How Are We Doing?

A Self-Assessment of the Quality of Services and Systems at NERSC (1999)

Issued May 2000 by the
High Performance Computing Department
William T. Kramer, Department Head

National Energy Research Scientific Computing Center

Ernest Orlando Lawrence Berkeley National Laboratory

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

EXECUTIVE SUMMARY

As we enter our second quarter century, NERSC continues to set high standards for anticipating users' scientific computing needs, then assessing and deploying new systems and services to address those needs. Since it was established in 1974 as the nation's first unclassified supercomputing center, NERSC has continually adopted the newest technologies and pioneered new ideas to redefine the field of high performance computing. Since the center was relocated at Berkeley Lab in 1996, NERSC has continued to push the parameters of computing resources, data storage, user services and scientific applications to accelerate the contributions of scientific computing across the spectrum of DOE-sponsored research. Based on reviews by outside experts, results of a user survey and our own internal assessment, NERSC retains its leadership position. This report details what the NERSC staff is doing to meet a set of goals established to ensure that our center retains its position as the nation's pre-eminent center for conducting unclassified computational science.

In 1999, the NERSC Policy Board was established to provide Berkeley Lab Director Charles Shank with overall policy direction and guidance. In its first report, the board noted, "The success or failure of any major scientific user facility is dependent ultimately on the degree to which the needs of the user community are satisfied. The Board was pleased to hear that NERSC has earned high marks from its users as indicated by the most recent user survey. We were also pleased with the user-oriented perspective that "we measure ourselves in the way our users measure us." Especially noteworthy is the user endorsement of a new allocation process involving the addition of a second peer review path operating in parallel with the DOE review path."

How are we doing? We're doing very well, thank you.

INTRODUCTION

In 1999, the U.S. Department of Energy's National Energy Research Scientific Computing Center achieved a significant milestone — 25 years as a leading center for high-performance computing — and mapped out a challenging future intended to maintain NERSC's leadership position.

Teraflops of computing power, terabytes of data storage and terabits of network capability are givens in today's world of high-performance computing and networking. The ability to routinely access data, computers and expertise is now an essential component of scientific research, but when the forerunner of NERSC was first proposed, skeptics said it couldn't be done and even if it could, there'd be no value in doing it.

Through the years, however, NERSC and ESnet (DOE's Energy Sciences Network) pioneered many of the computing and networking practices taken for granted today, and served as models for other agencies. These include remote access by thousands of users, high-performance data storage and retrieval, providing on-line documentation and around-the-clock support for users.

The center, originally called the Controlled Thermonuclear Research Computer Center, began providing cycles in July 1974, using a cast-off Control Data Corp. 6600 computer. Access was provided at 110 baud via four dial-up modems. Within a year, the center was home to a Control Data Corp. 7600. The center changed it name to the National Magnetic Fusion Energy Computer Center in 1976-77 and took steps to acquire a Cray-1 supercomputer, the first one delivered to Lawrence Livermore National Laboratory.

To take advantage of the Cray-1's capabilities, the center undertook a major software project to convert the 7600 operating system (LTSS or Livermore Time Sharing System), utilities and libraries to the new CRAY-1. The resulting system was called the Cray Time Sharing System (CTSS) and allowed interactive use of the CRAY-1. CTSS was later adopted by six other computer centers. Over the years, additional supercomputers such as the Cray-2 and Cray X-MP were installed at the center.

In the early years of NERSC, accessing archived data required an operator to retrieve the nine-track tape from a rack and load it, then notify the user that the data were available. Delivery of the Automated Tape Library in 1979 changed that, allowing hands-off access.

During the 1980s, the center broadened its research mission to support other programs within DOE's Office of Energy Research. In 1989, the name was changed to the National Energy Research Supercomputer Center to reflect the larger mission.

In 1991, several key decisions were made which shaped the future of NERSC. First, Bill McCurdy was chosen as the second permanent director of the center. His selection was note-worthy in that he came from outside the Lawrence Livermore community. Like his predecessor John Killeen, McCurdy was a research scientist — a chemical physicist. Within several months of McCurdy's arrival, NERSC canceled its contract with Cray Computer Corp. for the Cray-3 machine, and also halted development of operating system software. Together, these changes set the stage for NERSC to devote even more resources to advancing the state of computational science.

In the mid-1990s, NERSC and DOE began two transitions which would dramatically change the nature of computational science in the DOE community.

The first was the procurement of a Cray T3E-600, the first massively parallel processor-architecture machine to be installed at the center as a primary computational resource. The MPP architecture allows scientists to scale up their models to achieve larger, more accurate simulations. The MPP architecture, which marked a fundamental change in the computing environment, also led to the consideration of operational changes for the center.

In 1995, DOE made a decision to seek proposals for the future operation of NERSC. Berkeley Lab's proposal was selected and the decision to move the center was announced on Nov. 3, 1995. Among the reasons cited for the change were the integration of the center with ER-sponsored research at Berkeley Lab and the proximity to UC Berkeley and the opportunity to combine intellectual resources.

The physical move of NERSC and ESnet was achieved just two months later. During the move, at least part of the NERSC computing facility was continuously available to users. New Cray J90 supercomputers were installed and brought on line in Berkeley before machines in Livermore were shut down for the move. The result was a nearly transparent transition.

Since the move, NERSC installed the 128-processor Cray T3E and upgraded it to 696 processors. New storage systems were installed and the Cray J90s were upgraded to provide more powerful vector computing resources.

NERSC also developed new research programs and became integrated with existing computer science research efforts at Berkeley Lab and on the UC Berkeley campus. Among the areas of research are distributed computing, data storage, data management, computer architectures and data intensive computing. As in the past, NERSC's ability to bring varied intellectual resources to bear on difficult problems resulted in new approaches and innovative solutions.

For 25 years, the staff of NERSC and ESnet have continually defined and advanced the state of high-performance computing and networking for researchers across the nation. It has not always been easy to maintain the balance between the two elements of NERSC's mission — providing reliable, high-quality, state-of-the-art computing resources and client support in a timely manner, independent of client location, while wisely advancing the state of computational and computer science. But in the long run, these two goals reinforce each other and contribute to the advancement of science. Perhaps the largest overall contribution of the center has been to help prove the validity of computer simulations as a valid component of scientific research.

Conversion Completed to High Performance Storage System (HPSS)

In January 1999, NERSC completed the conversion of its two archival data storage systems to High Performance Storage Systems, or HPSS. The High Performance Storage System (HPSS) is a modern, flexible, performance-oriented mass storage system, designed and developed by a consortium of government and industry. Users are given HPSS accounts to enable them to economically save and access files that are too large or too infrequently used, to be kept in personal home directory spaces.

Since moving to Berkeley Lab, NERSC's archival storage capacity has been increased from 30 terabytes to nearly three-quarters of a petabyte. This increase of 200 terabytes per year far exceeds the increase projected under Moore's Law, which would have foreseen an increase from 30 to 180 terabytes in three-and-a-half years.

NERSC Selects IBM SP for Next Supercomputer

In April 1999, NERSC announced that it had selected an IBM RS/6000 SP system as the center's next-generation supercomputer. The IBM system was chosen based on its ability to handle actual scientific codes and tests designed to ensure the computer's capability as a full-production computing system in NERSC. These tests indicated that the system, when fully installed, will provide four to five times the total current computational power of NERSC, already one of the most powerful supercomputing sites in the world. This agreement, a fixed-price, five-year contract for $33 million, is the largest single procurement in the 68-year history of Berkeley Lab.

The new system, which will incorporate IBM's newest processor and interconnect technology, will be installed in two phases. When completed, the system will increase NERSC's computing capabilities by more than 400 percent. Phase I installation, which began in June 1999, consists of an RS/6000 SP with 304 of the two-CPU POWER3 SMP nodes that were recently announced by IBM. In all, Phase I will have 512 processors for computing, 256 gigabytes of memory and 10 terabytes of disk storage for scientific computing. The system will have a peak performance of 410 gigaflops, or 410 billion calculations per second.

Phase II, slated for installation no later than December 2000, will consist of 152 16-CPU POWER3+ SMP nodes, utilizing an enhanced POWER3 microprocessor. The entire system will have at least 2,048 processors dedicated to large-scale scientific computing. The system will have a peak performance capability of more than 3 teraflops, or 3 trillion calculations per second.

As part of the purchase contract, NERSC will work with IBM to develop computer-utilization benchmarks and methods to assess and improve the effectiveness of the SP system in a production environment. While the theoretical peak performance of supercomputers can be amazingly fast, that capability does not always represent real-world computing. To ensure that the new NERSC system is well-suited to the workaday world, NERSC and IBM have agreed to develop and test an "ESP" (Effective System Performance) benchmark for the new computer. This set of tests will measure how well the SP — and other MPP machines — deliver scientific work under a realistic workload.

Berkeley Lab Chooses Oakland as Site for New Facility

Construction is under way on Berkeley Lab's Oakland Scientific Facility, which will be the future home of NERSC's computing resources, as well as other computers operated by the Laboratory. The new center will be in downtown Oakland. The lease calls for 27,000 square feet of usable space, with an option to add another 35,000 square feet.

Berkeley Lab began looking for a suitable site for the new facility in early 1999, once it became clear that the increasing space and power demands of the new systems could not be reasonably accommodated on the Laboratory site. The new center is scheduled to open by fall of 2000.

New Allocations Policy Increases Peer Review, Broadens Access

In February 1999, the Department of Energy announced a new policy of broader scientific peer review in allocating use of NERSC resources. As proposals are submitted, they are subjected to peer review to evaluate the quality of science, how well the proposed research is aligned with the mission of DOE's Office of Science, and the readiness of the specific application and applicant to fully utilize the computing resources being requested.

The new policy is also expected to foster "start-up" or special projects that show promise. These are one-time allocations aimed at helping new projects get started, with a goal of applying for more time on NERSC's computers the following fiscal year through the normal allocations review process.

The *NERSC Program Advisory Committee* is responsible for the new scientific peer review process for proposals to access the facility's computers. This new process is used to allocate 40 percent of NERSC's computing resources. Members of the committee held their first meeting at

Berkeley Lab in September and spent two days discussing and deciding which proposals would be allocated time.

The peer review and resource allocation process for the remaining 60 percent of NERSC's computing resources is managed directly by the programs in the department's Office of Science, reflecting their mission priorities. Because DOE is a mission agency charged with carrying out specific programs related to national needs, the majority of NERSC's resources are focused on large-scale computational science programs.

To provide overall policy direction to the center and to help chart its future, Berkeley Lab established a *NERSC Policy Board*. The board reports directly to Lab Director Shank, following the approach of other major DOE facilities. The board, currently chaired by Al Narath, retired director of Sandia National Laboratories, met for the first time in December 1999 and submitted its report in January 2000.

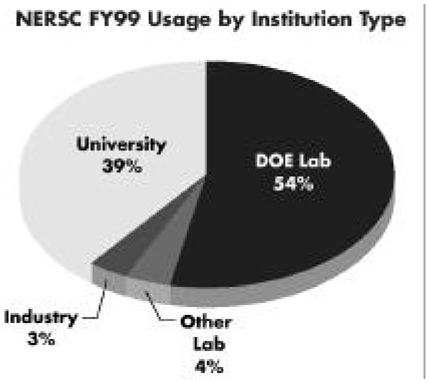
"It is clear that the LBNL management of the Center has succeeded in establishing an effective computational user facility for the various constituencies supported by SC. One of the key success factors has been the creation of an intellectual underpinning for the Center. This underpinning has not only provided valuable direct support to the NERSC user community. It has opened a window to the larger, and very dynamic, world of mathematics and computer science, thereby ensuring that NERSC can remain at the cutting edge of high-performance computing. The Board judges the balance between computing and intellectual services to be near optimum at this time," the board wrote in its report.

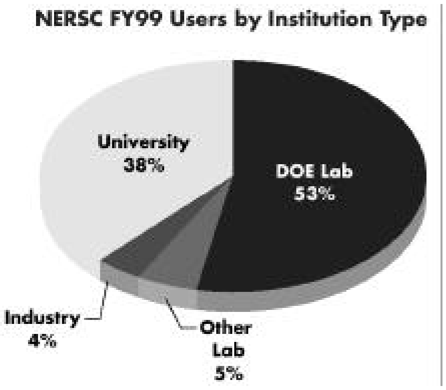
A Statistical Snapshot of NERSC Users

When talking about meeting the computational science needs of NERSC users, 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 users at DOE national laboratories, who account for 54 percent of the center's use. Universities account for 39 percent, other labs claim 4 percent and industry use is about 3 percent. By calculating usage by users at different types of institutions, the picture shifts a little. Users at DOE labs account for 53 percent of the total use, and university users account for 38 percent. Industrial users total 4 percent, and other federal users are 5 percent. These figures are based on FY99 allocations and use.



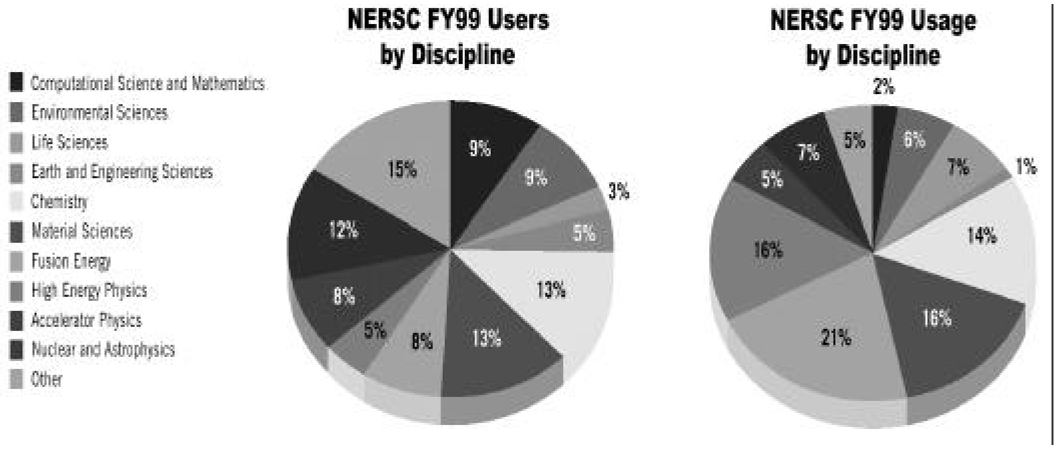


By National Laboratory

NERSC provides computational resources to users at more than 15 national laboratories. The largest number of lab accounts 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 accounts for 9 percent of the accounts. Argonne, Brookhaven and the Stanford Linear Accelerator Center each have 3 percent of the accounts, and DOE and the Naval Research Laboratory each have 2 percent. The National Renewable Energy Lab, NASA, Thomas Jefferson Accelerator Facility, Ames Laboratory and National Center for Atmospheric Research each have 1 percent.

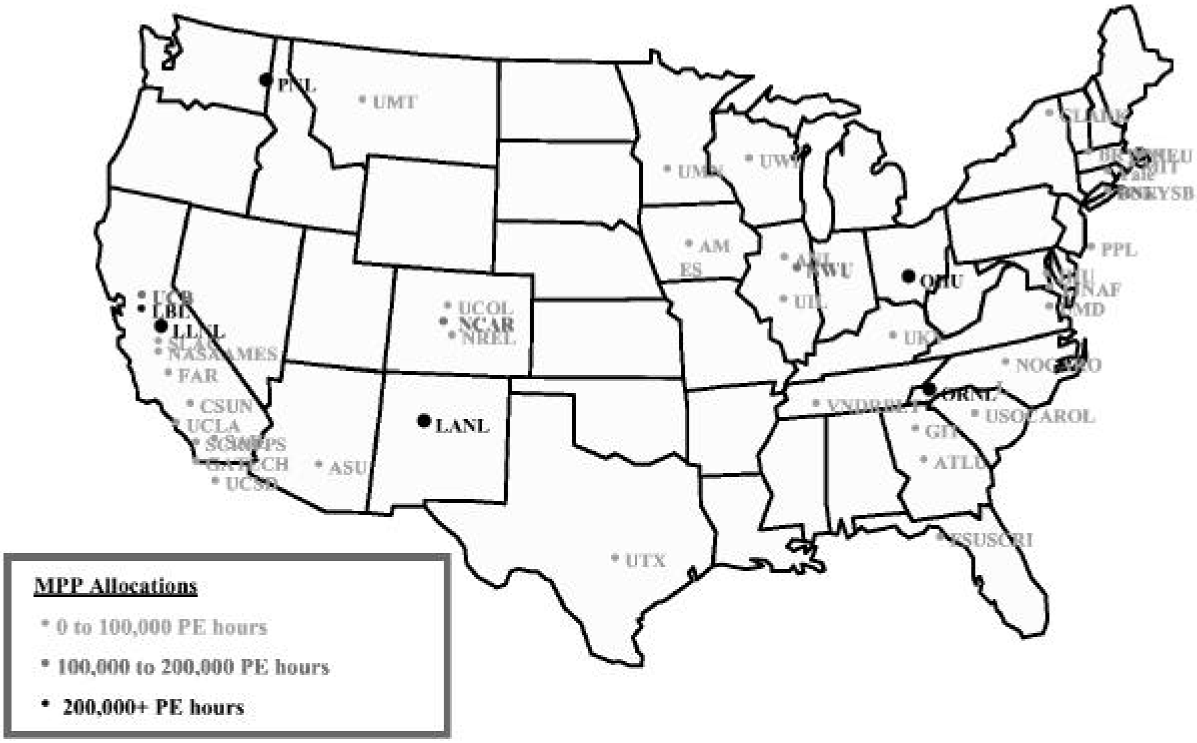
By Scientific Discipline

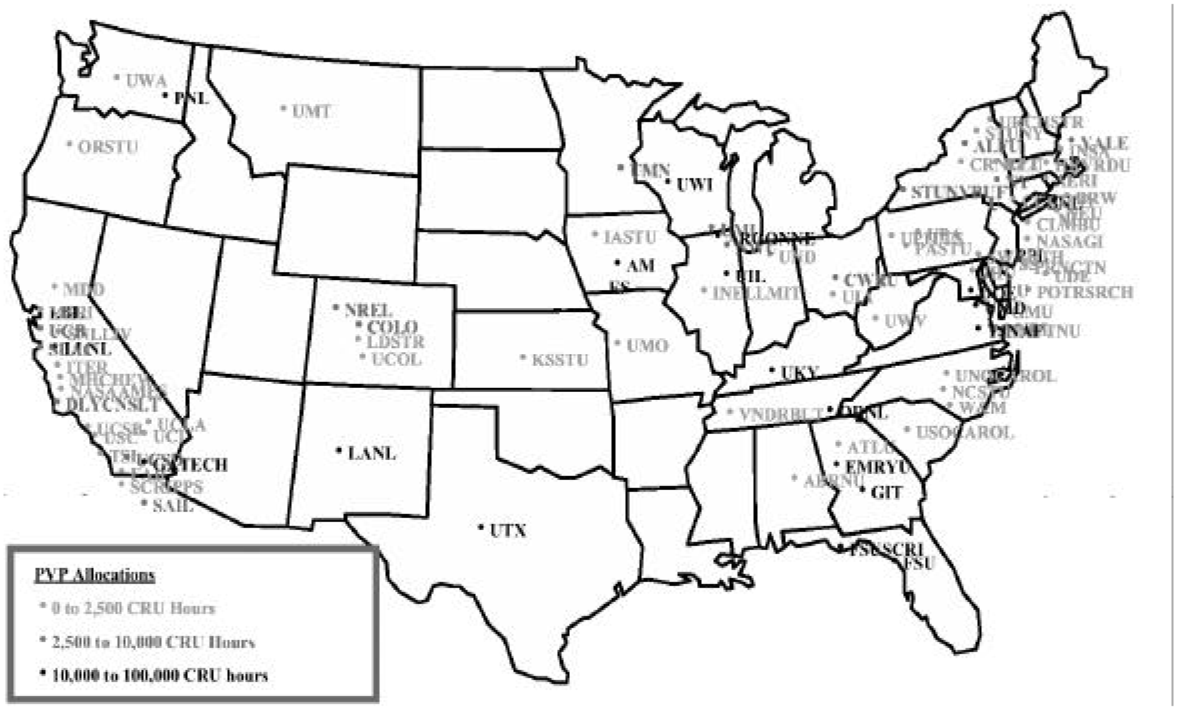
NERSC supports research across the scientific spectrum of programs in DOE's Office of Science. The breakdown of the total number of users by scientific disciplines shows Computational Science and Mathematics, 9 percent; Environmental Sciences, 9 percent; Life Sciences, 3 percent; Earth and Engineering Sciences, 5 percent; Chemistry, 13 percent; Materials Sciences, 13 percent; Fusion Energy, 8 percent; High Energy Physics, 5 percent; Accelerator Physics, 8 percent; Nuclear and Astrophysics, 12 percent; and Other, 15 percent. As to system usage by discipline, the breakdown is Computational Science and Mathematics, 2 percent; Environmental Sciences, 6 percent; Life Sciences, 7 percent; Earth and Engineering Sciences, 1 percent; Chemistry, 14 percent; Materials Sciences, 16 percent; Fusion Energy, 21 percent; High Energy Physics, 16 percent; Accelerator Physics, 5 percent; Nuclear and Astrophysics, 7 percent; and Other, 5 percent.



By Geographical Distribution

NERSC now serves users in 40 states. 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 systems, while the second map shows institutions which have allocations for using the parallel vector machines, the Cray SV1s and J90SE.


PNL
UMT
UMN
UWI
CLARK
AM
ES
ANL
NWU
OHU
UIL
UKY
PPL
NAF
UMD
UCB
LBL
LLNL
SLAC
NASAAMES
FAR
UCOL
NCAR
NREL
CSUN
UCLA
SCRIPPS
CALTECH
UCSD
ASU
LANL
NOC
ORNL
VNDRBLT
USOCAROL
GIT
ATLU
FSUSCRI
UTX
NYSB
MPP Allocations
* 0 to 100,000 PE hours
* 100,000 to 200,000 PE hours
* 200,000+ PE hours


UWA
PNL
UMT
ORSTU
UMN
UWI
IASTU
AM
ES
ARGONNE
UIL
INELLMIT
UND
CWRU
UWV
UKY
UMO
KSSTU
NREL
COLO
LDSTR
UCOL
LANL
UTX
MDD
LBL
UCB
LLNL
ITER
NASAAMES
DLYCNSLT
UCSB
UCLA
USC
CALTECH
SCRIPPS
SAIL
STUNY
ALFU
CRNELL
YALE
RDU
STUNYBUF
UPENN
PASTU
NASAGI
CLMBU
UDE
POTRSRCH
UMU
NAF
UNCAROL
NCSTU
WAM
USOCAROL
VNDRBLT
ORNL
ATLU
EMRYU
GIT
ABRNU
FSUSCRI
FSU
PVP Allocations
* 0 to 2,500 CRU Hours
* 2,500 to 10,000 CRU Hours
* 10,000 to 100,000 CRU hours

Setting Goals to Meet Users' Needs and Expectations

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

NERSC User Survey Finds Satisfaction Is High — and Getting Higher

For the second time since moving to Berkeley Lab, NERSC asked users to participate in a survey seeking feedback about every aspect of NERSC's operation. The survey was intended to help judge the quality of services, point to areas NERSC can improve, and show how NERSC compares to similar facilities.

This year 177 users responded to our survey, compared with 138 last year. In general, user satisfaction with NERSC was rated higher this year than last year, by an average of 0.6 on a 7-point scale across 27 questions common to the two years. Our overall satisfaction score was 6.25, an increase of 0.8 over last year. The biggest increases in satisfaction were with the allocations process, the HPSS system and the T3E.

Survey results are discussed further in the "Progress Measurement" section of this report.

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 groups within NERSC organization must work together to provide users with both the high-performance computing systems and the expert services for achieving research goals. To achieve this, NERSC takes a two-pronged approach. First, the NERSC staff is continually seeking out new techniques and technologies to anticipate and meet users' needs. Second, when a problem arises, we respond promptly to acknowledge, address and correct it.

NERSC Achieves Breakthrough 93 Percent Utilization on Cray T3E

In April 1999, the NERSC Computational Systems Group completed the final acceptance tests for the Cray T3E, completing an almost two-year effort to meet all conditions of the original purchase agreement. Key to completing the tests was the successful implementation of SGI's "psched" scheduling daemon. With all the features of psched running, and with NQS and "prime job" control scripts written by Computational Systems Group staff, the T3E has posted utilization figures of more than 93 percent, a level usually associated with capacity SMPs.

"This is an unprecedented level of utilization of a massively parallel machine in a general purpose computing environment," said group lead Jim Craw. "These impressive results are due, in part, to a lot of hard work by Mike Welcome, Brent Draney and Tina Butler of the Systems Group and by Steve Luzmoor and Bryan Hardy of Cray."

Efficiently scheduling a large MPP system is difficult. On the T3E, parallel applications are required to run on logically consecutive processors. Also, in the past, only one application could run on a range of processing elements in order to ensure synchronous scheduling. As applications entered and exited the system, the range of available processing elements would fragment, creating many small groups of unused processors. After a while, only small jobs could enter the system.

The psched load balancing feature automatically migrates running applications to collapse small holes of processors and thereby create large regions of available processors to run larger jobs. The psched gang scheduler will allow more than one application to run on a range of processors and will schedule them so that one application will run for awhile with complete control of the processors while the other is suspended. After a "time slice" is up, the applications switch roles and the suspended application gets to run. Another new feature of psched is the ability to designate a job with "prime" status. A prime job will preempt any other (non-prime) application and will be given preferred launching status to get in and running as soon as possible.

NERSC suggested to the T3E development staff a change to the ability to limit the amount of interactive work on the system at any point in time. In the past, this was achieved by designating a collection of processing elements to batch-only work, and a second region to interactive work. If there was no interactive work on the system at the time, the processors would idle. Now that all processors are available to run batch and interactive work, the staff schedules the entire system with batch work and allows interactive users to run with prime status. Interactive work is limited to 132 processors at a time.

Mike Declerck and Mike Welcome of NERSC and Steve Luzmoor of Cray developed a sophisticated set of PERL scripts to manage the batch system and control prime jobs. The scripts control queue and job activation such that Grand Challenge work runs in the evening, large jobs at night, and smaller jobs during the day. The scripts use the checkpoint/restart feature to halt one job mix and start the next.

The overall project was the result of an aggressive testing and problem reporting effort carried out with the help of on-site SGI/Cray analysts. In addition, forty UNICOS/MK kernels were built, loaded and booted on the system as NERSC continues to test software and bring new services into production. The resulting success benefits both NERSC and other centers with T3Es.

Providing Users with Reliable Resources in a Timely Manner

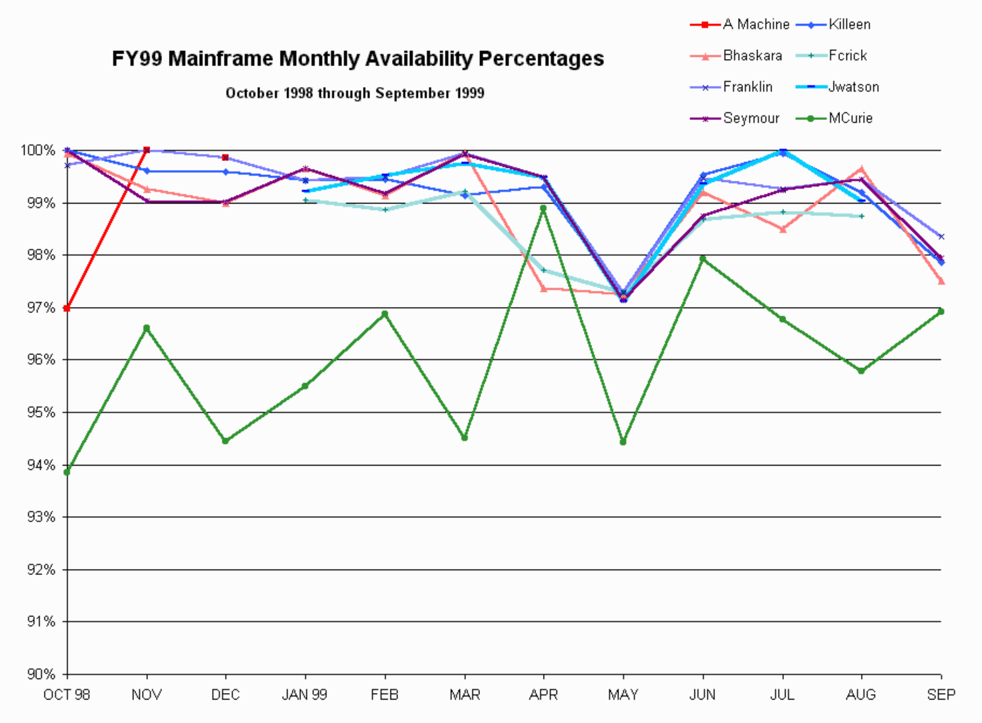
As the Department of Energy's largest unclassified computing center, NERSC provides computing systems and services to thousands of users day in and day out. Because requests for these resources exceed our capacity by a factor of four, NERSC staff do whatever it takes to keep our systems at the highest availability possible. This means that scheduled downtime is minimized and repair and restoration efforts after unexpected downtimes are given high priority. 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.

Note: Overall availability percentages are calculated using gross wall clock time. Outages/periods of unavailability (downtime) are subtracted from the total wall clock hours (accrued from beginning of FY99 or instantiation of the system) and presented as the percentage of the time the system was available to the users. The downtimes are scheduled and unscheduled periods of non-availability. Scheduled availability refers to the gross time minus scheduled outages.

SYSTEM METRICS FOR FY99
Measured (Goal)

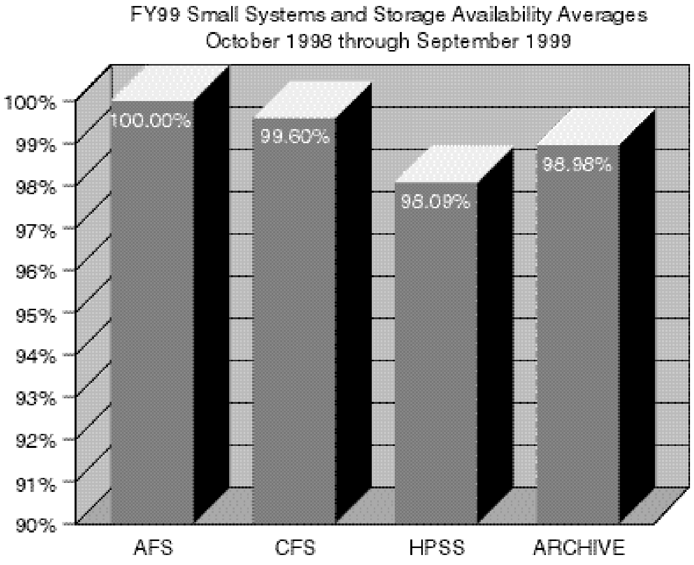
	% Availability		**Mean Time Between Interruptions**	**Mean Time to Repair**
Systems	**Scheduled**	**Overall**	**(Hours)**	**(Hours)**
Vector Systems	**99.58** (96)	**99.04** (95)	**361** (96)	**3.2** (4.0)
Storage Systems	**99.39** (96)	**98.63** (95)	**169** (96)	**2.8** (4.0)
Parallel Systems	**97.83** (96)	**96.02** (95)	**81** (96)	**3.1** (4.0)
Workstations				
Servers (fs/gw)	**100** (96)	**100** (96)	**NA** (340)	**NA** (8.0)
Clusters	**99.7** (96)	**99.5** (95)	**532** (4:0)	**1.7** (8.0)

This graph shows the monthly availability of each of the NERSC mainframe computers for the year from October 1998 through September 1999.



(Note: For detailed notes on this graph, go to http://hpcf.nersc.gov/computers/stats/AvailStats/FY99SysAvailStats.html#MainframeMonthly.)

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



Responding to Users' Requests for Assistance

NERSC's consulting staff is committed to responding promptly to all requests for help from our users. Our primary goal is to provide an initial response to all requests within four working hours. NERSC also strives to resolve as many requests as possible within two working days (at the most). Spot checks confirm that NERSC meets the goal of responding to problems within four hours. Moreover, during calendar year 1999, a total of 3,204 "trouble tickets" were created. Of these, 367 took more than two days to close. This translates to an 88.5 percent closure within the two-day target.

The annual user survey, in fact, found that NERSC's consulting services rated high in user satisfaction. On a scale from 7 (very satisfied) to 1 (very dissatisfied), the score for timely response to consulting questions was 6.6. Users also responded that one of the areas they are happiest with is consulting services.

Not all problems can be resolved within two days, and in some cases – in consultation with the affected users – these problems are put "on hold" as longer-term solutions are needed. Trying to resolve these problems quickly may improve the timeliness metric while shortchanging the users. 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. In 1999, for example, the NERSC staff identified 37 of the longer-term problems as being caused by bugs in the Cray software, which were forwarded to the company as Software Problem Reports, or SPRs.

Problems not resolved within 72 working hours are escalated for more in-depth review and NERSC managers receive weekly notification of all outstanding trouble tickets. This process ensures that outstanding problems are kept active and don't fall off the screen. 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.

A New Approach Proposed for Measuring Help-Request Response Times

In 1999, the User Services Group proposed a new approach to measuring the response to users' inquiries and problems. The new approach calls for different types of requests to be handled under different metrics. For example, some calls can be resolved in minutes over the phone, while others may require special actions by the NERSC staff or hardware vendor and this could take days, weeks or even months, depending on the nature of the problems. Given the broad variety of help requests cited above, it is not practical to expect them all to fit under the same metric – but there should be targets for time to closure whenever possible. The following categories of requests and associated metrics cover most or the requests received by NERSC:

Consulting Assistance Requests

These include help with coding, general assistance with libraries or data migration, and diagnostic assistance about a network connection, job status or compiler behavior. Such requests are typically handled entirely by the consulting staff, requiring little or no input from other parts of the NERSC organization. User Services proposes to institute a rough triage relating to the amount of the consultant's "thinking time" that must be devoted specifically to dealing with the problem:

- Short problems, requiring an hour or less of the consultant's time, should generally be closed within two business days;
- Medium problems, requiring between an hour and a day of the consultant's time, should be closed within five business days (normally one calendar week); and

- Long problems, requiring more than a day of the consultant's time, should be targeted for closure within a month.

Long-Term Collaborations

Some help requests become more extensive cooperative arrangements, in which NERSC personnel may be helping with algorithmic or logisitical issues on an ongoing basis. These requests could ranging from porting a large application program to run on the T3E architecture for the first time to a problem requiring new algorithms that are developed cooperatively with NERSC. Under the new metrics approach, the long-term collaborations are exempted from closure targets, since they may go on for months or years. User Services suggests that the existing collaborations at any one time be enumerated and tracked in monthly NERSC status reports.

System Service/Fix Requests

These are distinguished from the above requests by the fact that NERSC hardware or software must be changed or configured to address the request. This aspect usually means that cooperation from NERSC personnel outside User Services is necessary for a successful conclusion.

Routine service/fix requests involve some ordinary service or easy fix regularly performed by NERSC personnel that poses no unsolved technical problems or undecided policy issues, such as replacement of an accidentally deleted file or killing a runaway batch job. Routine service/fix requests should ordinarily come to closure within two working days. If some such requests are regularly taking longer, then we need to fix the process by which those requests are provided or decide to stop providing the troublesome service.

Complex service/fix requests are more difficult and usually either unsolved technical problems, unresolved policy issues or management decisions. Examples of complex service requests include creation of a special queue, purchase of a new piece of commercial software, or reconfiguration of the batch system on a machine to address a recognized problem or inequity. For complex service/fix requests, NERSC proposes a standard of one month for either a decision or a plan and/or workaround, with an update to the client within two weeks. By the end of the one-month period, the client should fully understand what NERSC will or will not do to address the request. There may be additional time to the fulfillment of a service request or completion of a fix if difficult technical issues must be faced. In this case the client should be given a completion target time when the positive decision is rendered.

Vendor Fix Requests

Requests may result in the identification of a weakness or deficiency in a vendor's product. These are similar to system fix requests, but NERSC may be unable to motivate the vendor to address the problem expeditiously. Fulfillment of these requests is frequently outside NERSC's control, so no single rule for fix completion can be set. However, NERSC can and should establish goals for the frequency with which an update is sought by NERSC (monthly seems about right). In some cases, contractual arrangements may make it possible to have time deadlines for response to a fix request, allowing some goals to be set for completion on a case-by-case basis.

NERSC believes that this new approach, combined with our longstanding commitment to a prompt response to every request, will ensure that the center's consulting services continue to satisfy users' needs.

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 being 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. Here are a couple of examples of our innovative service:

Special Queues to Meet Special Requests

NERSC strives to provide a fair and productive balance between meeting users' specialized computing needs and maximizing the availability and productivity of the center's resources. On a limited, by-request basis, NERSC staff members are able to utilize narrow windows of system availability to accommodate users' jobs which exceed system limits or scheduling parameters.

On the PVP batch systems, jobs are primarily limited by the amount of memory required and the length of computing time needed. The systems are set up for different mixes of these factors: typically longer jobs are run using less memory and those requiring more memory are given a shorter run time. NERSC staff members are constantly trying to find the ideal mix to achieve high utilization of the machines and prompt turnaround of jobs for users.

When a user comes in with a job that doesn't fit within the specified limits, or needs to have data for a conference presentation, the NERSC staff strives to accommodate the request without negatively impacting other users. To accommodate these special requests, User Services monitors the status of jobs on the systems and, when enough memory becomes available, can start the special job manually using a special queue. Jobs needing exceptionally long run times can also be accommodated occasionally, as when idle time on the interactive J90 is utilized to handle a job of more than 100 CPU hours.

Using the Right Solver Makes Shorter Work of Sparse Matrices

The old adage "It's not what you know — it's who you know" was borne out by a team of researchers using NERSC's Cray T3E to solve a fundamental problem of quantum physics — a complete solution to scattering in a quantum system of three charged particles. Using the Cray T3E-900 at NERSC, collaborators at Berkeley Lab, Lawrence Livermore National Laboratory and UC Davis obtained a complete solution of the ionization of a hydrogen atom by collision with an electron, the simplest nontrivial example of the problem's last unsolved component (see the chapter on High Performance Computing Center Leadership for more information about this research). Their findings were presented in the cover story in the Dec. 24, 1999, issue of Science magazine — and they credited the assistance of NERSC's Xiaoye "Sherry" Li in their article.

Sherry Li first discussed the problem, and possible means of solving it, with Bill McCurdy when she interviewed for a job at NERSC in 1996. McCurdy and his fellow researchers, all physicists or chemists, needed linear algebra software to solve the problem. However, they were using LAPACK, which is used for studying dense matrices, to study a problem of sparse matrices.

A sparse solver would be more efficient and save them a lot of time and memory, Sherry noted. The problem with using a dense solver is that it stores all the zeroes of a sparse matrix, which quickly reaches the computer’s memory limit. Since the group’s biggest problem was a 5-million-by-5-million matrix, every efficiency was essential. That’s where Sherry’s expertise entered the picture.

For her thesis at UC Berkeley, she developed SuperLU, a sparse solver. It takes its name from the fact that it uses “supernodal” technique to perform LU factorization. But since it was a relatively new mathematical software, many researchers hadn’t heard of it and still use 20-year-old software, which runs more slowly. SuperLU is not only faster, but since it uses less memory, significantly increases the size of problems which can be studied.

Since joining NERSC, Sherry developed a parallel version of SuperLU. She helped the group integrate SuperLU into their chemistry modeling code so they could run problems on NERSC’s Cray T3E. Her efforts have also made it possible for the researchers to now run fairly large problems on their laptop computers.

Since the group achieved significant results using the software, they are now spreading the word about its value to other scientists. Team member Mark Baertschy, who wrote his thesis about the problem and its solution, has since become a post-doc fellow in Colorado, where he is already converting new users to SuperLU.

NERSC Division Director Horst Simon has noted that in the arena of high-performance computing, if you’re standing still you’re really falling behind. NERSC strives not only to be in the race, but to also set the pace. We do this by not only providing the best systems, but by also delivering our services in an innovative manner.

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 changes 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.

NERSC Web Site Reorganized to Help Users to Find Resources

In early 1999, the NERSC web site at www.nersc.gov was completely redesigned and reorganized. One of the main goals of this effort was to subdivide the web site in such a way that different audiences could find quicker routes to the information they were seeking. For the NERSC user community, this resulted in the creation of hpcf.nersc.gov, a site focusing on NERSC's High Performance Computing Facility from a user perspective. In addition to general announcements and news about NERSC's systems and services, the site offers links to the support staff, information for new users, NERSC accounts, information on policies and security, running jobs, software, storage, training and visualization services. The main page also allows users to quickly check the status of all systems.

Providing Up-to-Date User Training

To help NERSC users make the most efficient use of the center's resources and to enable the highest quality of computational science, the NERSC staff provides regular training programs utilizing a variety of techniques. These include:

Classes: Short lectures and multi-topic classes are presented at the Berkeley Lab, remote sites, or as teleconference talks. These are announced by email and web postings, and their contents are usually updated each time a given topic is presented.

Online Materials: Materials from past lectures and some longer documents, intended as stand-alone introductory texts, are available as printable files and web documents. These materials provide good starting points in any search for concise survey information. More complete, in-depth coverage is available in the topic-specific areas of the NERSC web site.

Other Resources: NERSC's Training website includes a reference list of useful references and texts, as described and recommended by NERSC Staff and users; UNIX tutorials and sources of information on the Unix Operating System, in its various versions; and links to other HPC Resources, including other high performance computing sites and documentation.

Training sessions sponsored by NERSC were presented throughout the year and include:

On Jan. 21–22, 1999, at Oak Ridge National Laboratory, NERSC User Services and Oak Ridge presented a two-day set of lectures for new and intermediate T3E users.

On March 16–17, 1999, NERSC User Services presented classes for new and intermediate users of the J-90, T3E and auxiliary systems. The classes were held at Lawrence Berkeley National Laboratory. Topics covered during the 30-minute presentations included Introductions to the Cray J-90 and T3E computers, Batch Computing, NERSC File Systems, Debugging, Performance Tuning and Monitoring, the ACTS Toolkit, and Managing Secure Connections with SSH.

In April 1999, NERSC User Services presented a series of special-topic lectures via conference-call. The lectures were one to two hours in length, and covered material relevant to users of NERSC computational facilities. The lecture materials were made available in the Training area of the NERSC website prior to the lectures.

In conjunction with the ERSUG meeting held April 27–28 at Argonne National Laboratory, NERSC User Services presented classes for new and intermediate users of NERSC computational facilities.

Subsequent teleconference lectures were held during the rest of the year. The schedule was as follows:

May: Scripting and MPI
June: MPI and Make
July: MPI Tools: Tau and Vampir; Scripting
November: Introduction to Systems: T3E and J90
December: Introduction to Systems: SSH; File Systems; Batch

Getting the Word Out About NERSC Systems

NERSC uses several methods to inform users about changes and updates in its systems. These methods include email notification, Messages of the Day (MOTD) posted on the NERSC High Performance Computing Facility website, and announcements pertaining to specific systems. Here’s a breakdown of how many notices were issued using these methods.

Email notification: This is reserved for the most important announcements affecting a large number of users, such as a new system coming on line or a planned outage of all systems. These announcements are also archived on the announcements website.

Messages of the Day: Announcements of canceled downtimes, maintenance and other system interruptions are posted as they are issued, and can result in multiple postings on any given day. An estimated 500 MOTDs were issued in 1999.

Systems announcements: Changes and updates in specific systems are announced by posting them on the system-relevant web pages and at http://hpcf.nersc.gov/news/announcements/. In 1999, a total of 127 changes were announced in this way. The number of postings for each area were: 20 notices for the Cray T3E, 30 announcements for the PVP (J90 and SV1) systems, 14 notices for the Cray Programming Environment, 26 notices for the Programming Library, 22 announcements for Software Applications and Tools, seven notices for File Storage Systems, and eight general announcements.

NERSC’s Efforts Are Making a Difference to Users

NERSC’s commitment to providing users with timely and accurate information is working, according to the annual user survey. Results from the 1999 survey found that 95 percent of users said they felt adequately informed, compared with 82 percent in the 1998 survey.

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. NERSC now provides users with a 604-processor IBM SP, a 696-processor Cray T3E, three Cray SV1s, one Cray J90se, the PDSF cluster (112 CPUs and eight disk vaults with file servers) and the HPSS storage system. 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 has established the Advanced Systems and Future Technologies groups. These groups' 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.

NERSC Selects IBM SP for Next Supercomputer

In April 1999, NERSC announced that it had selected an IBM RS/6000 SP system as the center's next-generation supercomputer. The IBM system was chosen based on its ability to handle actual scientific codes, and tests designed to ensure the computer's capability as a full-production computing system in NERSC. These tests indicated that the system, when fully installed, will provide five or more times the total current computational power of NERSC, already one of the most powerful supercomputing sites in the world. This agreement, a fixed-price, five-year contract for $33 million, was the largest single procurement in the 68-year history of Berkeley Lab.

The new system, which incorporates IBM's newest processor and interconnect technologies, is being installed in two phases. When completed, the system will increase NERSC's computing capabilities by more than 400 percent. Phase I installation, which began in June 1999, consists of an RS/6000 SP with 304 of IBM's new two-CPU POWER3 SMP nodes. In all, Phase I has 512 processors for computing, 256 gigabytes of memory and 10 terabytes of disk storage for scientific computing. The system has a peak performance of 410 gigaflops, or 410 billion calculations per second.

Phase II, slated for installation no later than December 2000, will consist of 152 16-CPU POWER3+ SMP nodes, utilizing an enhanced POWER3 microprocessor. The entire system will have 2,048 processors dedicated to large-scale scientific computing. The system will have a peak performance capability of more than 3 teraflops, or 3 trillion calculations per second.

As part of the purchase contract, NERSC is working with IBM to develop computer-utilization benchmarks and methods to assess and improve the effectiveness of the SP system in a production environment. While the theoretical peak performance of supercomputers can be amazingly fast, that capability does not always represent real-world computing. To ensure that the new NERSC system is well suited to the workaday world, NERSC and IBM have agreed to develop and test an "ESP" (Effective System Performance) benchmark for the new computer. This set of tests will measure how well the SP — and other MPP machines — deliver scientific work under a realistic workload. (For more details on ESP, see the HPCC Leadership section of this report.)

Realizing the Potential of PC Clusters for Scientific Computing

For a number of years now, clusters of desktop computers have been heralded as the supercomputers of the future. Despite wide-ranging development programs, clusters continue to remain futuristic ideals. Tools for building and managing clusters are often the software equivalent of duct tape — ad hoc solutions that require significant attention and address symptoms without providing a permanent solution. Research groups that are attracted to clusters by low hardware costs find that they can spend many times the hardware cost on personnel to configure and manage the cluster. Those considering using clusters for production have found that features usually found in high-end systems, such as robust accounting, quotas, and security, do not exist.

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 scalable, full-featured environment similar to that of a traditional supercomputer. As a rule of thumb, software not designed from the beginning to be scalable cannot be made scalable by adding features — scalability has to be designed in from the beginning.

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.

Software being developed in the PC cluster project is being made available as the Berkeley Lab Distribution (BLD — see http://www.nersc.gov/research/bld). BLD development projects and products include:

- M-VIA 2, a second-generation, 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. It allows an application to get direct user-level access to the network, bypassing the operating system and the overhead associated with protocols such as TCP/IP. VIA will widen the space of parallel applications that can be efficiently executed by clusters, and has many commercial applications. The primary advantage of VIA is its strong industry backing, and the fact that it will be widely supported by hardware.
- MVICH, an implementation of MPI (Message Passing Interface) on VIA. MPI is the critical VIA "app" for high performance computing on clusters.
- 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.

DOE Establishes Probe Testbed for Storage-Intensive Applications

Scientific researchers are generating staggering volumes of data, with data sources ranging from climate simulations to experiments in high-energy physics. Add projects such as human genome mapping, with massive demands for rapid user access, and the need for optimizing data storage and retrieval is clear.

NERSC and Oak Ridge National Laboratory (ORNL) are tackling the storage challenge with Probe, a newly established distributed testbed for storage-intensive applications. It combines high-speed networking technology with the High Performance Storage System (HPSS). Launched in mid-1999, Probe will have significant installations at ORNL and NERSC, providing access to researchers around the country.

The Probe testbed is available for researchers from the scientific community to perform comparative evaluations of the latest technologies in storage hardware and software. By linking the two testbed systems together over the network, researchers will be able to evaluate the effects of network latency in remote storage access and develop new protocols for effectively using distributed storage systems. The testbed will also provide a platform for the developers of new storage and networking hardware and software to test their devices in high-demand facilities.

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 gave 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 (recently renamed Cray Inc. after its acquisition of Cray from SGI). 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. 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 as well as NERSC’s participation in the system evaluation.

One of the first products of NERSC’s evaluation of the Tera system earned the “Best Paper of SC99” award for Leonid “Lenny” Oliker of NERSC and Rupak Biswas of NASA Ames. Their paper, “Parallelization of a Dynamic Unstructured Application Using Three Leading Paradigms,” compared programmability and performance of their implementation of a mesh adaptation code on the Cray T3E, SGI Origin2000 and Tera MTA computers. They compared several critical factors of parallel code development, including runtime, scalability, programmability and memory overhead. Their overall results demonstrate that multithreaded systems offer tremendous potential for quickly and efficiently solving some of the most challenging real-life problems on parallel computers.

The award-winning paper can be found on the web at: http://www.nersc.gov/~oliker/paperlinks.html.

Preparing for the Computational Grid

In the last two years, the vision of a computational grid has gained broad acceptance. The grid is envisioned as a unified collection of geographically dispersed supercomputers, storage devices, scientific instruments, workstations, and advanced user interfaces, all operating over the Internet. As a partner in eight DOE networking and grid research projects funded in FY99, NERSC will be a key player in programs aimed at making the grid a useful tool for scientific experimentation and collaboration. Grid research is developing advanced networking technologies and revolutionary applications that require advanced networking, and is demonstrating these capabilities on testbeds that are 100 to 1,000 times faster end-to-end than today's Internet.

Berkeley Lab is a partner in three programs which support widely distributed visualization, and is the lead lab in two of them — developing a prototype environment for remote, collaborative visualization of large combustion simulation datasets, and developing visualization-sensitive network protocols. Current models for combustion modeling are not capable of scaling up to handle the huge amounts of data expected to be generated by the next generation of supercomputers. Being able to utilize the larger datasets, however, is important for helping collaborating researchers to develop insight into mechanisms of combustion and then apply these to solving engineering problems such as building cleaner, more efficient diesel engines.

One product that has already emerged from the Combustion Corridor project is Visapult, an image based rendering assisted volume rendering (IBRAVR) tool. Developed by Wes Bethel of the NERSC Visualization Group, Visapult enables distributed visualization of large data volumes, such as two gases mixing in a turbulent environment, on remote workstations. The fundamental idea behind Visapult is that large datasets are partially prerendered on a large computational engine close to the data, then final rendering is performed on a local workstation. Sharing the workload between a remote multiprocessor machine and the local workstation allows for some degree of interactivity on the local workstation without the need to recompute an entirely new image from all the data when the object is rotated by a small amount. IBRAVR was demonstrated at SC99 using data from two simulations, one involving a combustion modeling code and a second one from a cosmology model.

NERSC's expertise in grid technologies received another boost when Berkeley Lab's highly respected Data Intensive Distributed Computing Research Group, led by Bill Johnston, joined the NERSC Division as our Distributed Systems Department. The Distributed Systems Department conducts research and development into various components of the grid infrastructure, including collaboratory tools, computer security, distributed data intensive computing, and networking.

The broad-based expertise within NERSC's High Performance Computing Research Department and the new Distributed Systems Department positions us at the forefront of grid research and development, and ensures that our clients will be among the first to reap its benefits.

WISE TECHNOLOGY INTEGRATION

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

In an age when high tech firms constantly tout “New and Improved!” products and services, NERSC takes something of an old-fashioned approach. Newer isn’t always better, and merely installing the latest technology doesn’t ensure that it will meet users’ needs. To ensure that the systems it provides to clients are reliable, NERSC thoroughly evaluates the technology and tests systems off-line before deciding whether to implement the system for production, or file the results away under “lessons learned.” A common aspect of the later stages of such evaluations is inviting specific NERSC users with demanding applications to help put systems through their paces to see how equipment stands up to real-world demands. Feedback from users then helps the staff further improve the systems and services.

Here are some examples of this wise technology integration as practiced at NERSC during the past year.

NERSC Completes Conversion of Storage to HPSS

In January 1999, NERSC completed the conversion of its two archival data storage systems to High Performance Storage Systems, or HPSS. For the preceding three years, NERSC’s Mass Storage Group had been working to install new hardware, consolidate storage systems and provide clients with new capabilities. The primary goal has been to convert the two archive systems — CFS and UniTree — to HPSS. Because NERSC is one of the five HPSS development sites, we were well positioned to make the conversion without interrupting service or losing data. By phasing the transition, NERSC staff ensured that the change could be made with minimal inconvenience.

The High Performance Storage System (HPSS) is a modern, flexible, performance-oriented mass storage system, designed and developed by a consortium of government and industry. Users are given HPSS accounts to enable them to economically save and access files that are too large or too infrequently used to be kept in personal home directory spaces. HPSS is fundamentally different from a “normal” computer file system, and must be used and accessed differently. There are no file size limitations to its use, but there are practicality constraints imposed by the limited bandwidth that serves it.

Increasing capacity and automation of NERSC’s storage system while the center was at LLNL made it easier to move it all to Berkeley in 1996. The entire archive had been condensed to fit into three of the six silos before the transition. This allowed the three empty silos to be moved into place, so that only the tape cartridges had to be moved to bring the system back on line. This allowed the storage move to be accomplished in just three days, instead of weeks or months.

Since moving to Berkeley Lab, NERSC’s archival storage capacity has been increased from 30 terabytes to 750 terabytes – a 25-fold increase.

With HPSS, NERSC helped develop and regularly assess all aspects of the system, ensuring that it would meet the needs of users in a demanding production environment. This work to continue developing storage technologies continues in such projects as Probe, a partnership described in the New Technologies section of this report.

The Parallel Distributed Systems Facility (PDSF) Builds for the Future

From the computer cast-offs of the now-canceled Superconducting Supercollider (SSC) has risen a powerful, flexible and reliable cluster-computer system supporting a variety of high energy and nuclear physics research across the country and in Europe. Called the Parallel Distributed Systems Facility, or PDSF, the system is a “Cluster of Clusters” managed by NERSC’s High Energy and Nuclear Physics Systems Group, with help from the NERSC Computational Systems Group, and the Berkeley Lab Physics and Nuclear Sciences Divisions.

The PDSF is a networked, distributed computing environment used for the detector simulation and data analysis requirements of large-scale High Energy Physics and Nuclear Science investigations. The system originated from the Particle Detector Simulation Facility for the SSC. The obsolete PDSF equipment was thoroughly reconfigured with new workstations featuring Pentium processors and the Linux operating system. Although the technology has been completely updated, the PDSF system will provide functionality well beyond what was originally intended and is expected to provide a strong computing platform for numerous future HENP experiments. To accommodate the massive amounts of data to be generated by the various experiments, PDSF has been provided with 3 terabytes of networked disk space in Linux data vaults, and been integrated with NERSC’s HPSS. During the course of 1999, CPU power on PDSF was more than doubled and disk space was increased fourfold.

PDSF now serves 20 different experiments and research groups with more than 300 users carrying out large-scale computational requirements of High Energy Physics and Nuclear Science investigations. 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. NERSC will continue to learn from and evolve the PDSF as a growing, expanding computer resource targeted at the unique computing needs of HENP experimental physics programs.

Cray SV1 Computers Pass NERSC Acceptance Tests With Flying Colors

In August 1999, NERSC officially accepted three Cray SV1 computers, which perform on average three times as fast as the Cray J90se machines they replaced. NERSC was one of the first centers to receive certified production models of the SV1. In order to pass the acceptance tests, the SV1s had to run without interruption and be available more than 99 percent of the time for 30 days. After a week of staff testing and another week of use by selected NERSC users, the machines were made available to the general NERSC community on July 27. The 20-processor machines “Seymour” and “Bhaskara” posted availability figures of 99.73 percent and 99.88 percent respectively. A third machine, the 28-processor “Franklin,” was available 100 percent of the time.

The SV1 is Cray’s newest computer and is described as the company’s first scalable vector supercomputer, combining powerful processors with the “scalability” features necessary to link large numbers of processors together

Stephen Jardin, the principal investigator at the Princeton Plasma Physics Laboratory, has been a long-time user of the NERSC supercomputers and is presently chair of NERSC’s Program Advisory Committee. He and others in his field have been anxiously looking forward to using the SV1 since its procurement was announced. “Many of our most complex design and simulation codes employ algorithms that are particularly well-suited to the SV1’s parallel vector architecture.” said Jardin. “We expect to be able to make very effective use of these machines immediately, and are excited about the new research possibilities that they open for us.”

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 NUG meetings and post statistics covering a wide range of our systems and services 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.

User Survey Results

Essential to meaningful measurement is taking the perspective of our client community. In 1999 we conducted our second annual survey of all our users. 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 us to areas for improvement, and show how NERSC compares to similar facilities.

This year 177 users responded to our survey, compared with 138 last year. In general, user satisfaction with NERSC was rated higher than last year, by an average of 0.6 on a 7-point scale across 27 questions common to the two years. Our overall satisfaction score was 6.25, an increase of 0.8 over last year. The biggest increases in satisfaction were with the allocations process, the HPSS system and the T3E.

On a scale from 7 (very satisfied) to 1 (very dissatisfied), the average scores ranged from a high of 6.6 for timely response to consulting questions to 4.0 for PVP batch wait time. The areas users are happiest with this year are consulting services, HPSS reliability and uptime, as well as PVP and T3E uptime. Areas of concern are batch wait times for both PVP and T3E systems, visualization services, the availability of training classes, and PVP resources in general.

The areas of most importance to users are the overall running of the center and its connectivity to the network, the available computing hardware and its management and configuration, consulting services, and the allocations process. Access to cycles is the common theme.

In their verbal comments, users focused on NERSC's excellent support staff and its well-run center with good access to cycles (although users wish we could provide even more), hardware and software support and reliable service. When asked what NERSC should do differently, the most common response was "provide even more cycles." Of the 52 users who compared NERSC to other centers, half said NERSC is the best or better than other centers, 23% simply gave NERSC a favorable evaluation or said they only used NERSC, 19% said NERSC is the same as other centers or provided a mixed evaluation, and only 2% said that NERSC is not as good. Several sample responses below give the flavor of these comments.

> “I have found the consulting services to be quite responsive, friendly, and helpful. At times they went beyond the scope of my request which resulted in making my job easier.”

> “Provides reliable machines, which are well-maintained and have scheduling policies that allow for careful performance and scaling studies.”

> “Provides a stable, user-friendly, interactive environment for code development and testing on both MP machines and vector machines.”

> “It would be nice if there were fewer users, so turn-around time could be faster.”

> “NERSC provides a well-run supercomputer environment that is critically important to my research in nuclear physics.”

NERSC made several changes in 1999 based on the responses to the 1998 survey:

- We more frequently notified users of important changes and issues by email. This year 95% of users said they felt adequately informed, compared with 82% last year.
- We changed the way we present announcements on the web. (We have no comparison rating between the two years.)
- We restructured the queues on the Cray T3E. The satisfaction rating for T3E batch queue structure went up by one point (from 4.5 to 5.5).
- We added additional debug queues on all the Crays. Last year we received 2 complaints in this area, this year none.

In next year’s edition of this self-assessment, we expect to report back to you about additional changes made as a results of our 1999 user survey. Stay tuned.

The survey itself and complete results are available at http://hpcf.nersc.gov/about/ survey/99/.

Systems Statistics Web Pages

An important way NERSC holds itself accountable to our clients is by tracking monthly system statistics and posting them on the Web at http://hpcf.nersc.gov/computers/stats/. Both current and historic data are available. Statistics on system availability, downtime and outages include:

- Individual Mainframe Availability Averages
- Mainframe Monthly Availability Averages
- Small Systems and Storage Availability Averages
- Small Systems and Storage Monthly Availability Averages
- Overall Systems Availability Averages
- Overall Systems Availability Details

Other available figures include queue run and wait statistics and user account statistics.

Taken together, these figures paint a picture of an efficiently and effectively run computer center. Highlights of 1999’s statistics are given in the “Reliable and Timely Service” section of this report.

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 has continually provided leadership and helped shape the field of high performance computing since the center was established in 1974. Twenty-five years ago, the then-Controlled Thermonuclear Research Computer Center opened new frontiers in computational science by providing researchers across the nation will remote access to a supercomputer. The center also developed the Cray Time Sharing System, which became a standard at other U.S. supercomputing centers. In the mid-1980s, when the National Science Foundation was developing its own supercomputer centers, many of the systems and services developed here were adopted in San Diego and Urbana-Champaign. Finally, when DOE and NERSC reinvented the concept of the supercomputing center in 1995 by creating a model which provided both cycles and specialized intellectual resources, other centers also adopted this model.

In 1999, NERSC continued to provide leadership both in the areas of operating a national supercomputing facility and in advancing computational science. Here are some examples of NERSC's expertise and knowledge-based leadership in enhancing large-scale computing in 1999:

Making the Most of Computing Resources

NERSC's Revised Allocations Policy Increases Peer Review, Access

In January 1999, the U.S. Department of Energy announced a new policy for allocating time on NERSC's computing resources. The change makes NERSC the most open supercomputing center in the nation and reflects the ever-broadening role of NERSC since it was founded in 1974 to provide computing resources for magnetic fusion research. Since then, the facility has expanded its scope to include high-energy physics, materials science, computational biology, astrophysics, energy research, chemistry and climate modeling.

The new policy will optimize the scientific results of NERSC for the Department of Energy by opening up the facility to a wider range of proposals. As proposals are submitted, they will be subjected to peer review to determine whether:

- The quality of science and scientific promise is appropriate to utilize high-performance computing resources:
- The proposed research is aligned with the mission of DOE's Office of Science; and
- The specific application and applicant are at a stage where they are ready to fully utilize the computing resources being requested.

The revamped allocations policy also responds to suggestions by the federal Office of Management and Budget that proposals for utilizing NERSC's resources involve a broader range of peer review.

Because DOE is a "mission agency" charged with carrying out specific programs related to national needs, the majority of NERSC's resources will be focused even more on the largest-scale computational science programs, NERSC's "Class A" projects. The new procedures will better position NERSC for meeting its goal of accelerating scientific discovery through large-scale, multi-disciplinary research projects.

The new policy is also expected to foster “start-up” or special projects which show promise. These are one-time allocations aimed at helping new projects get started, with a goal of applying for more time on NERSC’s computers during the next fiscal year.

The *NERSC Program Advisory Committee* is responsible for the new scientific peer review of proposals for access to the facility’s computers. Time will be allocated based on a combination of new peer review (40 percent of resources) and direct DOE guidance reflecting its mission priorities (for 60 percent of the computing resources at NERSC). The 14-member Program Advisory Committee held its first meeting ever at the Lab in September 1999 to discuss requests from researchers for FY00 computing resources at NERSC. In all, more than 300 proposals were reviewed on their scientific merit and decisions made on how many computing hours (on the massively parallel or parallel vector machines) and how much archive storage capacity would be allocated.

The 1999 Program Advisory Committee was chaired by Steve Jardin, a long-time NERSC user at the Princeton Plasma Physics Lab. Other members are David Bailey, NERSC liaison; Ian Foster, Argonne National Lab; Steven Hammond, National Center for Atmospheric Research; Bruce Harmon, Ames Lab; Robert Harrison, Pacific Northwest National Lab; Dale Koelling, Argonne National Lab; Jean-Noel Leboeuf, UCLA; Gregory Newman, Sandia National Laboratories; Robert Ryne, Los Alamos National Lab; Shankar Subramaniam, UC San Diego; Robert Sugar, UC Santa Barbara; Doug Swesty, State University of New York, Stony Brook; and Mary Wheeler, University of Texas, Austin. NERSC staff reviewers were Craig Tull, Charles Leggett, Julian Borrill, Jonathan Carter, Andrew Canning, Osni Marques, Tom DeBoni, Chris Ding, Adrian Wong and Esmond Ng. Francesca Verdier and John McCarthy of NERSC User Services Group provided technical support for the meeting.

To provide overall policy direction to NERSC and to help chart its future, a *NERSC Policy Board* was also established. This approach follows that of other major DOE facilities, including the Advanced Light Source at Berkeley Lab. The NERSC Policy Board, established in early 1999 as part of NERSC’s revised allocation policies, and met at Berkeley Lab for its first meeting in December 1999. Berkeley Lab Director Charles Shank chartered the board to advise him on the operation and future strategies of NERSC as a DOE national facility. Among the topics open for discussion are: What opportunities does NERSC have to increase its value to the scientific community? What can be done to raise the level of NERSC’s involvement in the national computational science community and to increase the level of support of the scientific community for NERSC? What is the future of centralized computing resources, and what should NERSC’s role be in the nation’s scientific computing effort?

The members of the Policy Board are: Professor Fred E. Cohen, Department of Medicine, UCSF; Professor Larry L. Smarr, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign; Professor Robert J. Goldston, Princeton Plasma Physics Laboratory; Dr. Warren M. Washington, Climate and Global Dynamics Division, National Center for Atmospheric Research; Professor John Hennessy, Provost, Stanford University; Professor Bastiaan J. Braams (Ex-Officio, NERSC User Group Chair), Courant Institute of Mathematical Sciences, New York University; Dr. Paul C. Messina, Defense Programs, U.S. Department of Energy; Dr. Stephen Jardin, (Ex-Officio, Chair of Program Advisory Committee), Princeton Plasma Physics Laboratory; Dr. Albert Narath (Retired), Lockheed Martin Corporation.

NERSC Computer Scientist Wins “Best Paper” Award at SC99

Leonid “Lenny” Oliker, a post-doc in NERSC’s Scientific Computing Group, was awarded “Best Paper of SC99” for a paper he co-authored with Rupak Biswas of NASA Ames Laboratory. Their paper, “Parallelization of a Dynamic Unstructured Application Using Three Leading Paradigms,” implemented a mesh adaptation code on the Cray T3E, SGI Origin2000 and the Tera MTA computers comparing programmability and performance. They compared several critical factors of parallel code development, including runtime, scalability, programmability and memory overhead. Their overall results demonstrate that multithreaded systems offer tremendous potential for quickly and efficiently solving some of the most challenging real-life problems on parallel computers.

A PDF version of paper can be found at <http://www.nersc.gov/~oliker/papers/sc99.pdf>

Effective System Performance (ESP): A New Benchmark

System effectiveness has been largely ignored to date in the high performance computing community, where peak performance is the most quoted statistic; yet this factor can make a significant difference in the total usefulness of the system. NERSC has proposed a new benchmark test that measures Effective System Performance (ESP) in a real-world operational environment. We hope that this test will be of use to system managers and will help to spur the community (both researchers and vendors) to improve system efficiency.

The importance of system effectiveness can be seen in the improved utilization of NERSC’s Cray T3E. Improving T3E utilization from 80% to 90% would be equivalent to adding more than $2 million in additional hardware for the following reasons:

- 644 processing elements (PEs) running at 90 percent are equivalent to 725 PEs running at 80 percent
- 81 PEs are needed to make up the difference
- a PE costs ~$50,000 list, $25,000 discounted
- 81 × $25,000 = $2,025,000.

In fact, over the course of 18 months from October 1997 to July 1999, NERSC increased T3E utilization from around 55 percent to over 90 percent — a value of $10.25 million (Fig. 1). This is almost equivalent to the improvement in processor/price performance from Moore’s Law.

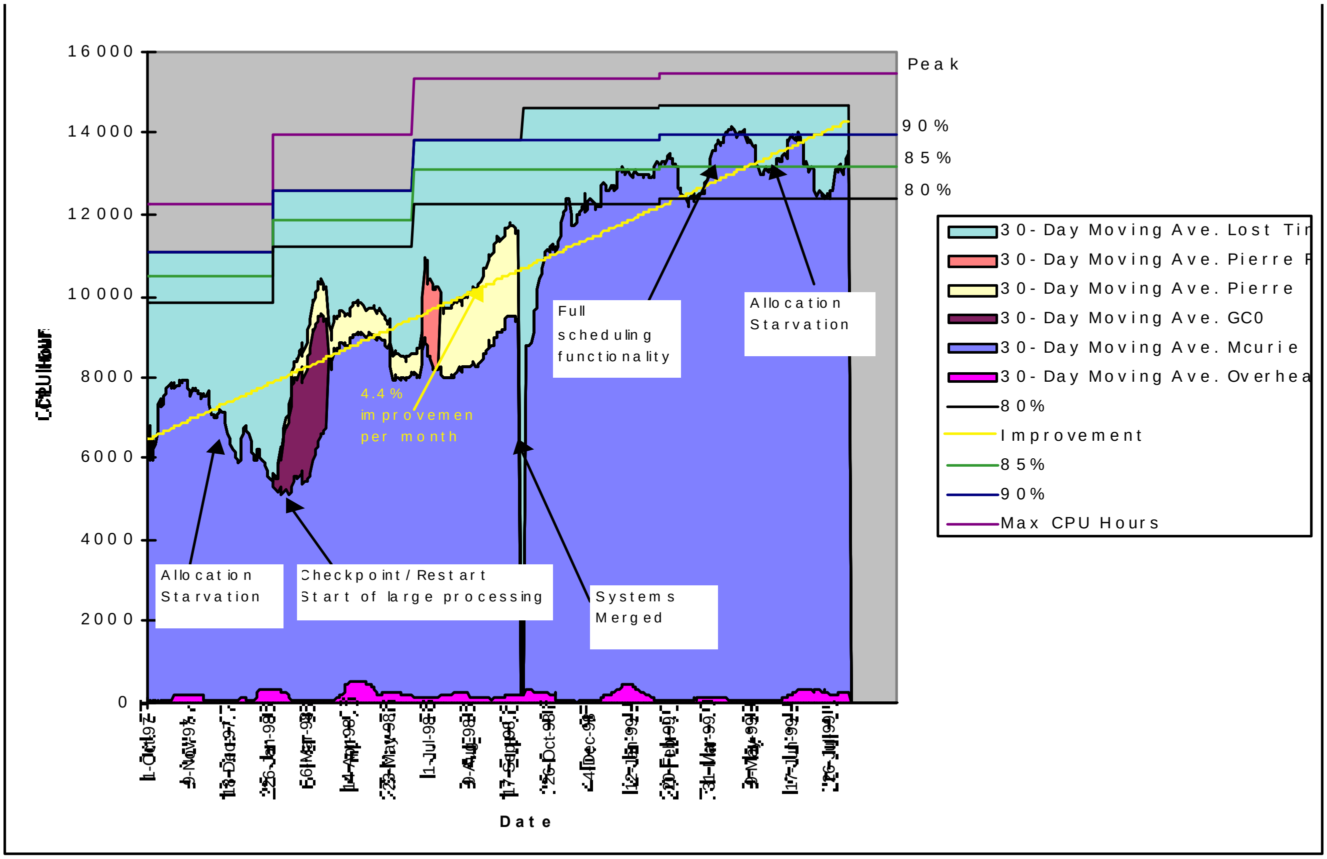


Fig. 1: Over 18 months, NERSC increased Cray T3E utilization from ~55% to ~90% — a value of $10.25 million.

Peak operations per second, the statistic used most widely to compare high performance systems, says nothing about how specific scientific codes will perform. The percentage of peak performance achieved on NERSC's Cray T3E varies widely depending on the benchmark (Table 1 and Fig. 2).

Table 1
Benchmark Results for NERSC's 644-PE Cray T3E

Benchmark	System Performance	Single Processor Performance	% of Peak
Theoretical peak	580.0 Gflop/s	900 Mflop/s	100.0%
Linpack	444.2 Gflop/s	~690 Mflop/s	76.6%
LSMS code (Locally Self-consistent Multiple Scattering), 1998 Gordon Bell Prize-winning application	256.0 Gflop/s	~398 Mflop/s	44.1%
Average of seven major NERSC applications	67.0 Gflop/s	~104 Mflop/s	11.6%
NAS Parallel Benchmarks	29.6 Gflop/s	~46 Mflop/s	5.1%

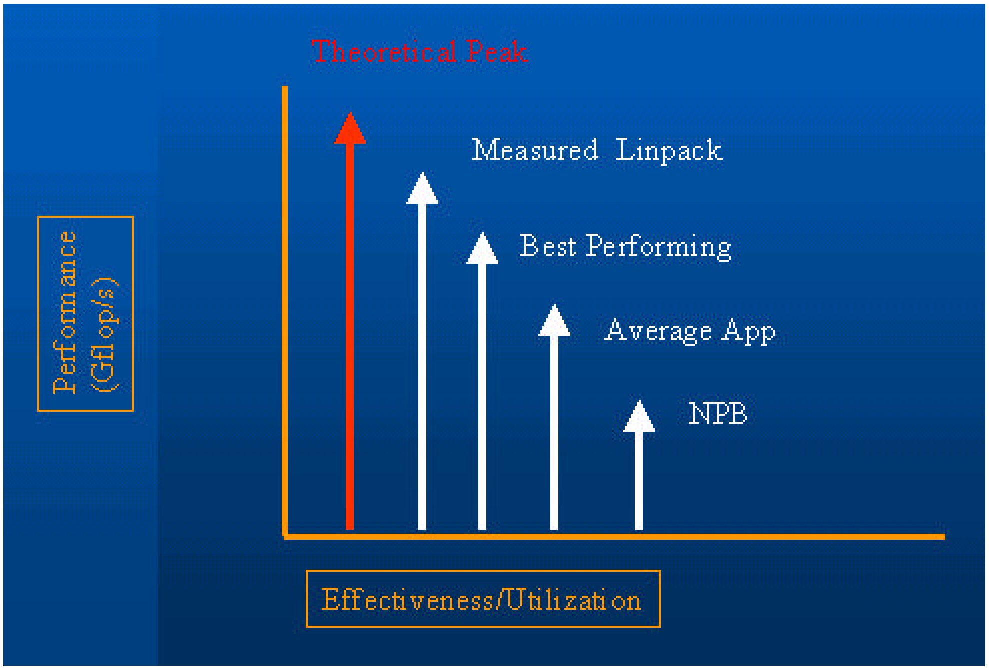


Fig. 2: Percent of peak performance achieved varies widely on NERSC's Cray T3E, depending on the benchmark.

Effective System Performance (ESP) is a new metric designed to evaluate systems for overall effectiveness, independent of processor performance. The ESP test suite simulates "a day in the life of an MPP" by measuring total system utilization. Results take into account both hardware (PE, memory, disk) and system software performance. Developed by NERSC as part of the NERSC-3 procurement process, ESP is designed to predict the effectiveness of a system before purchase, as well as to evaluate system changes before implementation.

The goals of the ESP benchmark include:

- Determining how well an existing system supports a particular scientific workload
- Assessing systems for that workload before purchase
- Providing quantitative effectiveness information regarding system enhancements
- Comparing different systems on a single workload or discipline
- Comparing system-level performance on workloads derived from different disciplines
- Comparing different systems for different workloads.

The NERSC ESP scientific application test suite is a set of 82 individual jobs, two of which are full-configuration jobs, namely, calculations that use the entire T3E system. The test can easily be modified to make it more appropriate for other systems.

Advancing Computational Science

Solved at Last: A Fundamental Problem of Quantum Physics

For over half a century, theorists have tried and failed to provide a complete solution to scattering in a quantum system of three charged particles, one of the most fundamental phenomena in atomic physics. Such interactions are everywhere; ionization by electron impact, for example, is responsible for the glow of fluorescent lights and for the ion beams that engrave silicon chips. Using the Cray T3E-900 at NERSC, collaborators at Berkeley Lab, Lawrence Livermore National Laboratory and UC Davis obtained a complete solution of the ionization of a hydrogen atom by collision with an electron, the simplest nontrivial example of the problem's last unsolved component. Their findings were presented in the cover story in the Dec. 24, 1999, issue of Science magazine.

Their breakthrough employs a mathematical transformation of the Schrödinger wave equation that makes it possible to treat the outgoing particles not as if their wave functions extend to infinity — as they must be treated conventionally — but instead as if they simply vanish at large distances from the nucleus.

"Using this transformation we compute accurate solutions of the quantum-mechanical wave function of the outgoing particles, and from these solutions we extract all the dynamical information of the interaction," says Bill McCurdy, Berkeley Lab's Associate Laboratory Director for Computing Sciences and a principal author of the Science article. The article was co-authored by Thomas Rescigno, a staff physicist at Livermore Lab, doctoral candidate Mark Baertschy of UC Davis and postdoctoral fellow William Isaacs of Berkeley Lab.

NERSC Sponsors Computational Science Lecture Series in Washington, DC

In an effort to raise the awareness of the role of computational science in federally funded research, NERSC presented a series of colloquia in Washington, D.C., showcasing some of the more exciting and promising areas of computational science. The eight colloquia were presented in April, May and June 1999 at Berkeley Lab's Washington, D.C., office.

The titles and speakers were "Re-inventing the Supercomputer Center at NERSC — Again," NERSC Division Director Horst Simon; "Supercomputing and the Fate of the Universe," Saul Perlmutter, Supernova Cosmology Project Leader (Co-recipient of Science Magazine's 1998 Breakthrough of the Year); "Computational Challenges in Structural and Functional Genomics," Teresa Head-Gordon, Berkeley Lab Staff Scientist; "Computational Fluid Dynamics and Combustion Modeling," Phil Colella, NERSC Applied Numerical Algorithms Group Leader (Recipient of IEEE's 1998 Sidney Fernbach Award); "Data Intensive Computing at NERSC," Robert Lucas, NERSC High Performance Computing Research Department Head; "NERSC Terascale Production Computing in the Next Decade," Bill Kramer, NERSC High Performance Computing Department Head; "Handling Large Datasets in Biology," Manfred Zorn/Sylvia Spengler, Co-directors of NERSC's Center for Bioinformatics and Computational Genomics; and "Materials Science at the Teraflop Level," Andrew Canning, NERSC Scientific Computing Group (Co-recipient of 1998 Gordon Bell Prize).

Center For Bioinformatics and Computational Genomics Hosts International Conference

NERSC's Center for Bioinformatics and Computational Genomics hosted the "Objects in Bioinformatics '99" conference in San Jose on August 19–21, 1999. The conference followed two very successful meetings held at the European Bioinformatics Institute (an Outstation of the European Molecular Biology Laboratory) in Hinxton, U.K.

Objects in Bioinformatics focused on the role of object-oriented technology, reusable software components, design patterns and distributed computing in bioinformatics and computational genomics. The conference was aimed at those who are interested in, are developing, or have developed object-oriented software that will be of use to the bioinformatics and genomics community in both academia and industry.

Conclusion

In summary, NERSC aims to be a leader both 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 and together they have redefined the notion of services to our user community. By developing new tools, refining existing ones and working actively with specialized groups within the high performance computing community, NERSC staff members are sharing their expertise with an ever-broadening community. For example, the NetLogger tools for diagnosing bottlenecks in distributed systems is now helping to improve performance at more than 20 research sites across the country. Here are some other examples of how we're sharing our technological expertise with our users and the high-performance computing community:

Invited Talks:

"Experiences with TCP/IP over an ATM OC12 WAN," Brian Tierney, Gigabit Workshop at IEEE InfoCom Conference, New York, NY, March 1999.

"Computational Aspects of Modular Ocean Model Development," Yun "Helen" He, invited talk at the Jet Propulsion Laboratory, Pasadena, CA, April 1999.

"Cosmic Microwave Background Data Analysis," Julian Borrill, Inner Space Outer Space II conference, Chicago, IL, May 1999.

"Building the Teraflops/Petabytes Production Supercomputing Center," Horst D. Simon, William T. C. Kramer and Robert F. Lucas, EuroPar '99, Toulouse, France, September 1999.

"MADCAP: The Microwave Anisotropy Dataset Computational Analysis Package," Julian Borrill, 5th European SGI/Cray MPP Workshop, Bologna, Italy, September 1999.

"COMBAT - Cosmic Microwave Background Analysis Tools," Julian Borrill, NASA Advanced Information Systems Research Projects Meeting, Boulder, CO, September 1999.

"Results of the BOOMERanG North America Test Flight," Julian Borrill, Key Tests For Cosmogenic Theories, Cambridge, England, December 1999.

Conference Presentations and Proceedings:

"Symbolic aspects of sparse orthogonal factorization," Esmond G. Ng, Conference on Linear Algebra: Theory, Applications, and Computations, Wake Forest University, Winston-Salem, NC, January 1999.

"A Scalable Sparse Direct Solver Using Static Pivoting," Xiaoye "Sherry" Li and J. W. Demmel, Proceedings of the 9th SIAM Conference on Parallel Processing and Scientific Computing, San Antonio, TX, March 1999.

"Incomplete Cholesky parallel preconditioners with selective inversion," Esmond G. Ng, the 6th International Symposium on Solving Irregularly Structured Problems in Parallel, San Juan, Puerto Rico, April 1999.

"Data Intensive Distributed Computing: A Medical Application Example," Jason Lee, Brian Tierney and William Johnston, Proceedings of the 7th International Conference on High Performance Computing and Networking, Europe, Amsterdam, the Netherlands, April 1999.

“Comparison of Parallel Schwarz Methods and Substructuring Methods for a Metal Forming Process,” Andreas Mueller, P. Adamidis, High Performance Computing, San Diego, CA, April 1999.

“Structural results for sparse orthogonal factorization,” Esmond G. Ng, Minisymposium on Recent Advances in Sparse Matrix Techniques, 1999 SIAM Annual Meeting, Atlanta, GA, May 1999.

“System Utilization Benchmark on the Cray T3E and IBM SP,” Adrian T. Wong, Leonid Oliker, William T. C. Kramer, Teresa L. Kaltz and David H. Bailey, Fifth Workshop on Job Scheduling, May 1999.

“Re-Revisit ICCG,” Esmond G. Ng, International Conference on Preconditioning, Minneapolis, MN, June 1999.

“An Embedded Boundary / Volume of Fluid Method for Free Surface Flows in Irregular Geometries,” Phil Colella, Dan T. Graves, David Modiano, E.G. Puckett, M. Sussman, ASME Paper FEDSM99-7108, in Proceedings of the 3rd ASME/JSME Joint Fluids Engineering Conference, San Francisco, CA, July 1999.

“Analysis of Substructuring in a Metal Forming Process,” Andreas Mueller, P. Adamidis, Proceedings of the 11th International Conference on Domain Decomposition Methods, Greenwich, July 1998.

“High performance sparse Cholesky factorization,” Esmond G. Ng, Sun Microsystems, Menlo Park, CA, July 1999.

“A Network-Aware Distributed Storage Cache for Data Intensive Environments,” Brian Tierney, Jason Lee, Brian Crowley, M. Holding, J. Hylton and F. Drake, IEEE International Symposium on High Performance Distributed Computing, Redondo Beach, CA, August 1999.

“NERSC,” Esmond G. Ng, Partners in Scientific Collaborations, Oak Ridge, TN, August 1999.

“High-Performance Data Intensive Distributed Computing,” Brian Tierney, Seminar at CERN on DPSS, NetLogger, and Data Grids, Geneva, Switzerland, and at the Large Hadron Collider Workshop, Marseilles, France, September 1999.

“Update on NERSC PScheD Experiences,” Tina Butler and Michael Welcome, Cray User Group T3E Workshop, Princeton, NJ, October 1999.

“Bad I/O and Approaches to Fixing It,” Jonathan Carter, Cray User Group T3E Workshop, Princeton, NJ, October 1999.

“I/O and Filesystem Balance,” Tina Butler, Cray User Group T3E Workshop, Princeton, NJ, October 1999.

“Ordering sparse matrices,” Esmond G. Ng, Stanford University, Palo Alto, CA, November 1999.

“Efficient Parallelization of a Dynamic Unstructured Application on the Tera MTA,” Leonid Oliker, with Rupak Biswas, SC99, Portland, OR, November 1999.

“Data Organization and I/O in a Parallel Ocean Circulation Mode,” Chris H.Q. Ding and Yun He, SC99, Portland, OR, November 1999.

“DeepView: A Channel for Distributed Microscopy,” Bahram Parvin, John Taylor and Ge Cong, SC99, Portland, OR, November 1999.

"A Parallel Implementation of the TOUGH2 Software Package for Large Scale Multiphase Fluid and Heat Flow Simulations," Erik Elmroth and Chris Ding, with Yu-Shu Wu, SC99, Portland, OR, November 1999.

"Parallelization of Radiance for Real Time Interactive Lighting Visualization Walkthroughs," David Robertson, Kevin Campbell, Stephen Lau and Terry Ligocki, SC99, Portland, OR, November 1999.

"Numerical simulation of incompressible viscous flow in deforming domains," Phil Colella and D. Trebotich, Proceedings of the National Academy of Sciences of the United States of America 96, 1999.

"A blocked incomplete Cholesky preconditioner for hierarchical-memory computers," Esmond G. Ng (with Barry Peyton and Padma Raghavan), proceedings of Iterative Methods in Scientific Computations IV, ed. David R. Kincaid and Anne C. Elster, IMACS, 1999.

"Coupling and Parallelization of Grid-based Numerical Simulation Software," Andreas Mueller, P. Adamidis, A. Beck, U. Becker-Lemgau, Y. Ding, M. Franzke, H. Holthoff, M. Laux, M. Muench, B. Steckel and R. Tilch, Proceedings of High Performance Computing in Science and Engineering - The Second Result and Review Workshop of the HPC Center Stuttgart (HLRS), 1999.

"The Challenge of Data Analysis for Future CMB Observations," Julian Borrill, Proceedings of the Rome 3K Cosmology Euroconference, 1999.

"MAXIMA: An Experiment to Measure Anisotropy in the Cosmic Microwave Background," Adrian T. Lee, Julian Borrill, et al., Proceedings of the Rome 3K Cosmology Euroconference, 1999.

"MADCAP: The Microwave Anisotropy Dataset Computational Analysis Package," Julian Borrill, Proceedings of the 5th European SGI/Cray MPP Workshop, 1999.

Workshops and Tutorials:

"NetLogger visualization and analysis tools," Dan Gunter, San Francisco State University, San Francisco, CA, April 1999

"NetLogger Tutorial," Brian Tierney, DOE Next Generation Internet workshop, Berkeley, CA, July 1999

"NetLogger Tutorial," Brian Tierney, HPDC Conference, Redondo Beach, CA, August 1999.

"NetLogger Tutorial," Brian Tierney, Stanford Linear Accelerator Center, Palo Alto, CA, November 1999.

"Production Linux Clusters: Architecture and System Software for Manageability and Multi-User Access," William Saphir and Patrick Bozeman with R. Evard and P. Beckman, SC99 Tutorial, Portland, OR, November 1999.

"High Performance Computing Facilities for the Next Millennium," William Kramer, James Craw, Keith Fitzgerald, Francesca Verdier, Tammy Welcome and Robert Lucas, SC99 Tutorial, Portland, OR, November 1999.

"Computational Biology and High Performance Computing," Horst Simon, Sylvia Spengler, Manfred Zorn, Teresa Head-Gordon, Adam Arkin and B. Shoichet, SC99 Tutorial, Portland, OR, November 1999.

Publications:

“Performance of greedy ordering heuristics for sparse Cholesky factorization,” Esmond G. Ng (with Padma Raghavan), SIAM J. Matrix Anal. Appl. 20, 1999.

“Parametric study of reactive melt infiltration,” Phil Colella and E.S. Nelson, in R. M. Sullivan, N. J. Salamon, M. Keyhani, and S. White, eds., “Application of porous media methods for engineered materials” AMD-Vol 233, American Society of Mechanical Engineers, 1999. Presented at the 1999 ASME International Mechanical Engineering Congress and Exposition, Nashville, TN, November 1999.

“A conservative finite difference method for the numerical solution of plasma fluid equations,” Phil Colella, M. R. Dorr, and D. D. Wake, Journal of Computational Physics, 149, 1999.

“A projection method for low speed flows,” Phil Colella and K. Pao, Journal of Computational Physics, 149, 1999.

“A conservative adaptive-mesh algorithm for unsteady, combined-mode heat transfer using the discrete ordinates method,” L. H. Howell, R. B. Pember, Phil Colella, W. A. Fiveland, J. P. Jessee, Numerical Heat Transfer, Part B: Fundamentals, 35, 1999.

“Thick-restart Lanczos Method for Electronic Structure Calculations,” Kesheng “John” Wu, Andrew Canning, Horst D. Simon and Lin-Wang Wang, Journal of Computational Physics, 1999.

“Parallel Empirical Pseudopotential Electronic Structure Calculations for Million Atom Systems,” Andrew Canning, Lin-Wang Wang, A. Williamson and A. Zunger, Journal of Computational Physics, 1999.

“Direct Numerical Simulation of the Developing Region of Turbulent Planar Jets,” Scott A. Stanley and S. Sarkar, AIAA Paper 99-0288, January 1999.

“An Adaptive Level Set Approach for Incompressible Two-Phase Flows,” Mark M. Sussman, Ann S. Almgren, John B. Bell, Phillip Colella, Louis H. Howell, Michael Welcome, Journal of Computational Physics, 148, 1999.

“Adaptive Mesh and Algorithm Refinement,” John B. Bell, William Y. Crutchfield, A. L. Garcia, B. J. Alder, Journal of Computational Physics, 1999.

“Multiple scales analysis of atmospheric motions: Impact on modeling and computation,” Ann Almgren (with N. Botta and R. Klein), Proceedings of the ENUMATH99 Conference, July 1999 (refereed).

“Dry Atmosphere Asymptotics,” Ann Almgren (with N. Botta and R. Klein), PIK (Potsdam Institute for Climate Impact Research) Report, September 1999.

“Large Eddy Simulation of a Plane Jet,” Scott A. Stanley (with C. Le Ribault and S. Sarkar), Physics of Fluids, Vol. 11, No. 10, October 1999.

“Influence of Nozzle Conditions and Discrete Forcing on Turbulent Planar Jets,” Scott A. Stanley and S. Sarkar, accepted for publication in AIAA Journal, December 1999.

“A Supernodal Approach to Sparse Partial Pivoting,” Xiaoye “Sherry” Li (with James Demmel, S. Eisenstat, J. Gilbert and J. Liu), SIAM J. Matrix Anal. Appl., vol 20(3), 1999.

“An Asynchronous Parallel Supernodal Algorithm for Sparse Gaussian Elimination,” Xiaoye “Sherry” Li (with James Demmel and J. Gilbert), SIAM J. Matrix Anal. Appl., vol. 20(4), 1999.

SuperLU Users’ Guide, James W. Demmel, John R. Gilbert, and Xiaoye S. Li, September 1999.

TRLAN (Thick-Restart Lanczos Method) User Guide, Kesheng “John” Wu and Horst Simon, March 1999.

“Portable Parallel Programming for the Dynamic Load Balancing of Unstructured Grid Applications,” Leonid Oliker (with R. Biswas, S. K. Das, and D. J. Harvey), 13th International Parallel Processing Symposium, 1999.

“Experiments with Repartitioning and Load Balancing Adaptive Meshes,” Leonid Oliker (with R. Biswas), Grid Generation and Adaptive Algorithms, IMA Volumes in Mathematics and its Applications, Vol. 113, Springer-Verlag, 1999.

“Global Earth Structure: Inference and Assessment,” Osni Marques with D. W. Vasco and L. R. Johnson, Geophysical Journal International, 1999.

“Evaluating System Effectiveness in High Performance Computing Systems,” Adrian T. Wong, Leonid Oliker, William T. C. Kramer, Teresa L. Kaltz and David H. Bailey, November 1999.

Conference Organization

Along with the Mathematical Sciences Research Institute in Berkeley, NERSC co-hosted a “Conference on Self-Assembling Geometric Structures in Material Science: The Geometry of Interfaces in Mesoscopic Materials” in April 1999. The conference brought together experimentalists, theorists and mathematicians to discuss newly discovered microstructured material systems, computational modeling of materials and the relevant advances in the understanding and classification of embedded periodic surfaces that satisfy variational and symmetry constraints. The study of geometric structure and the desire to predict and control properties of self-assembling materials are themes common to other parts of material science. The meeting was scheduled to take place after the Spring meeting in San Francisco of the Materials Research Society.

At SC99 in Portland, Ore., NERSC staff contributed their talents to making the conference one of the best-attended in history, from chairing technical committees to assisting with communications.

NERSC and Berkeley Lab hosted the 1999 DOE Computer Graphics Forum in May at Lake Tahoe, CA. The 24th annual DOE Computer Graphics Forum continued the tradition of its predecessors by providing a forum for DOE Computer Graphics researchers to get together, exchange ideas and share information in a casual atmosphere.

Thomas DeBoni of NERSC User Services provides audio-visual, computer and network support for DOE’s annual meeting on high performance computing held each spring in Salishan, Ore.

El Cerrito High School Students Learn About NERSC

A group of 17 math and science students from El Cerrito High School visited the Lab in May 1999 to hear about career options in high-performance computing and learn about various programs in Computing Sciences. The students listened to presentations and asked good questions about topics ranging from salaries to programming languages.

Computing Sciences staff members who met with the group were Tom DeBoni, NERSC User Services Group; Sherry Li, NERSC Scientific Computing Group; Denise Wolf, Center for Bioinformatics and Computational Genomics; Ann Almgren, Center for Computational Sciences and Engineering; and Ravi Malladi, Mathematics.

STAFF EFFECTIVENESS

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

An ongoing measure of the effectiveness of NERSC's staff is the fact that NERSC is able to offer more computing power with a smaller staff than other computing centers with comparable budgets. Although our staff members certainly do work hard, we think working smarter is what makes the crucial difference. The interdisciplinary expertise of so many staff members makes a major contribution to our effectiveness — their in-depth knowledge of a scientific field as well as computational methods and programming techniques allows them to understand the needs of researchers and make the most effective use of computers to solve scientific problems.

One of our primary objectives is to increase the efficiency and effectiveness of scientific computation, not only for NERSC clients, but within the entire computational science community. Here are some examples of how we're achieving that goal.

NERSC Staff Ports Global Climate Model to Cray SV1

The Climate System Model (CSM), a fully coupled, global climate model that provides computer simulations of the Earth's past, present, and future climate states, has been ported to run on NERSC's newly installed Cray SV1 computers, thanks to the efforts of Harsh Anand, Chris Ding and Helen Yun He. The effort will allow NERSC users to use the National Center for Atmospheric Research-developed model to simulate atmosphere, ocean, sea ice, and land surface conditions, either coupled together or as individual components. The model is already being used by a research group led by Prof. Inez Fung, director of the Center for Atmospheric Sciences at UC Berkeley.

"It was a humongous job — we were the first ones to port CSM to run on the Cray SV1 and J90," said Harsh, who works in the User Services Group. She credits Dr. Lawrence Buja of NCAR with making the job possible. She also drew upon the contacts she developed over the 12 years she worked at NCAR in Colorado.

The model was developed over the years by several groups at NCAR, and Dr. Buja had to consult with members of those groups over the course of the project. Additionally, each of the model's five components — land, sea ice, ocean, atmosphere and coupler — were written using different methodologies, and all were hard-wired to run in NCAR's computing environment.

Chris Ding of NERSC's Scientific Computing Group, who is the main coordinator for climate research projects and provided his expertise to the project, said, "The CSM is the biggest thing happening in the climate modeling community. Scientifically, it's the best coupled model in the world." Because of the size and complexity of the model, running it requires a lot of time and effort and expertise. "If you use CSM, you're running a monster," Chris said. "For this reason, we also provided CPU and memory usage information."

To ensure that the CSM would run smoothly on NERSC's SV1 and J90 machines, Helen He of the Scientific Computing Group served as a tester. Helen, who earned her doctorate in marine science from the University of Delaware (and earned the "Best Dissertation Award") drew upon her background in oceanography. As a general user, she helped discover necessary modifications in the setup and model scripts which are needed for a specific user environment.

"Although it was a big effort, it was worth it," Harsh said. "By being the first to do this, we've provided a good place to start for other centers wishing to port the CSM."

For more information on CSM, go to http://hpcf.nersc.gov/software/apps/climate/.

New Database Seeks Out Products of Alternative Gene Splicing

In its first half year of operation, a new database that identifies clusters of proteins arising from alternative gene splicing has received more than 35,000 requests from researchers in genetics and cell and developmental biology around the world. The Alternative Splicing Data Base, or ASDB, was created by Inna Dubchak, Igor Dralyuk, and Manfred Zorn of NERSC’s Center for Bioinformatics and Computational Genomics, in collaboration with M. S. Gelfand of the Institute of Protein Research at the Russian Academy of Sciences.

“In the first weeks after ASDB went on line in January, requests for data went from an average of a few dozen per day to hundreds,” says Dubchak. “One day in May, we got more than 6,000 requests.”

The worldwide demand for alternative gene-splicing data is confirmation that Dubchak and her colleagues have hit upon one of the most exciting and important problems in contemporary biology — which suits her fine: “We want to help biologists solve their hardest problems by computational methods.”

Dubchak and her colleagues spent a year and a half assembling the ASDB, which currently contains some 1,700 protein sequences. It can be searched to find out how many known proteins can be derived from a single gene sequence (some can generate up to 64 variations of messenger RNA!) or to find all known products of alternative splicing in a given organism, such as the fruit fly, mouse, or human, or in a particular tissue such as muscle, heart, or brain.

The ASDB is available on the Web at http://cbcg.nersc.gov/asdb.

Vortex-Flame Simulation Is First to Reproduce Experimental Results

As computing tools become more powerful, computational simulation will play an increasingly important role in the design of combustion devices such as more efficient gasoline engines or less-polluting diesel engines. Researchers in NERSC’s Center for Computational Sciences and Engineering (CCSE) are working toward a key component of this goal with the development of high-fidelity numerical simulation capabilities applied to turbulent combustion processes such as furnaces and engines.

One of the most difficult issues in the modeling of turbulent combustion is the coupling between chemical kinetic processes and the small-scale eddies in the flow. The computational challenge arises from the need to resolve numerically a wide range of spatial and temporal scales associated with the flow field, while at the same time employing complex models for the fundamental chemical processes. CCSE is developing an adaptive block-structured refinement approach which allows overall computational effort to be focused in localized, time-evolving regions of the domain, such as the zone near a burning flame. By minimizing unnecessary computation in less critical regions of the domain, they can incorporate more detail in the fluid and chemistry components of the model. Implementing such software requires a variety of design and implementation expertise, including software infrastructure design, detailed algorithm development, physical model validation, parallel computing, and complex visualization issues.

CCSE recently tested their methodology on a set of vortex-flame interactions, an important prototype for premixed turbulent combustion. They studied the effect of fuel stoichiometry on the interaction of a counter-rotating vortex pair with an initially flat premixed methane flame. The simulation was based on a well-diagnosed, highly reproducible, two-dimensional vortex-flame experiment by Q.-V. Nguyen and P. H. Paul at Sandia National Laboratories. This experiment posed a challenge to existing numerical combustion models, which could not correctly predict the time-dependent behavior of a number of intermediate species produced by the combustion process. CCSE's simulation was the first to reproduce some key results of the experiment.

CCSE conducted numerical simulations using a configuration similar to the Sandia vortex-flame experiment, in terms of fueling characteristics and the strength and shape of the imposed vortices. Simulations over a range of inlet stoichiometry and vortex characteristics indicated that the vortex not only stretches and strains the flame, but also scours material from the cold region in front of the flame. The scouring effect is strongly dependent on the spatial distribution of various key flame radicals, and therefore is strongly affected by the inlet fuel equivalence ratio. This latter observation helped to explain previously observed computational results which seemed to otherwise disagree with experiment, and underscores the benefit of efficient computing methods that can provide results over a range of similar scenarios. CCSE is continuing research to further improve the fidelity of the detailed fluid dynamical simulations, and is working with combustion chemists at UC Berkeley and LBNL's Environmental Energy Technologies Division to develop more complete chemical mechanisms for combustion.

Parallel Version of Flow and Transport Code Runs 60 Times Faster

For challenges ranging from cleaning up groundwater contamination to increasing the flow from oil and natural gas fields, understanding the movement of liquids and gases in the subsurface is essential for earth scientists, and computer simulations give them insight into otherwise inaccessible regions. The Earth Sciences Division at Berkeley Lab has developed a code for simulating multiphase flow and transport processes in fractured-porous media. Called TOUGH2 (Transport of Unsaturated Groundwater and Heat), the code can model one-, two-, and three-dimensional flows of multiple phases, such as gas, aqueous liquids, and oil, and multiple components, such as water, air, organics, and radionuclides. The code is used by over 150 organizations in more than 20 countries for large-scale, multi-component flow simulations in environmental remediation, nuclear waste isolation, and geothermal reservoir engineering.

NERSC's Scientific Computing Group has developed a parallel implementation of TOUGH2 that enables it to run on high performance systems. This will benefit researchers such as the Yucca Mountain nuclear waste isolation project. Currently, the Yucca Mountain modeling group runs their flow model on about a dozen workstations 24 hours a day, 7 days a week. But they need to study grid blocks of 100,000 to 1 million, which is impossible on even the fastest workstations. NERSC's MPP systems will allow the model resolution to be increased significantly, and will provide a complete flow picture in a timely fashion.

TOUGH2 uses a finite-volume method to solve the mass-energy balance equation. The most computationally demanding part is to solve a large, unsymmetric, non-positive, linear equation. NERSC staff are developing a parallel implementation of the package and integrating two key software components, the domain partitioner and the linear solver. To optimize the code, they will study both parallel computing related issues such as efficiency, scalability, etc., and numerical issues such as the stiffness of the Jacobian matrix involved in solving the highly non-linear equations. Typically these equations are very stiff and difficult to solve. The effectiveness of the preconditioner and iterative methods when applied to such large-scale problems will be investigated.

Results to date look promising. The codes have been restructured, domain decomposition is completed, and the Aztec solver from the ACTS Toolkit has been integrated into the package. On a real application of 17,584 grid blocks with 3 components (52,752 equations), the parallel codes solved the problem 60 times faster on the T3E than the original sequential codes did on workstations.

Electronic Structure of Million-Atom Systems Can Now Be Calculated

The electronic, optical, transport, and structural properties of semiconductor nanostructures (films, quantum dots, and quantum wires) have recently been under intense study. This interest arises because of the novel physical properties of these systems and their potential application to a whole new set of nanoscale devices such as lasers, sensors, and photovoltaics.

Before scientists and engineers can begin to design nanoscale devices with custom-made electronic and optical properties, they must have a detailed understanding of the underlying physical phenomena. In nanoscale systems whose sizes vary from 1 to 50 nanometers, these phenomena are controlled by quantum mechanical effects and can only be understood by solving Schrödinger's equation. Performing quantum mechanical calculations on systems containing thousands or millions of atoms requires state-of-the-art numerical techniques and computing resources.

Lin-Wang Wang and Andrew Canning of NERSC's Scientific Computing group, in collaboration with Alex Zunger's research group at the National Renewable Energy Laboratory, have developed a Parallel Empirical Pseudopotential method for electronic structure calculations. This code allows the calculation of the electronic structure (for a small number of electronic states) of systems of up to 1 million atoms on the T3E at NERSC. It uses pseudopotentials for the single-electron Hamiltonians, which are commonly used for accurate *ab initio* total energy calculations. It expands the wavefunctions in planewaves, thus requiring fast Fourier transforms to convert the wavefunction from reciprocal space to real space. The number of basis functions in such a million-atom system is about 50 million. A "folded spectrum" algorithm developed by Lin-Wang Wang is used to calculate a few physically interesting states in the middle of the energy spectrum without the calculation of all the other states.

Previous methods were not able to give accurate information on the electronic structure of systems larger than 1000 atoms. The Parallel Empirical Pseudopotential program opens a new approach in this field by enabling accurate atomistic calculations for million-atom nanosystems. This parallel code is now used by many materials science research groups and has resulted in publications in the areas of quantum dots, quantum wells, superlattices, alloys, composition modulations, ordering, and defect states.

Easier Access to Massive HENP Datasets

NERSC's Scientific Data Management Group (SDM) is involved in various projects including tertiary storage management for high energy and nuclear physics (HENP) applications, data management tools, and efficient access to mass storage data. One of their recent accomplishments is the Storage Access Coordination System (STACS), which was developed to support the Mock Data Challenge tests of the Grand Challenge Application on HENP Data.

STACS coordinates file caching from tape to a shared disk for a large number of concurrent HENP applications. The software supports simultaneous scheduling of multiple files, incorporates the NetLogger file tracking system developed at Berkeley Lab, and produces online dynamic resources usage profiles, such as disk cache in use, file transfers pending, etc. Despite its complexity, STACS is robust, with clean interfaces and efficient functionality. It performed so well in tests that the STAR and PHENIX projects at Brookhaven National Laboratory plan to use STACS in their data analysis framework, CERN is adopting the STACS index method, and several Next Generation Internet projects are considering using concepts developed in the STACS project.

The work of the SDM group is unique among supercomputing centers, and we are not aware of a comparable research effort elsewhere. NERSC's research efforts in data storage and management will result in efficient new tools that our clients can use to extract scientifically significant information from their petabyte datasets.

NERSC, CERFACS Jointly Studying Hybrid Ordering Algorithms

In January 1999, Sherry Li of NERSC's Scientific Computing Group began working with three computer scientists at CERFACS in France to study hybrid ordering algorithms for both direct and iterative solvers. A $10,000 grant will be used primarily to fund collaborative visits between NERSC and CERFACS, the Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (or European Center for Research and Advanced Training in Scientific Computing) in Toulouse. The grant is provided by the France-Berkeley Fund, which was established in 1993 by the French Embassy in the United States and UC Berkeley to support scientific and scholarly exchanges. The money is viewed as seed money for getting new collaborations started, rather than as funding for long-term research. In addition to Sherry, Iain Duff, Patrick Amestoy, and Jean-Yves L'Excellent of CERFACS will be working on the project.

In their proposal, the group wrote that they plan "to develop new ordering algorithms for large sparse linear systems of equations. The hybrid schemes are expected to reduce the number of interface variables, hence the amount of inter-processor communication, in a domain decomposition method. We will apply our new algorithms to the large-scale industrial and scientific applications from France and LBNL."

At the center of the collaboration are two of the world's top codes for solving large linear equations. One code, called SuperLU, was developed by Sherry Li and Jim Demmel. Duff and Amestoy created a code called MUMPS. Each code takes a quite different approach to the problem, and each has its own strengths. The original goal was to create a hybrid of the two, but it's not easy to blend the codes, Duff said. One focus is the ordering of the codes — using the same ordering in both codes produces quite different effects. Both codes are useful, Duff notes, as they "enable great things — they really are at the heart of scientific computation."

"The great thing is we have people working individually and bringing different competencies and experiences to these efforts," Duff said. "The goal is to combine them to benefit both organizations and science in general."

PROTECTED INFRASTRUCTURE

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

Cycber-security became a DOE-wide watchword in 1999 due to incidents involving information on classified systems. Although all work conducted using NERSC's systems is unclassified, it is essential that we strike a balance between providing access for users across the country while ensuring that our systems are secure and protected.

One of the key steps to enhancing security at NERSC was the installation in 1999 of "Bro," an intrusion detection system developed at Berkeley Lab. Written by Vern Paxson, Bro (named for the ever-watching "Big Brother" in George Orwell's "1984") monitors and actively responds to incoming data packets and searches for unusual patterns which could indicate an attack. If such a pattern is detected, a real-time notification is sent to cyber-security personnel, and, when appropriate, BRO automatically blocks future intrusion attempts.

BRO is a layered system that seeks out certain types of network traffic. The first layer is a general packet filter, which decides which data packets should be examined. The second layer is an "event engine," which takes the first-level packets and pieces them together into "events," such as the beginning or end of a connection; or, for some applications such as FTP (file transfer protocol), high-level events such as identifying user names. Above that is the policy layer, which interprets scripts, written in a specialized language, that define how to respond to different events. Should the policy layer detect information amounting to an attempted security breach, the system notifies computer security people in real time, or if need be, can automatically and proactively block the suspect activity until it can be analyzed by staff. BRO also archives summaries of the network traffic into and out of Berkeley Lab in a permanent record.

BRO's powerful tools provide a strong security blanket for both NERSC and Berkeley Lab. And not only does Bro detect intrusions, it also compiles extensive logs of the traffic, which have proven useful in tracking down hackers.

While NERSC continues to assess and meet cyber-security issues, hackers continue to look for vulnerabilities to exploit. In fact, one such hacker attempted to end 1999 with an attack on NERSC. Brent Draney of NERSC's Computational Systems Group detected and rebuffed one hacking try early in the evening of Dec. 31, but otherwise NERSC rolled smoothly into the Year 2000.

Although the calendar rolled over into Y2K with barely a ripple, NERSC was well prepared for the event. In the midst of the hysterical claims of computational chaos resulting from Y2K, NERSC developed a realistic strategy to ensure that our systems and users would not be adversely affected by the date change. This strategy was devised so that NERSC could continue to provide systems and services to our users.

NERSC conducted extensive tests of all of our systems well in advance of Y2K, and also shared our experience and results with other sites. NERSC was the first site to completely test and report the results for Y2K compliance by the Cray T3E supercomputer and UNICOS operating system. NERSC was also the only site to fully test HPSS (the High Performance Storage System). Finally, NERSC was the only site to fully test the IBM SP system for Y2K compliance.

The results of our tests were then used by others. For example, NERSC's HPSS tests were used by the entire HPSS collaboration to run tests at other sites running the system.

NERSC was also one of the first facilities to complete Y2K testing of the IBM SP supercomputer and our test procedures and results were shared with other sites with IBM machines.

CONCLUSION

In the words of Bill McCurdy, director of NERSC from 1991-1995 and now Associate Laboratory Director for Computing Sciences at Berkeley Lab, NERSC is a vision fulfilled. Here are Bill's observations on NERSC's 25th anniversary:

"In 1974 the staff of the Controlled Thermonuclear Research Computer Center, NERSC's original name, began a revolution in computing. Their new vision was to provide a supercomputer to a user community spread over the nation, instead of only to local users, and they set about creating and adapting technology to do so. Over the years, the staff of NERSC and ESnet have delivered on that original concept, and continued to set the standards for the supercomputer centers and research networks that followed.

"Twenty-five years ago, the word "supercomputer" was not yet in current use, and we were just scientists; "computational scientist" was not yet a respectable or even understandable term. Computer science was only beginning to gain recognition as an independent discipline. In those early days, scientists ordinarily sat in a computer center, keypunching decks of computer cards for batch submission, and then waiting for fan-folded stacks of printout. That was what it meant to use a computer center. Interacting with a large computer via a printing teletype or a "dumb" terminal was still a novelty for most of the country's scientists.

"The idea that interactive scientific computing could be provided to a national community from a central facility was truly revolutionary. The early NERSC Center achieved that goal and built the modern aesthetic of supercomputing, which allows scientists to interact with the machines as though they were in the same room, visualizing and manipulating results immediately. The aesthetic was so strong that for a while a delay was put in the connections for local users to give them the same level of access as users on the other side of the country. That arrangement ensured that the pressure for better service for the entire national community would be the dominant influence on the Center.

"The concept worked. Numerical simulation in plasma physics and fusion research advanced quickly and set a new standard for scientific discovery. These scientists were the vanguard of the supercomputer revolution. When NERSC broadened its mission to serve a larger scientific community in the late 1980s, new discoveries and advances over the entire spectrum of scientific research ensued.

"Another important result from the center was development of the Cray Time Sharing System. CTSS became the standard at other centers, and when the National Science Foundation Centers in Illinois and San Diego were established, they also adopted the NERSC (then called the MFECC) model of a supercomputer center, including CTSS.

"In its 25-year history NERSC has witnessed a stunning change in the technology of computing. In 1974 the speed of a supercomputer was measured in megaflops and its memory was large if it had sixty–four thousand words. Today our measures are teraflops for speed and terabytes for memory – million fold increases over the standards of 25 years ago. The history of computing technology is not NERSC's history though. NERSC's story is about invention and scientific discovery, and it is the story of the computer scientists and scientists of the center whose accomplishments created its influential past and are creating its future."

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 the same determination and commitment which we have demonstrated over the past 25 years. We plan to continue working to improve on our systems and services, and will report back to you on our further progress next year.