clients is also being improved with an eye toward greater effectiveness. In addition to on-site training, we have also implemented a web-based training program for the Cray T3E and are exploring the use of videoconferencing tools for remote training sessions. This effort resulted in the availability of videotaped presentations for viewing via the web.

We also hold monthly conferences with ERSUG members, producing more frequent two-way communication with representatives of our key client groups.

Finally, the NERSC staff began revising the entire NERSC web site in summer of 1998 to better meet the needs of various audiences for information about the center.

We believe we have "built a better qualified staff" in Berkeley and have achieved a more flexible workforce to quickly support new technologies. For example, in 1994, NERSC's 79-member staff included 42 with degrees, 21 with advanced degrees and 11 Ph.Ds. In 1998, 44 of the 62 staff members have degrees, with 24 having advanced degrees and 21 having Ph.Ds. Although having a degree is not alone a measure of effectiveness, the total picture is one of an intellectual center positioned to help clients in new areas and with new perspectives.

The bottom line is that we believe NERSC continues to offer excellent service to clients, whether it's an advisory phone call or the provision of new software applications. The staff is constantly looking for new ways to meet our clients' needs with approaches that are cheaper, better and faster.

## PROTECTED INFRASTRUCTURE

***Goal: Provide a secure computing environment for NERSC clients and sponsors.***

The security of NERSC's systems is a paramount concern to the Berkeley Lab staff, our DOE sponsors and our national client base. To ensure the ongoing security of the NERSC systems, we have taken the following steps:

- Hired a person to develop and support a standardized workstation configuration that will keep all NERSC workstations updated with the latest security patches.
- Hired a Computer Security Analyst to formalize a NERSC security plan and develop/implement a security program.
- Installed a security monitor to warn of suspicious incidents on the NERSC network.

Much of the benefit of these steps will be realized next year; however, our informal security infrastructure has allowed us to finish the year without a significant security incident.

## CONCLUSION

When NERSC was relocated to Berkeley Lab in 1996, it marked the rebirth of the center. Established in 1974, NERSC was one of the nation's first unclassified supercomputing centers and was a reliable supplier of supercomputing cycles. But the Department of Energy determined that the center could provide more effective leadership in the field of scientific computing and sought new ideas. The result was the move to Berkeley Lab, where there is greater interaction with UC Berkeley. NERSC also provides intellectual resources in addition to computing resources. Since its arrival in Berkeley, NERSC has added seven large-scale computers, a high-speed storage system and more than 30 new employees from some of the top computing science organizations in the country.

Scientific computing is on the threshold of a new era, made possible by the availability of teraflops computers and petabyte data storage systems. Although the results of this new era in computing are expected to dramatically improve our health, our economy, our education system and our scientific leadership, achieving them will take determination and commitment. We believe many of the essential components are already in place at NERSC and new ones are being added.

To document the quality of our work and demonstrate our commitment to our clients, we have instituted the system of metrics outlined in this report. We believe the data demonstrate that we're doing a good job in meeting clients' needs, and that there is also room for us to do better. We plan to do just that, and will report back to you on our further progress next year.

## ACKNOWLEDGMENT

This work was supported by the Department of Energy's Office of Science, Office of Advanced Scientific Computing Research, Mathematical, Information and Computational Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

This report was compiled by Jon Bashor. For more information, call 510-486-5849, or write to JBashor@lbl.gov. To learn more about the National Energy Research Scientific Computing Center, visit our web site at: http://www.nersc.gov.

ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY
ONE CYCLOTRON ROAD | BERKELEY, CALIFORNIA 94720

Prepared for the U.S. Department of Energy under Contract No. DE-AC03-76SF00098