Minutes for the Advanced Scientific Computing Advisory Committee Meeting
March 13-14, 2003, Crowne Plaza Hotel, Washington, D.C.

ASCAC members present:

John W. D. Connolly, Vice Chair
Roscoe C. Giles
Helene E. Kulsrud
Gregory J. McRae
Juan C. Meza
Karen R. Sollins
Ellen B. Stechel
Stephen Wolff
Margaret H. Wright, Chair

ASCAC members absent:

Jill P. Dahlburg
William A. Lester, Jr.

Also participating:

Melea Baker, Office of Advanced Scientific Computing Research, USDOE
James Corones, President, Krell Institute
James Decker, Principal Assistant Director, Office of Science, USDOE
James Glimm, Chair, Department of Applied Mathematics and Statistics, Stony Brook University; Director, Center for Data-Intensive Computing, Brookhaven National Laboratory
Daniel Hitchcock, Senior Technical Advisor, Office of Advanced Scientific Computing Research, USDOE
Michael Holland, Office of Science and Technology Policy, Executive Office of the President
Gary Johnson; Program Manager; Advanced Computing Research Testbed; Mathematical, Information, and Computational Sciences Division; ASCR, USDOE
Kathleen Kingscott, Director of Public Affairs, IBM, Washington, D.C.
Alan Laub, Director, SciDAC, ASCR, USDOE
Reinhold Mann, Deputy Laboratory Director for Science and Technology, Pacific Northwest National Laboratory
Anthony Mezzacappa, Physics Division, Oak Ridge National Laboratory
Frederick O’Hara, ASCAC Recording Secretary
C. Edward Oliver, Associate Director, Office of Advanced Scientific Computing Research, USDOE; ASCAC Designated Federal Officer
Aristides Patrinos, Director, Office of Biological and Environmental Research, USDOE
Walter Polansky, Acting Director, Mathematical, Information, and Computational Sciences Division, Office of Science, USDOE
Jeannie Robinson, Oak Ridge Institute for Science and Education
Charles Romine, Program Manager, Applied Mathematics, MICS, ASCR, USDOE
Rachel Samuel, Office of Management, Budget, and Evaluation, USDOE
Stephen Scott, Research Scientist, Computer Science and Mathematics Division, Oak Ridge National Laboratory
Horst Simon, Director, National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory

Rick Stevens, Director, Mathematics and Computer Science Division, Argonne National Laboratory
Gloria Sulton, Office of the General Counsel, DOE
Linda Twenty, Program Analyst, Office of Advanced Scientific Computing Research, USDOE
Roy Whitney, Chief Information Officer, Thomas Jefferson National Accelerator Facility
Thomas Zacharia, Director, Center for Computational Sciences, Oak Ridge National Laboratory

About 35 others were in attendance in the course of the two-day meeting.

Thursday, March 13, 2003

Chairwoman **Margaret Wright** called the meeting to order at 8:29 a.m. and announced that Warren Washington had resigned from the Committee because of his new responsibilities as the chair of the National Science Board. The DOE FY04 budget request is now public and is being reviewed by Congress. Also, the FY05 budget request is open for discussion. She called the Committee's attention to a *New York Times* article about the chancellor of a school system who was embroiled in controversy when he chose schools to be exempt from a new curriculum. Veteran educators and political consultants attributed the ensuing imbroglio to "a stubborn insistence on doing what he thought was right regardless of the consequences." They said he acted on what he believed to be the merits of the case. "This is always a danger. ... Never let your candidates get in a situation where they have to choose winners because the only people who remember are the people who lose."

Wright commented that questions sometimes arise about what ASCAC members are allowed to do. For example, can they testify to Congress about the budget? Edward Oliver had invited **Rachel Samuel**, Deputy Advisory Committee Management Officer of DOE, and **Gloria Sulton**, Office of General Counsel, to review the agenda, roles, and responsibilities of advisory committees. Samuel pointed out that DOE has a Committee Management Officer to supervise advisory committees and that ASCAC has a designated federal officer (DFO). Advisory committee members are selected for their diverse knowledge and opinions and serve without compensation. The Federal Advisory Committee Act (FACA) was enacted in 1992 and amended in 1997. Title 41, CFR Part 102-3, Federal Advisory Committee Management (as amended in 2001) provides rules for advisory committees. Also, DOE M 510.1-1, Advisory Committee Management, lays out the activities of advisory committees. Advisory committees are to conduct business openly; they are solely advisory; and they are to advise on the development, implementation, and evaluation of policies and programs in a defined DOE subject area. Minutes are required and must be certified by the chair of the committee and deposited in the DOE Reading Room. DOE advisory committees can be established by presidential order, by Congress, and under general agency authority. DOE has 22 federal advisory committees: 7 statutory and 15 discretionary.

The concerns and sensitivities regarding advisory committees include conflict of interest, scope and objectives (which are listed in the committee's charter), membership, and balance/representation. A commitment to the needs and purposes of the committee is expected from the committee members along with frankness and a sincere effort to avoid conflict of interest and its appearance.

Samuel turned the floor over to Sulton, who provided the Committee with written guidelines on ethics and conflict of interest. Committee members who determine that they have a conflict of interest are asked to recuse themselves from any vote in writing and on the record. In addition to potential personal or financial gain, DOE worries about the use of membership for personal gain, using inside information improperly, and accepting gifts and gratuities. She provided copies of the guidelines for the acceptance of gifts by federal employees.

Giles asked the definition of “lobbying.” Sulton said that all people have First Amendment rights, so one can speak for oneself before Congress or any other body, but one may not speak for the Committee or make representations on behalf of this Committee.

Wright asked if a group from this Committee could visit a federal employee (e.g., in the Office of Management and Budget). Sulton voiced the opinion that that would probably be construed as lobbying. Wright asked if she was representing the university community when she spoke out at an ASCAC meeting. Sulton responded that representing the university community would not be a conflict of interest; one’s financial interest is tied to specific university. However, if only two universities would be able to benefit from a Committee action and a Committee member’s university was one of those two, the conflict-of-interest question becomes fuzzy. Disclosure of affiliation will often obviate the question of conflict of interest.

Stechel asked whether it would be lobbying if some members of the Committee wanted to understand where a federal employee stood and then to bring that understanding back to the Committee. Sulton suggested that the federal employee be asked to address the full Committee publicly. One undermines the Committee’s deliberations if one interviews another person and then synthesizes that person’s views. Sollins pointed out that all the members of the Committee cannot do everything; there has to be delegation; that is why there are subcommittees. Sulton suggested that those requests be filtered through the Committee’s DFO. With conference calls that include the DFO, one can bring discussions into the public arena.

McRae asked what role the Committee can play in setting the charter and agenda of the Committee. Sulton replied that the charter is set by the Secretary of Energy. The agenda is set by the Department. McRae asked what the mechanism is for changing the charter. Samuel responded that it is up to the Secretary and Department to change the charter and agenda of the Committee. They determine what they want advice about. Sulton added that it would be inappropriate for the Committee to discuss matters outside the scope of its work. An individual, on the other hand, can recommend to the Department that other issues be discussed by the Committee.

Wright asked how one resolves questions of bias that are raised. Sulton replied that the DFO resolves such questions with the General Counsel’s Office. Openness protects everyone on the Committee.

Wright introduced **Thomas Zacharia** to talk about the Cray X1 and Black Widow evaluation and plans. Juan Meza recused himself from the discussion.

Zacharia listed the high-level goals of the Center for Computational Sciences (CCS):

- Evaluate new hardware for science through the development and evaluation of emerging and unproven systems and experimental computers.
- Deliver leadership-class computing for DOE science by offering specialized services to the scientific community, focused on the biological, climate, and nanoscale sciences. It has been the principal resource for the Scientific Discovery Through Advanced Computing (SciDAC) program. For the future, it looks forward to improving performance on major scientific simulations by a factor of 50 by 2005 and a factor of 1000 by 2008.

- Educate and train the next generation of computational scientists.
- Operate as a designated user facility (since 1994).

The equipment it has used includes the Intel Paragon (no longer around), the IBM Power3 (a terascale system), the IBM Power4 (which operates at a peak of 4.5 Tflop/s), and the Cray X1 (which will soon be evaluated).

54% of CCS's resources are dedicated to SciDAC. An example of applications running at the Center is the all-orders spectral code in 3D for fusion simulation. A broader view of usage reveals that biology accounts for 11% of CCS usage, chemistry 20%, climate 17%, materials science 16%, fusion 13%, and physics 17%. The Center has made significant increases in expertise in biology; each of the four Genomes to Life laboratories has large computational requirements.

The CCS is available to a national user community. Four types of user agreements are available:

- a nonproprietary, no-cost agreement for commercial users,
- a nonproprietary, no-cost agreement for educational users,
- a nonproprietary, cost-required agreement for all users, and
- a proprietary, cost-required agreement for all users.

It is one of 20 user facilities managed by Oak Ridge National Laboratory (ORNL). More than 500 user agreements are in place. Cray is the most recent.

To understand the context of the Cray X1 proposal, one must realize that the Earth Simulator is about 13 times faster than the top Office of Science (SC) machine. The last time SC had the fastest machine was in 1995. The SC computing capability is significantly behind the leadership-class machines; it must keep the leaders in sight. In May 2002, ASCAC stated that "without a robust response to the Earth Simulator, the United States is open to losing its leadership in defining and advancing frontiers of computational science as a new approach to science. This area is critical to both our national security and economic vitality."

The CCS held a series of workshops and meetings with users and vendors [Cray; Hewlett-Packard (HP); International Business Machines (IBM); Silicon Graphics, Inc. (SGI); and others]. The assessment was that, even though clusters of general-purpose symmetric multiprocessors (SMPs) dominate U.S. high-performance computing (HPC), the science community and users found a need to increase processor speed, parallelism, and algorithm efficiency to meet the increased computational requirements for scientific simulation. As a result increases have been seen in relative memory and interconnect latencies, power consumption, heat generation, system complexity, and software complexity. A more balanced system is needed.

The Cray X1 provides a relatively well-balanced system for science applications as judged from quarternary plots of processor performance, memory, cache, and interconnect. Cray X1 is the commercial name for the SV2 project that Cray has been building for the National Security Agency (NSA) for more than 4 years. It combines multistreaming vector processors with a globally addressable memory similar to T3E. It offers the best opportunity for a leadership-class system for delivered performance in scientific applications, such as climate, materials, astrophysics, and fusion. A proposal was written by ORNL entitled "Reasserting U.S. Leadership in Scientific Computation"; it was submitted to DOE on July 4, 2002, and called for evaluating and deploying the Cray X1. The proposal was funded in August 2002.

The peak-10-Tflop/s systems (the CCS-3, NERSC-3E, and CCS-4) among the Office of Advanced Scientific Computing Research (ASCR) computers were compared in terms of number of processors, number of processors per cabinet, maturity and stability of the systems, and cost. The costs are in the same ballpark; the Cray processors are much smaller than the IBM ones. There is a need to ensure that software can be migrated from one machine to the next; this is covered by a memorandum of agreement (MOU) with Cray. It is planned that the projected Cray X2 (Black Widow) will share interconnect technology with the projected Red Storm (a Linux-service machine). ORNL's CCS is a member of the Defense Advanced Research Projects Agency (DARPA) High-Productivity Computing Systems (HPCS) program, which is looking at shared memory locales, heavyweight processors, and processors in memory (PIM) for light-weight processes (LWP).

A four- phase evaluation and deployment of the Cray X1/ X2 is planned:

Phase 1	3.2 Tflop/s, 1 TB, 20 TB
Phase 2	8.192 Tflop/s, 2.621 TB, 24 TB
Phase 3	40.96 Tflop/s, 13.107 TB, 102 TB (an Earth-Simulator-class machine)
Phase 4	120 Tflop/s, 40 TB, 400 TB

The first cabinet of the Cray X1 was being shipped the week of this meeting and was to be installed the following week. It is expected that, in summer 2003, the system will be operating at 3.2 Tflop/s, with 256 processors, with 1 TB of shared memory and 32 TB of disk space in eight cabinets.

A detailed evaluation plan was developed in concert with user community and is available at www.csm.ornl.gov/meetings/. It is an evaluation and benchmarking effort that is driven by the science applications. Projections of the resources that will be needed by climate simulations during the next 10 years indicate that machine speed will have to increase from 3 to 750 Tflop/s and data production per run will increase from 1 to 250 TB to support the applications. In climate science, CCS is collaborating on the

- Community Climate System Model (CCSM),
- Community Atmospheric Model (CAM),
- Community Land Model (CLM), and
- Parallel Ocean Program (POP).

Porting strategies to the Cray X1 and NEC SX have been developed and scheduled for the CCSM. The Cray X1 is the fastest processor cluster available to the climate-research community. It is a factor of 5 faster than other machines for delivering the science. For dynamical cluster approximation/quantum Monte Carlo (DCA/QMC) calculations of strongly correlated electronic materials, the Cray X1 performs Monte Carlo analyses a factor of 15 times faster than other available machines. The Cray X1 looks like it has promise for running Boltztran, which calculates the Boltzmann neutrino transport, but that code is, as yet, an immature software system.

In summary, the Cray X1 offers a balanced architecture for science. It provides architectural diversity and unprecedented performance on SC applications. Access is allowed to the operating-system source code to tune the system for DOE applications. A strong collaborative partnership has been forged that will provide an opportunity to guide the development of the next-generation Black Widow system based on DOE applications.

Wright introduced **Horst Simon** to talk about the recent upgrade at the National Energy Research Scientific Computing Center (NERSC) and its plans for the future. The facility serves about 2000 users working on about 400 projects. It has a staff of 78. A strategic proposal was submitted in 2001 that defines NERSC as a general-purpose, full-service-capability center that focuses on high-end computing requirements.

In FY02, NERSC 3 (Seaborg) was upgraded to 10 Tflop/s and the high-performance storage system (HPSS) storage capacity was increased to 7 PB. It reached >95% utilization and received excellent ratings from users. It will be an IBM test site for Grid computing.

Demand for NERSC resources has increased dramatically. The 5-year plan projects a more than eight-fold increase in peak teraflops, calling for upgrades every 3 years. No machines are available that meet the needs and expectations. To deliver the cycles to the users in a timely manner, NERSC 3 will be upgraded and doubled in size instead of a NERSC 4 being installed.

The upgraded system has 416 16-way Power 3+ nodes with each CPU at 1.5 Gflop/s, producing a total peak performance of 10 Tflop/s (1.4 Tflop/s sustained) and an aggregate memory of 7.8 TB. IBM was selected as the vendor, but a Power 4 was not selected because of a lack of sustained performance.

The system is being used for:

Accelerator physics: 4%
Astrophysics: 13%
Chemistry: 9%
Climate and environmental sciences: 9%
Computer science and mathematics: 1%
Earth and engineering sciences: 1%
Fusion energy: 24%
Lattice gauge theory: 19%
Life sciences: 3%
Materials science: 14%
Nuclear physics: 3%

DOE laboratories use 56% of the resources, universities 36%, other laboratories 5%, and industries 3%.

DOE initiated a new allocation process for FY03. A lot of data are gathered, and the Computational Review Panel makes allocation decisions, including "Big Splash" allocations (72%; these are a small number of projects that require lots of resources to develop their computational science). The number of projects at NERSC has significantly decreased; at the same time, the amount of hours has significantly increased. This situation reflects a trend toward high-end computing.

Within six weeks of the upgrade, the system was running at 95% utilization. High priority is being given to large jobs that use a lot of time to accommodate high-end computer needs. Hardware developments will produce a divergence of peak and sustained performances as machines get bigger and faster. A joint workshop was held by NERSC, Argonne National Laboratory (ANL), and IBM to let vendors know what scientific users need. Usually the processor architects hold his type of workshop with companies like PeopleSoft and Oracle. This is the first time these designers have met with scientific users. IBM has incorporated the input from this workshop into their product line.

The science results from the NERSC 3E HP Linpack benchmark were 7.21 Tflop/s at 72.2% of peak, more than double what was produced on a smaller system. High sustained performance, in some cases up to 60% of peak, was also demonstrated on applications related to

- Electromagnetic-Wave–Plasma Interactions by Don Batchelor of ORNL,
- Terascale Simulations of Supernovae by Tony Mezzacappa of ORNL,
- Accelerator Science and Simulation by Kwok Ko of Stanford Linear Accelerator Center (SLAC) and Robert Ryne of Lawrence Berkeley National Laboratory (LBNL),

- Quantum Chromodynamics at High Temperatures by Doug Toussaint of the University of Arizona,
- Cosmic Microwave Background Data Analysis by Julian Borrill of LBNL and the University of California at Berkeley, and
- A Parallel Climate Model by Warren Washington of the National Center for Atmospheric Research and Albert Semtner of the Naval Postgraduate School.

The last application invalidated an analysis of observational data that were interpreted to say there was no global warming.

In summary, NERSC implemented an upgrade to 10 Tflop/s successfully and is delivering a new capability to the DOE SC community, it has excellent scalability on many large-scale applications, it has a high sustained performance on levels comparable to those of the Earth Simulator, and it has produced new science results.

Wright declared a break at 10:10 a.m. She called the meeting back to order at 10:30 a.m. for questions addressed to Zacharia and Simon. Sollins asked Zacharia what operating systems were planned to be used in the upcoming machines and how well they would be tuned to the architectures. Zacharia noted that Steve Scott, who is knowledgeable about the architecture of the Cray X1, was present. What was sought was a machine that can do the science with as little pain as possible. A meeting was held for users and computer manufacturers. The manufacturers said that the scientists did not know how to program, and the users disputed that allegation. Scott said that the operating system for the Cray X1 was developed from that of Silicon Graphics, enhanced with features from UNICOS (the UNIX Cray Operating System) and other capabilities. Sollins asked what would not be served by this procedure. Scott responded that some things that did not run well on previous machines will run well on these machines (e.g., sparse-matrix problems). The Cray compilers are based on those used for 25 years and are very efficient. Things that run well on clusters (e.g., dense linear algebra) might not run well on these machines.

Connolly noted that one cannot make everyone happy. Both NERSC and ORNL's CCS want to get science done, but they have different roles. Zacharia noted that both centers support the DOE mission and they continually adapt to that mission. One center cannot exist without the other, though. The size of a job and other variables determine which center's resources should be used for a given project or application. The ORNL center brings technology to maturity. Simon added that, today, computational scientists know their code intimately, having written it many times for different machines. People want to get science done; they do not want to spend their time rewriting codes.

Kulsrud asked if they could qualify the mixture of architectures to indicate what percentage of applications belong on a new machine rather than an older one. Zacharia replied that users tend to be agile. The ORNL center tries to give preference to people who need large-scale computing. The climate community has a schedule for advances and computer use. The centers have to make a judgement on what the climate (and other) scientists will need in the future and make sure that that it is available when they need it. These classes of computer give people the capability to do fundamental research. The hardware features of the Cray make it especially apropos for biological research. Kulsrud asked what percentage of users at his site will move over to new machine. Zacharia said that he expected a large percentage to do so.

Kulsrud asked what happened to the vector-computer users at NERSC. Simon answered that they had about 160 projects on the vector machines. Most had dual accounts and had already transitioned when the vector machines went away. The other users were able to move to workstation resources; they no longer needed high-end resources. In terms of different architectures, NERSC has a second machine, a 500-PC Linux cluster, that is used in a production setting for high-end, high-energy and nuclear physics users. That is an example of putting the right applications on the right machines. Kulsrud asked if some of the people who converted their vector codes to run on a scalar machine might want to go back to an X1. Simon replied that some might. He expected that whatever runs well on the IBM SP will run well on the Cray X1. Some users will make the effort, and some will not. Some will simply keep running their scalable, parallel-architecture applications on cluster-based or SMP machines.

Stechel asked if it was fair to say that ORNL is a prototype plant and NERSC is a production plant. Zacharia replied that the complementarity was more significant than that. There *are* hardware differences; some of NERSC's projects started out at the CCS. It is about the science. With a large number of users, NERSC has to be more robust than the CCS has to be. Both centers have to make resources available to users in a timely fashion, but they do it in different ways. The science will drive architecture changes (and vice versa). Some users use both NERSC and CCS because they offer different capabilities. Simon noted that DOE makes all the allocation decisions for NERSC, and that this limits flexibility. Zacharia noted that the two centers have different missions and are managed differently, accordingly.

McRae noted that it can be painful to go from one architecture to another and asked what the cost for such a transition is and whether a problem can be mapped from one architecture to another with software. Simon said that he did not know what the cost is. The second part of the question is being addressed through SciDAC, giving the scientists tools to change from architecture to architecture. McRae asked if DOE is under-investing in the effort to address the problem of the divergence between peak and sustained performance. Simon responded that some advances can be made through software and algorithms; however, part of the problem is what is being measured and how it is being measured and another part is that it is an inappropriate statistic. Zacharia added that, in his view, DOE was under-investing in *both* hardware *and* software. Both hardware and software limits exist that should be addressed. Peak performance is not the best measure; four full-time staff are needed to port data for the basic U.S. climate models. Simon agreed that DOE is under-investing. Getting SciDAC back to its original intent would help. Architectures are not bandwidth focused. More money must be put into meaningful benchmarking and performance evaluation.

Giles noted that a lot of roadmaps and plans have been drawn up and asked how they all fit together in a sustained and effective manner. Simon said that NERSC has put together a local 5-year plan. There is little coordination among the laboratories. A coordinating meeting was held in January. DOE is developing a strategic (20-year) plan for all of SC this year. Zacharia said that the CCS responds to requirements set down by DOE, which is guided by its own strategic plan. He has looked back on the projections that were made 15 to 20 years ago and has been amazed at how accurate those projections have been. He also pointed out that the fact that the CCS and NERSC have been funded is indicative that they are operating in conjunction with the DOE strategic plan.

Wright asked them what the optimal strategy should be for the different laboratories to work together. Zacharia replied that he looks to the Department's needs and what its management says. Computational science has to play a key role in scientific advance. DOE should aspire to be a leader in computational science in all areas of science. That requires investment in software, algorithms, and mathematics, but something is needed to run that software on. A balanced investment is needed. Wright asked if the way that facilities and architecture are managed now is the right way. Simon responded that, in High Energy Physics (HEP), the High Energy Physics

Advisory Committee (HEPAC) has produced a report on what is likely to happen in HEP during the next several decades. Nanoscience has gotten three to five centers established across the country. Genomes to Life has a coordinated effort to sell its plan. Computer science needs the same type of planning and coordination. Zacharia pointed out that SciDAC has a 5-year plan, but it has not gone forward. DOE should be looking beyond that SciDAC document; it has to evolve and change and be used actively.

Connolly commented that computing in SC seemed fractionated and uncoordinated among the laboratories. Zacharia acceded that that was a valid perspective. The different projects have different needs and requirements. The Environmental Molecular Sciences Laboratory (EMSL), for instance, has unique needs and roles, so it has its own user group and facilities. Simon commented that the level of planning in SC should be the same as that in HEP.

Wright thanked Zacharia and Simon and turned the discussion toward the capabilities and capacities of computing centers. Giles noted that there are needs for building capabilities to support ongoing research. There is also a need for breaking new ground in computer science. Wright commented that, when people feel starved for resources, they start to shift perspectives. Giles said that that means new access to funding.

Wolff asked what the reaction has been to the new allocation scheme at NERSC. Simon said, so far, so good. It was developed with user input, but has only been in use for a few weeks. The budget language requires high utilization, and that impacts turnaround time. Zacharia said that he believed that resources should be managed according to the science performed, not the utilization rate. Oliver interjected that the General Accounting Office (GAO) said differently. Wright asked if they were constrained by this form of management. Oliver answered, no. Hitchcock commented that the Department's friends at the GAO think that processors are sitting idle, they should be able to be put to work. These metrics distort how one manages the information-technology (IT) resource. The situation is not one that DOE can fix. Sollins noted that many things do not get counted; all that is considered is useful cycles. Hitchcock said that it is worse than that, unfortunately. The vendors' numbers for maximum performance become the *de facto* benchmarks used. DOE had this battle with its own inspector general (IG) in 1988 when the IG did not understand queuing theory. Giles suggested using an analogy to military strength might be helpful. There, a standing army is maintained of the strength needed to mount full campaigns in one or two theaters, even in peacetime. Analogously, machines with high peak performance are needed even though that peak performance is not continuously needed or used. Hitchcock suggested that some advice from this Committee about appropriate metrics might be helpful. Wright suggested that the Committee invite some representatives from these agencies [e.g., the Office of Management and Budget OMB)] to attend a future meeting to discuss appropriate and useful metrics. Oliver agreed that that might be helpful.

Connolly asked if the metrics used by OMB are written down. Hitchcock responded that that was a complicated question. Wright asked Michael Holland (formerly of OMB) to comment. He said that the only metric that he remembered said that facilities (like a light source) had to be used at least 50% of the time, and it was left up to the discretion of the agency's management to determine how that criterion was to be met.

Stechel said that the Department needs to develop a meaningful metric to offer to the managing agencies.

Kulsrud asked if the personnel at OMB have workstations and, if so, do they apply these same metrics to their workstations. [Laughter.]

McRae commented that this is a complex issue and that the Committee should not get hung up on one, inappropriate metric. Wright said that not all the problems can be solved with more money; those that cannot need serious consideration. Stechel stated that the problem is more complex than just funding availability. The Committee and the computer-science community need to spend more time figuring out how to get more out of constrained resources.

Wright adjourned the meeting for lunch at 12:02 p.m. She called the meeting back into session at 1:45 p.m. and introduced **James Decker** to present an update on the Office of Science (SC). He thanked the Committee members for their work on the various charges put to them by the Director of SC.

He reviewed the information for the FY04 budget, pointing out that $86 million is freed up by the rolloffs produced by completion of the Spallation Neutron Source (SNS) and other projects, resulting in about a 4.5% increase in funding for science. In the FY03 appropriation, Congress added about 30 projects and funding for about 20 of them in addition to other unfunded Congressional directions. Also, $20 million of SC's funding was transferred to the Office of Homeland Security. As a result, the program of the Office of Biological and Environmental Research is down about $50 million.

SC provides more than 40% of the federal support for the physical sciences and constructs and operates large scientific facilities for the U.S. scientific community (e.g., accelerators, neutron sources, and light sources). About half of its nonfacility funding goes to the national laboratories, and the other half to universities.

The priorities in the FY04 SC budget are

- International Thermonuclear Experimental Reactor (ITER; $12 million)
- Next-Generation Computing Architecture ($15 million)
- Nanoscale Science, Engineering, and Technology ($196 million)
- Genomes to Life ($67 million)
- Climate Change Research Initiative ($25 million)
- SciDAC ($62 million)
- Workforce development ($6 million)
- Facility upgrades at Fermilab and SLAC ($447 million)

The re-entry into the ITER was announced by the Secretary and the President late in the budget process. On Feb. 6, 2003, the President said "the United States will work with Great Britain and several European nations, as well as Canada, Japan, Russia, and China, to build a fusion test facility and create the largest and most advanced fusion experiment in the world." It is a $5 billion experiment that will produce 500 to 700 MW for 400 seconds to 1 hour with a burning plasma (fusion energy will heat the fuel). Its objective is to demonstrate the scientific and technological feasibility of fusion energy for peaceful purposes.

In nanoscience, DOE is providing funding for the establishment of four Nanoscience Research Centers [at ORNL, LBNL, Sandia National Laboratories/Los Alamos National Laboratory (SNL/LANL), and (BNL)] and $10 million for equipment for the ANL center, which is being constructed with $36 million of state funding.

In Genomes to Life, the objective is to build on advances in sequencing, molecular science, and computing to understand and harness microbes to address DOE's energy, environmental, and national-security missions. Genomes to Life will continue the complex task of characterizing all of the multiprotein molecular machines and their associated regulatory networks in microbes of importance to DOE's missions. A $24 million increase in funding will focus on characterization of molecular machines, on development of broad capabilities for large-scale protein production and diverse molecular-imaging approaches, and on DNA sequencing of individual microbes and microbial communities.

The climate-change program will expand field measurements of clouds' effects on Earth's radiation balance and will improve terrestrial carbon-cycle models and our understanding of carbon cycling to identify and quantify the North American carbon sink.

In workforce development, the laboratory Science Teacher Professional Development Program will help teachers become leaders in their profession by bringing grade- and high-school teachers into ANL.

A "good-news" story is that the operating time at user facilities has been improved to between 83% and 100% of maximum as a result of a $38 million investment in these facilities.

McRae noted that computing is vital to the success of all of these programs and asked if DOE had a strategic plan for computing. Decker said that SC is putting together a 20-year plan and is looking at all of the United States' high-performance computing as part of the FY05 budget process.

Wright asked about the status of the bills currently before Congress for increasing the funding for science. Decker replied that those bills seem to be moving forward. Wright asked if that could affect the FY04 budget. Decker responded that it is difficult to say. Wright asked if SC had any contingency plans on how to spend that money, should it be appropriated. Decker replied that they certainly do.

Stechel asked him to list the parts of the strategic plan that are needed. Decker answered that SC needs to look at the research needs across all the programs. Oliver commented that SC has plans from many projects, and those need to be stitched together and the overall plan filled in. Decker said that it is a very dynamic process.

McRae pointed out that the semiconductor industry has been very successful in developing strategic plans [see the International Technology Roadmap for Semiconductors (ITRS) 2002 Update at http://public.itrs.net/Files/2002Update/Home.pdf] and asked why DOE does not have a program like that. Decker responded that strategic planning has been conducted in a lot of projects and in parts of DOE; the process is not foreign to DOE.

Wright observed that this meeting had been shifted to March so comment could be made on the ASCR budget. She called upon **Edward Oliver** and **Walter Polansky** to comment on the ASCR program and budget. Polansky reviewed the staff assignments and noted that there are three vacancies.

The mission of ASCR is to discover, develop, and deploy the computational and networking advances that enable researchers in the scientific disciplines to analyze, model, simulate, and predict complex physical, chemical, and biological phenomena important to the Department. ASCR carries out this mission with a two-prong program: by supporting a broad research portfolio in advanced scientific computing (including applied mathematics, computer science, networking, and collaboratory software) and by operating supercomputers, a high-performance network, and related facilities.

He analyzed the budget crosscuts from previous years. The budget has been flat but indicates a major shift (between national laboratories and universities) in allocations. This apparent shift resulted because the FY01 appropriation was late. As a result, the laboratories were funded for only three months in FY01, and that deficit was made up in FY02. The universities were forward-funded for 21 months in FY01.

Slicing the budget differently shows that from the FY03 request to the FY04 request, base research is flat, SciDAC is flat, facilities are down, and next-generation architecture is up. In FY03, the President's request was $163.557 million, and the appropriation was $164. 480

million; in FY04, the request is for $170.490 million.

The Mathematical, Information, and Computational Sciences Division (MICS) activities reported at the October 2002 ASCAC meeting were updated:

- A workshop and eight town meetings were conducted to evaluate the Earth Simulator's impact. A science case has been prepared; it includes 15 working documents; 11 final release documents are available on the Web at http://www.ultrasim.info.
- An Early Career Principal Investigator activity was launched to strengthen the core research program. In FY02, 17 awards were made; the FY03 call for proposals closed February 20, 2003.
- An ASCAC-BESAC (Basic Energy Sciences Advisory Committee) workshop on computational nanoscience was held. The results of that workshop served as the basis for a call for proposals (http://www.science.doe.gov/grants/Fr03-17.html), which is now on the street.
- A workshop on networking requirements for future science was conducted. Discussions are continuing, leading to a consideration of a "network environment."
- Several Genomes to Life workshops on applied mathematics and computer science were conducted. Collaboration is continuing and is revealing many opportunities for ASCR to work cooperatively with the Office of Biological and Environmental Research.
- An ESnet backbone upgrade was initiated to increase transmission from 622 Mb/s [OC12 (optical carrier, level 12)] to 10 Gb/s (OC192) in order to service the increased networking requirements of science. The northern route was at OC192 as of March 5, 2003; the southern route was at OC48 en route to OC192. This level of performance is about one year ahead of original projections because of market conditions.

The planned activities during FY03 are to

- Initiate reviews of applied mathematics and collaboratory pilot research activities (laboratory activities will be reviewed in October; collaboratories were being reviewed at the same time as this meeting)
- Initiate a review of the SciDAC portfolio [the computer-science Integrated Software Infrastructure Centers (ISICs) were under way; the mathematics ISICs will be reviewed in May]
- Continue workshops and town meetings to assess ultrascale simulation needs (no recent activity)

With the FY03 appropriation, Advanced Architectures got $3 million per Congressional intent. This funding will be used for lease payments on the Cray X1 and will allow an early start on a meaningful evaluation. Base research and SciDAC were continued at FY02 levels. Computational nanoscience was funded at $3 million in ASCR with a similar amount in Basic Energy Sciences; 64 preproposals have been submitted. Genomes to Life was continued at FY02 levels. The Early Career Principal Investigators program was funded; 65 grant applications were submitted and are now under review.

The objective of the Next-Generation Architecture program is to identify and to address major architectural bottlenecks (e.g., internal data movement in very large systems) in the performance of existing and planned DOE science applications. The major activities anticipated are research on the impact of alternative computer architectures on application performance, research to improve application performance and system reliability through software development, and evaluation with hardware testbeds of sufficient size to understand key issues.

The ASCR programs are driven by the applications community. Applied Mathematics, Computer Science, Network Environment, Scientific Applications, and Genomes to Life are going to expand. ASCR is looking forward to the future when simulation needs will require a sustained computational capability of from 50 to more than 100 Tflop/s. An example of such an application is the need to understand the behavior of a design (like a fusion torus or an accelerator beamline). Such an application requires input from mathematicians for

- the writing of a mathematical description that is scientifically accurate across multiple scales with proper boundary conditions;
- discretization with mesh technology, functional analysis, robustness, efficient computability, and proper treatment of boundaries; and
- a computational solution that uses high-performance computing with accurate, efficient, scalable, tunable, robust, modular, and fast numerical algorithms.

As one looks at advanced architectures and systems, one needs to consider memory management that is intelligent, dynamic, adaptive, and under the programmer's control through new language and/or compiler techniques. One must also consider the legacy issue: SMP clusters have driven the development of system software, libraries, and applications for the past 10 years. As a result, one might want to think about multiple architectures. In a projected science application, speed goes from today's 1 Gb/s or 100 Gb/s, a bottleneck would occur in the site-security process, and a research opportunity would occur in the end-to-end performance gap.

Discussions with SC staff and researchers at universities and national laboratories have led to the identification of a number of opportunities for ASCR:

- To establish/strengthen strategic partnerships with other SC programs (e.g., in Genomes to Life, nanoscience, and fusion energy)
- To embark on a sustainable path to provide high-performance computers for science through industry partnerships (involving all of the vendors with multiple acquisitions and strategies) and architecture research (which will play a major role)

These efforts will be made successful by providing scientific foundations in applied mathematics and computer science in areas that are barriers to world-leadership in computational science (e.g., in multiscale mathematics and in operating systems and programming environments). It would be desirable to restore the vitality of the base research to FY92 levels and to build on the SciDAC success

What the agency is looking for down the road is set forth in the highlighted passage in "Research and Development Funding" in http://www.whitehouse.gov/omb/budget/fy2004/pdf/spec.pdf (see page 181).

Wright declared a break at 2:59 p.m. She called the meeting back into session at 3:30 p.m. for members to address questions about the ASCR budget and directions to Polansky and Oliver.

Connolly asked about the "high-end core technologies" cited in the "Research and Development Funding" writeup. Polansky replied that this effort involves the National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), Accelerated Strategic Computing Initiative (ASCI), Department of Defense (DoD), and DOE; more would be presented about this effort in Hitchcock's presentation later in this meeting. Connolly asked specifically about the reference to "ultrascale simulation." Polansky responded that it referred to the town meetings (the results of which are available on the Web at http://wwwultrasim.info) and "the matrix," which appears on the Web at http://www.appsmatrix.info.

Kulsrud asked what activities would be funded by the $3 million in FY03. Polansky said the money will go toward lease payments for the Cray X1; it is unsure what that money will be used for in FY04; that is in the planning stage.

Wright asked what he meant by "planning." Polansky responded that the funds will be disbursed through competitive calls for proposals, so the actual funding will be determined by the results of the proposals submitted. Kulsrud noted that in his slide No. 11, Polansky had given the teraflops needed rather than the total computation time needed. She stated that that was not meaningful and asked for more information. Oliver said that ASCR would like the application community to say what metric should be used.

Meza noted that Polansky had expressed a desire to go back to the funding level of FY92 for base research and asked what that level would be. Polansky replied that, to meet the needs of SC, ASCR needs an increase in funding of a factor of 2. Meza asked if there was a roadmap or plan to restore that funding. Oliver noted that the slide referred to had a list of good, representative activities that had been put forward but not funding.

Stechel said that, in order to build on the success of SciDAC, one needs to know why SciDAC has been a success and why prior efforts were not successful. Polansky said that SciDAC will be a rousing success because the MICS support has been joined with the application experts. The program was able to be sold because it had an element that resonated with the funders. Previous, similar attempts to get funding were accompanied by large requests for hardware. There are many competing reasons why projects get funded. Laub commented that SciDAC was successful partly because money was dangled before people to get them to do critical tasks cooperatively. It also came across as an SC initiative, which carried more clout than an ASCR initiative.

Giles was amazed that there were no results behind the town meetings and other activities that have called for growth of ultrascale computing and other computing. Polansky answered that it is crucial to make the case on science grounds and on the basis of the SC mission. That is where the Committee can help. Often, a good idea does not get sold the first time; one has to keep plugging away at it. Oliver noted that ASCR has to make the best scientific case for investment in advanced computing; it cannot control the other influences on funding success.

Sollins referred to slides 14 and 15 in the presentation and said that they did not make a case for networking. She pointed out that some fields of research do not analyze 90% of the data collected; others ship disks and whole computers by UPS because they do not have the requisite networking. That pipeline for funding has to be built year after year. Research is needed on protocols. Grids are going to make the congestion problem even worse. A significant research agenda is needed and does not appear in this presentation. Wolff added that the budget today is one-sixth that for the NSF's networking research in 1996. Sollins went on: without adequate networking, the use of multiple machines with different capabilities will be impossible. Too many things have been lumped together. The resulting request is not understandable. ASCR should ask for $20 million or $30 million rather than for $7 million for networking research. DARPA is a disaster for networking research. NSF is overrun with proposals for networking research. Oliver asked, if ASCR got an additional $20 million in its budget, how it should be allocated. Sollins said that she would give it all to networking. Polansky said that, if the network-research community feels the need for increased funding, they should approach DOE with a rationale for that increase. Sollins responded that they will not come to DOE because DOE does not have any money for networking research. Polansky said that someone has to make the case that networking research is necessary to SC.

Wolff noted that ASCR is projecting a 4-Tb ESnet; that capability is being produced now by others.

Oliver stated that the Office needs the numbers by May to make a case for the FY05 budget.

Wolff suggested that a workshop be held.

Wright said that she cared deeply for the Computational Science Graduate Fellowship (CSGF) Program and noted that it had not received an increase in funding. She asked what other programs had not been mentioned and why. Polansky offered as one specific example the applied mathematics base program, which has proposals requesting four times the available funding. The Genomes to Life activity had a call in FY01; 10 or 11 projects were funded for $3 million. Many proposals passed peer review but were not funded. Many credible nanoscience proposals will be received, and the reviewers will need to make some tough decisions. Many opportunities exist to work with Fusion Energy, but funds are not available to underwrite such activities this year.

Wright noted that mathematicians and computer scientists bring a different perspective to science. This Office should explain the importance and effects of algorithms and mathematics to Congress and others involved in the funding process. Oliver said that he could not agree more. Networking and other research interests are important, also. In addition, facilities need to be supported and expanded.

Connolly noted that Pacific Northwest National Laboratory (PNNL) has great computational needs that are not being met by the centralized facilities, so they are building their own and asked if this was the way to go. Polansky responded that a stronger argument could be made for coordinating the resources of EMSL, NERSC, and CCS better.

Kulsrud stated that someone has to take the initiative in seeing that new architecture is developed and made available and asked how that could be brought about. Hitchcock responded that architectural research is done across the federal government, and DARPA does a good job at that. DOE needs to invest in niches that are not being funded and to leverage off the investments of others. Kulsrud replied that that strategy resulted in too long a time frame and would be only one effort. Hitchcock pointed out that DOE had not been successful in getting funding for architecture development in the past; maybe it will be successful in the future. In the meantime, it is trying to work with other agencies to fill the gap. That strategy has certain risks, but it is in accord with the President's management strategy. Kulsrud said that DOE needs to know what its requirements really are. Wright interjected that it is dangerous to say the an application is equivalent to an algorithm is equivalent to a code that runs on a computer. An interesting new architecture that a number of people can use would be a good investment. Kulsrud pointed out that it is known that problems have been solved when new architectures came along because people had not thought about the problem in those terms before.

Giles said that someone has to ask the questions of what needs to be done, when, at what level of funding, and with what relationship to other funding needs. In other words, a plan is needed.

McRae pointed out that, in real terms, the budget of the Office is going down. He asked, if computing is crucial for SC's science, what needs to be done to get the budget ratcheted up to the needed level. Polansky responded that the agency needs to make a case for a funding set point to meet its mission in tandem with theory and experiment. At the SC level, it has to be articulated on a 10-year time frame what new understanding will be produced and how high-performance computing will contribute to that result. Also, the research agenda has to be advanced. McRae asked who has to be influenced to get this job done. Oliver responded that, first, one has to convince Orbach (Director of SC), then one has to convince the Undersecretary, then one has to get the Secretary's approval; then get OSTP's approval, and the OMB's approval. Each one has different interests.

Connolly asked how much the other offices in SC spend on computing. Hitchcock responded that the number is $70 million to $80 million per year for scientific systems only; this number does not include what the universities spend with their grants.

Wright asked Ari Patrinos, if there was some way that this Committee could convey the importance of computing to the research community. Patrinos commented that the percentage of good ideas that get funded is very small. All of the parties involved at the Office level must come to closure on the one thing that will be absolutely important for scientific computing to do to help SC achieve its mission. That will improve the chances of funding tremendously. The Chair should also empower members of the Committee to learn what is available across the federal government rather than relying on tidbits of information that may or may not be reliable, allowing the Committee to put its initiative in a national perspective. All of the players must act collaboratively.

Kulsrud asked what mechanism should be used to give advice. Oliver suggested that the Committee form a subcommittee and submit a report. Individual members can also comment in committee on the information presented by speakers, saying what is good, what is not, and what is missing. Patrinos noted that any committee member can make a motion about what should be the highest priority for ASCR in the FY05 budget. The Committee can speak with one voice right now. Wright asked him if that is what the Biological and Environmental Research Advisory Committee (BERAC) does. Patrinos responded that BERAC does not take that path usually but has at times. BERAC functions mostly through standing subcommittees because of the breadth of scientific subject matter that it covers.

Sollins asked if the Committee was limited in what it can do to what it is charged to do by Oliver or Orbach. Wright responded that the charter allows the Committee to initiate actions.

Wright noted that the Committee could return to the FY05 budget on the following day. Giles asked if there would be a vote on funding priorities. Wright offered that the Committee could talk about priorities in the executive session at the end of the meeting. She opened the floor to public comment.

James Corones said that the fusion community has issued a report on planned research and simulation. The report is available at http://www.isofs.info./isofs/index.shtml. That report is a strong endorsement of the collaboration between applied mathematics and the fusion community.

Connolly asked if Orbach was expecting a response from each office of SC on the 20-year strategic plan for facilities. Wright responded that he was and that the Committee was going to discuss its response at the following day's session.

Reinhold Mann commented that the EMSL computer facility is an integral part of EMSL and has been a great success as a user facility. What is important is an overall-enterprise perspective. One must ask how each facility fits in with every other.

Roy Whitney commented that other advisory committees have representatives of national laboratories that provide staffing support for subcommittees. This Committee could benefit from such membership. The Committee's charter provides the right to make recommendations about such membership. Also, the Committee could ask each member to come to the next meeting with a plan about what to do about the FY05 budget. The national laboratories are making plans, and the Committee could ask to see what those plans are as they relate to scientific computing.

Hitchcock noted that, by the next meeting, all of DOE's plans for the FY05 budget would be embargoed, but the Committee could ask about the strategic plan or about DOE's roadmap for future activities. Oliver said that, at this time, the SC is soliciting advice about what the FY05 budget should be. Once those numbers are set and under internal review (around September), the budget is embargoed, and the numbers and details cannot be discussed. The Secretary wants the freedom to look at the whole portfolio and consider the big picture without constituency pressures.

Whitney pointed out that other advisory committees consider plans that have general priorities and scopes. This Committee could ask ASCR to prepare and present such a plan, and ASCR could turn around and ask this Committee to volunteer to write that plan.

Alan Laub noted that there are plans around. The High-End Computing Task Force was meeting the following week to ask each agency what it will need and do in high-end computing.

Rick Stevens said that this Committee is the most well-positioned committee to determine what this country is going to do in high-end computing. This Committee needs to form subcommittees to develop plans. Those subcommittees should be populated with experts from inside and outside this Committee And this Committee should ask hard questions of those subcommittees. Think big; do not bicker over $20 million here and $3 million there. This Committee is advising the Office on behalf of the community. The members of this Committee are in the driver's seat. It should be able to put together a plan that will sell itself. The window is open right now to provide input for the FY05 budget. Wright pointed out that, for the FY03 budget, this Committee advocated what has become ultrascale computing. It got no response. Stevens said that that was because the Committee did not push that proposal. The Committee has to form a community-neutral forum, have strong support from the program office, and come up with a coherent set of priorities. In the end, one has to get all the wood behind the arrowhead; politics is part of the game; everything has to be treated together. Giles noted that, if the Committee did that, its proposals would be less likely to run aground, but it would be a huge task.

Kathleen Kingscott observed that many of the comments that were offered on network architecture lie in the industrial community, and the Committee should seek and hear the perspectives of that community. The previous week, the National Academy of Sciences (NAS) held a workshop on the future of supercomputing that included industrial presentations. A similar industrial perspective might inform the Committee as it considered programs that are being put forward through the Office of Science. Her employer, as a member of the industry, would be happy to share such a perspective with the Committee in a cooperative mode.

James Glimm complimented Stevens on his call for leadership and for a plan for high-performance computing. The broad outline of what such a plan would look like had emerged. It is essential to have high-performance computing hardware to solve the science problems. It is clear that multiscale analysis is needed. Collaboration is needed between the application scientists and the computer scientists. Uncertainty quantification is also needed to assess the validity of what is said in the plan. The next step would be to form study groups on high-performance computing, multiscale science, uncertainty quantification, and collaboration.

There being no further public comment, Wright adjourned the meeting for the day at 5:31 p.m.

Friday, March 14, 2003

Wright called the meeting to order at 8:28 a.m., thanked the Oak Ridge Institute for Science and Education (ORISE) personnel for their support in conducting the meeting, and introduced **Daniel Hitchcock** to speak about ASCR's joint activities with other agencies. He started by pointing out that the language in the R&D portion of the FY04 President's Budget focused on "computing core technologies, a federal high-end computing capacity and accessibility improvement plan, and ... federal procurement of high-end computing systems." One problem is gaining advice about procurements without violating FACA. In terms of agency coordination, ASCR does a lot with NNSA, with $17 million of research funded at National Nuclear Security Administration (NNSA) laboratories and the development of Red Storm, all planned under formal coordination documents. It also cooperates with the DoD and the Deputy Under Secretary of Defense for Science and Technology. In the 2001 Defense appropriations bill, there was a charge to conduct a study on high-end computing. SC was invited to participate. The Intergovernmental High-End Computing study did not deal with buying systems and letting users on them; it only dealt with research. It is the most important unreleased report in the government. In addition, under NAS funding, ASCI was required to perform a study on how to coordinate efforts if both ASCR and ASCI were to be funded to go forward on larger machines; SC was also planning to conduct such a study, so it is being conducted jointly. The report is due out in July. With DARPA , the issue is whether one is the right number of agencies to conduct experiments on high-performance computing. DARPA is *the* steward of HPC for the federal government. DOE's computing requirements are similar to those of NSA, so it has several interactions with that agency on benchmarking, on Uniform Parallel C (UPC, which will run on SV1, clusters, and other high-performance computers), and on the development of the Cray SV2/X1 (Black Widow, which was discussed earlier in this meeting by Zacharia). ASCR also participates in the government-wide High-End Computing and Computation Working Group (HECCWG).

The Performance Evaluation Research Center (PERC) is part of SciDAC and is a partnership of national laboratories and universities to understand the performance of high-performance machines. The goal is to make the tools developed in this effort available to everyone; however, copyright restrictions are becoming an issue.

The DARPA High-Productivity Computing Systems Program (HPCS) is designed to provide a new generation of economically viable high-productivity computing systems for the national-security and industrial user communities between 2007 and 2010. This program is going forward into Phase II now. The program has also been active in the evolution of computing metrics. These metrics are desirable so one can be assured that the machines purchased do what is needed of them.

The Next-Generation Computer Architecture initiative is in the FY04 budget and has started the architecture evaluation. Its goal is to identify and address major hardware and software architectural bottlenecks to the performance of existing and planned DOE science applications. A roadmap will be produced, and a study on Red Storm is needed to see the differences between the use of commercial processors and the addition of vector processors.

In network and middleware, ASCR is involved in an elaborate federal coordinating group that has three teams:

- A Joint Engineering Team links federal and academic research networks; this has been going on since 1986 and is critical to making scientific computing work.
- The Network Research Team recently conducted a joint workshop on cybersecurity.
- The Middleware and Grid Infrastructure Coordination (MAGIC) Team sets standards to make sure that certificate authorities exist and work; significant research issues exist here.

The Particle Physics Data Grid (PPDG) and the Grid Physics Network (GriPhyN) are focusing on high-data-generation problems (up to 1 PB/s) that involve hard issues of data caching and distribution. DOE has always been a small player in networking research; it focuses on niche needs that were not covered by investments by agencies that were larger players.

Sollins asked if standing committees would avoid the concerns he expressed about meeting the requirements of FACA. Hitchcock answered that that concern was under discussion. Another way to avoid conflicts with FACA is to use materials that individuals have produced and are publicly available. It is important to get advice from outside the government, but it is also important to do it the right way. Laub said that the task force dealing with this issue was meeting the following week. Not much will be done on research; the discussion will focus on policy.

Kulsrud noted that several data-management areas were not touched on and asked if anything was going on in those areas. Hitchcock said that individuals are looking at such issues as data mining and data management in support of scientific efforts. There is not a coordinated program in data management. Much of that work is done in programs like Genomes to Life.

Wolff asked Hitchcock if he had gotten the idea of expanding DOE's role in network research from the interagency network-research community. Hitchcock answered, partly from there and partly as cochair of the Large Scale Networking Committee. This is inherently an interagency topic. Some thought has also gone into next-generation transfer protocols and into the Tier-One network. Wolff asked what mechanism should be used to make these decisions. Hitchcock replied that that is primarily the job of Thomas N'Dousse, who is the ASCR-MICS Program Manager for network research.

Sollins suggested that the focus of the Department should be a little broader. Hitchcock responded that the hope is that most large data-movement problems will be dealt with by NSF. Sollins pointed out that such an approach does not address the specialized, highly complex systems. Hitchcock pointed out that ASCR's budget for network research is smaller than NSF's was in 1986.

Wright noted that Hitchcock had mentioned the President's management plan and asked if the philosophy of acceptable redundancy is going to be maintained. Hitchcock replied that DOE battles with this problem regularly. If everything was perfect, one would not need redundancy because all decisions would be correct. But the world is not perfect, and one has to take actions that will reduce risk at the leading edge in the future. Sometimes the Department wins those arguments, and sometimes it does not. Wright noted that this Committee would be glad to comment on behalf of the Department. In industry today, strategies for R&D are misnomers; industries are looking to the government to do the research.

Wright introduced **Helene Kulsrud** to speak about the ASCAC Subcommittee on the Future of Large Facilities for High-Performance Computing. She reviewed the charge to the Subcommittee from Ray Orbach:

> Join me in taking a new look at our scientific horizon and to discuss with me what new or upgraded facilities in your discipline will be necessary to position the Office of Advanced Scientific Computing Research at the forefront of scientific discovery. Please start by reviewing the attached list of facilities assembled by Dr. Ed Oliver and his team, subtracting or adding as you feel appropriate, with prudence as to cost and time frame. For this exercise, please consider only facilities/upgrades requiring a minimum investment of $50 million.
>
> Provide me with a report that discusses each of these facilities in terms of two criteria:
>
> A. The importance of the science that the facility would support. To what extent it would answer the most important scientific questions, etc. Organize the facilities in three tiers such as: absolutely central, important, and don't know enough yet.
>
> B. The readiness of the facility for construction. Answer in three tiers: ready to initiate construction, significant challenges to resolve, and mission and technical requirements not yet fully defined.

The question arose whether A and B should be answered in terms of both (1) discovery and potential achievements in applied mathematics, numerical analysis, high-end computing, algorithms, networking, distributed systems, etc. and (2) connections between these areas and "classical DOE" problems like fusion, high-energy physics, materials, etc. plus newer ones like biology. The Subcommittee assumed the answer to this question to be yes. The time frame was restricted to 5 to 10 years.

Ed Oliver supplied a list of current ASCR facilities: NERSC ($50 million to $99 million), ESnet, ORNL CCS ($50 million to $99 million), and Ultrascale Scientific Computing Capability (>$1 billion).

Kulsrud reviewed the membership of the Subcommittee. The Subcommittee decided to examine six topics:

- The computer centers,
- ESnet,
- Hardware,
- Software,
- Joint projects within the Office of Science, and
- Joint studies with other government agencies.

Each member took one topic. A one-day workshop was held that covered the six topics. A summary of the findings follows.

The computer centers are doing well what they are supposed to. A sustained commitment is needed to support and develop center facilities. There are questions of capability and capacity. A strategic plan is needed. These centers need more funding. The question should be asked whether adding another center would help or hinder the effort; after all, two of the centers are adding floor space.

With flat funding and staff, ESnet will be able to meet the needs of the user community through 2005. After that, the cushion between capacity and demand would be gone. ESnet users are beginning to desert. Reliance on bigger bandwidth and corresponding routers would be prohibitively expensive. Will ESnet be needed in the future? If ESnet is to be kept, then substantial research must be put in to create something like the three-tier network.

In terms of hardware, it is difficult for the government to influence the vendors; its scientific computer users are just a small portion of the market. Projects such as DARPA HPCS offer a chance to have a change of direction. More knowledge about the plans of the vendors is needed. DoD/DOE currently have some projects that could influence the development of architectures. More must be known about the specific hardware needs of the DOE applications.

Assumptions about software are everywhere, but explicit planning has been scant. There seems to be knowledge of the work needed to solve problems but not of the computation times needed. If ASCR is to have a leadership role in developing and deploying software, it must have a plan for sustaining a production environment that works at petaflop scales. SciDAC is addressing the research for applications, and the computer centers support of the users, but there is a gap between the two. There is almost no software support for the larger, emerging hardware. Training for the next generation of HPC software developers needs to be part of the plan.

Joint studies within SC that will use large amounts of computing resources include nanoscience, fusion, and biotechnology. Although large programming and computer use will be required, few computer resources are being provided. The amount and cost of programming and computer use are not even known for nanoscience and biotechnology, and the magnitude of computer use is not known for fusion.

Six recommendations are being considered:

- Each of the three computer centers should receive a budget in the range $50 million to $100 million to cover both hardware and software, and it might be a good idea to add one or more additional centers.
- ESnet should plan on the three-tier project doing the necessary research so it can go on beyond 2005.
- Industry projects to create new or special hardware should be supported.
- An ultrascale computing environment plan should be developed that includes a major component of support for sustaining a petascale software environment.
- A stand-alone software and system development facility with its own support hardware should be established.
- The three centers and ESnet should be managed as one facility with a strong tri-laboratory management structure.

The Subcommittee has identified some large facilities. More knowledge is needed, raising the question of how to get it. Two possibilities are to hold a town meeting (perhaps by e-mail) or to hold additional Subcommittee meetings.

Wright commented that the Subcommittee's recommendations were well balanced. Oliver noted that, years ago, Alvin Trivelpiece had looked at the need for large facilities. That study is being redone, and all SC offices are being asked to list their facility needs. The Earth Simulator has raised the bar monetarily, and Orbach wants all offices to look at their needs in the context of that new standard.

Giles asked how set the deadline for this report was. Oliver responded that six of these reports have been requested; two have already been turned in. Drafts had been requested by the day of this meeting.

Sollins commented that people need to work in a nonproduction (but realistic) environment in order to develop new software. She asked what level of granularity needed to be represented in such a facility. Kulsrud responded that the Subcommittee had tried to work out such issues, so the question was well taken. Sollins noted that overlays that sit above the network and use resources on the network were being looked at. Such an arrangement will move this concern along more quickly. Kulsrud answered that there would certainly be multiple architectures, but the Subcommittee has not yet had time to think about the problem.

Wright suggested that the Subcommittee could say that a unified effort in networks, data management, ultrascale computing, etc. is desirable. That would meet the $50-million criterion.

Connolly asked if it would be possible to get from the user and application communities an estimate of their needs for computer resources. Meza called attention to the need that the fusion community has estimated for simulation. BER is also drawing up such estimates, but not much is known about other areas. Connolly said that it seemed that estimates would be essential. Wright noted that not much time was available to hold town meetings to obtain such information.

Kulsrud asked what the Subcommittee should do next. Stechel suggested that it look at high-performance computing as a facility that needs (say) $300 million to support SC's science. To be compelling, this report should list the science that would not occur without such an investment. The case is there, but it must be made.

Wright noted that computer scientists' research is not always looked upon as *real* science. That case should be made, also.

Sollins suggested picking some examples out of (e.g.) high-energy physics data, nanoscience, etc. The other reports are going to talk about their tokamaks etc.; this report needs to talk about the computing needs. Giles said that the report needs to emphasize the capabilities, costs, and number of users that make computation a big investment over time. Stechel said that the case also has to be made for the upgrades and increases in capacity and capability of these facilities. The report needs to show what the projected needs are and how the current facilities are inadequate to meet those needs.

Wright declared a break at 10:17 a.m. She called the meeting back into session at 10:52 a.m. and introduced **Alan Laub** to present an update on SciDAC. SciDAC is an integrated program to create a new generation of scientific simulation codes that take full advantage of the extraordinary capabilities of terascale computers, using the mathematical and computing systems software to enable scientific simulation codes to effectively and efficiently use terascale computers and creating a collaboratory software environment to enable geographically distributed scientists to work effectively together as a team and to facilitate remote access. It is not just hardware; it is mathematics and computer science, also. A key observation is that advances in algorithms have led to as much advance in science as has experimentation.

SciDAC has been under way for about a year and a half. The first principal-investigator (PI) meeting was held in January 2002 in Washington, D.C. The second one was held March 10-11, 2003, in Napa, Calif.; its theme was assessing SciDAC progress. An overview of the state of the art and science of SciDAC can be obtained at the SciDAC website, www.science.doe.gov/scidac, where 76 “two-pagers” are now available; these research writeups give an insight on what the investment in SciDAC is producing, including the outputs of ASCR.

At the recent PI meeting, Ray Orbach and Jack Dongarra gave keynote addresses. Oral presentations were given, poster sessions were held, and spirited panel discussions on Closing the “Performance Gap” and Future SC Computing/Infrastructure Needs were conducted. Some recurring themes included

- the need for advances in storage and data-handling capability (terascale computing produces huge amounts of data in a hurry),
- the need for advances in interconnect and memory technology, and
- the need for and the problems involved in dealing with architectural diversity.

The SciDAC concept is really working. Teams and collaborations are yielding new science that would not otherwise have been obtained. The concept has been proven. Many discipline scientists are (re)discovering new value in mathematics, computer science, and computer science and engineering research through application to their problems and codes. A cultural change is emerging.

Computer science and engineering is team-oriented, and forming interdisciplinary teams is very difficult. The normal reward structures in academia focus on the individual, making team forming incompatible with traditional academic practices. SciDAC will help break down the barriers and lead by example. In this process, the national laboratories are a critical asset.

The initial SciDAC focus is on software, but new hardware will be needed within the next 2 years. The Japanese Earth Simulator has been a wake-up call in computer architecture. A lot of synergistic partnerships can leverage off the success of the SciDAC model. Both capability and capacity computing needs are evolving rapidly; SciDAC members are a large portion of the NERSC and CSS users. But the architectural options available in the United States today are limited. Mathematics and computer-science research will play a key role; they allow scientists and engineers to do their jobs. Nonetheless, this is not a big bottom-line item for vendors.

The SciDAC program is ripe for expansion. Many important SC research areas (e.g., materials/nanoscience and functional genomics/proteomics) are not yet included in SciDAC. It may expand to include the nanoscience research centers and Genomes to Life. Moreover, it would seem to be prudent to have a high-performance computer and simulation capability to support the investment in ITER.

Laub turned the floor over to **Anthony Mezzacappa**, a researcher from ORNL who heads the TeraScale Supernova Initiative, a SciDAC project that is investigating the gravitational collapse and subsequent explosion of massive stars (supernovae). This interdisciplinary effort involves 12 institutions, 17 PIs, and 42 people; it is part of a larger collaboration of 121 people from 24 institutions. The purpose is to ascertain the mechanism(s) of core collapse in supernovae and to understand supernova phenomenology (including element synthesis and neutrino- and gravitational-wave signatures).

Supernovae produce all the elements from oxygen to iron and half of those higher than iron. Knowing how they work is essential to understanding our place in the universe. Fundamental nuclear physics will also come from this effort. This is not a topic that is amenable to experimentation in the laboratory; models must be the experimental mode. The modeling requirements are severe and will be pertinent to other investigations. This is a new model for theory; it requires common-component architecture, good planning, debugging, and a long-term philosophy.

The project has ties to

- the Relativistic Heavy Ion Collider (RHIC) for understanding the properties of high-density nuclear matter, which is critical to our understanding of stellar-core bounce dynamics;
- the Sudbury Neutrino Observatory (SNO), which will play a pivotal role in future supernova-neutrino detection and analysis;
- the Rare Isotope Accelerator (RIA) for a better understanding of the nuclear physics of the *r*-process, a primary justification for the construction of this facility; and
- the proposed National Underground Science Laboratory, which would be the site of a next-generation supernova-neutrino-detection capability.

The effort will require tera/peta-scale 3D, general relativistic (macroscopic-scale) physics, radiation magnetohydrodynamics, state-of-the-art nuclear (microscopic-scale) and weak-interaction physics. In transport calculations, one has to deal with tera- and peta-scale sparse linear systems of equations. Hydrodynamic calculations produce 1 Gb per write and 1 to 10 Tb per variable per simulation, raising questions about how one manages, analyzes, and renders such large amounts of data. Weak-interaction calculations among nuclei and their energy levels require terascale nuclear "structure" computation.

What is being treated here is a massive star that is larger than the diameter of the Earth's orbit around the Sun. Over the course of a million years, the gravitational pull of this mass produces a superhot, superdense core that is smaller than the Earth and is made up mostly of neutrons. That core collapses and explodes in a matter of seconds, producing a supernova. The explosion of the core radiates four types of neutrinos (in matter/antimatter pairs) and produces shock waves in the surrounding mass of the star. Those shock waves deposit energy behind them as they propagate through the star's mass. To the modeler, calculating the accumulated energy is a radiation-transport problem. As the shock waves propagate through different layers of mass, they are partially reflected back to the center of the explosion, complicating the calculations. They also produce convection cells, requiring hydrodynamic calculations. The heating rate depends on the angular distribution of the neutrinos and their energy spectrum. After exploding, the star will go on to form a black hole.

To calculate the heating rate requires solving the Boltzmann constant, which is most easily done in spherical geometry. However, axisymmetry and nonsymmetrical components make the solution a 6D problem. This proliferation of dimensions is almost overwhelming and takes up huge amounts of memory, thrash caching, and big bandwidth. The initiative has performed spherically symmetric models (3D) and needs to move on to axisymmetric models (5D) and models with no imposed symmetry (6D).

The process being modeled depends on the rate of electron capture, which itself depends on the structure of the nuclei. Therefore, the initiative first carried out 1D-collapse models with Boltzmann neutrino transport, an ensemble of nuclei, and state-of-the-art electron-capture rates. It was discovered that shock-wave instability may aid the explosion and define the explosion's "shape." The instability also plays a role in inducing explosions.

The *r*-process (rapid neutron capture) occurs in the wind given off by the explosion. It produces half the elements heavier than iron. Originally, it was believed to *require* neutron-rich conditions; now it is believed that it *may* be produced in a proton-rich environment if the nuclear reactions are out of equilibrium.

If funded, neutrino-nucleus cross-section measurements at the SNS will tie into the neutrino calculations. Under a DOE-NSF partnership, the initiative will carry out multidimensional stellar collapse simulations from which gravitational-wave signatures can be post-processed, allowing the inclusion of general relativity in the models.

The initiative is working on transport developments with the Terascale Simulation Tools and Technology (TSTT) Center, where alternative transport techniques are being developed for discrete ordinates (currently used) and discontinuous Galerkin (under development). Another approach is under investigation that involves adaptive quadratures (direction cosines) for multidimensional radiation transport.

The production of 20-TB datasets requires new ways to look at the data, such as order-of-magnitude reduction with principal-component-analysis (PCA) techniques, raw data, dimensional compression, PCA, and feature extraction. The integration of data analysis and visualization has raised many issues that have to be implemented in software engineering at large. The initiative is using off-the-shelf technologies but also has to develop custom visualization. It has been integrating visualization with data analysis and networking. The initiative has been working with University of Tennessee Logistical Networking and ORNL networking groups to significantly improve data-transfer rates between "nodes" of the initiative for local, remote, and collaborative visualization.

The PERC assesses code performance on parallel platforms and identifies code optimizations to increase performance. The hydrodynamics and neutrino-transport codes have been implemented; they will prove very effective. The initiative also has several approaches to verification and validation. It has participated in United States–Japan Computational Science Roundtable and has submitted a proposal for a United States–Japan collaboration on supernova dynamics on the Earth Simulator. This project has been selected as a testbed application for the ORNL Cray X1 evaluation, and some of the codes [AB NES, AB, V2D, and GENASIS (a new 2D Boltzmann code)] are running on the Eagle, Cheetah, Seaborg, and X1.

The science outlook is that the community now has exploding models, but no realistic 2D/3D models. The fundamental ingredients in a supernova model are the neutrinos [it must represent multifrequency, accurate neutrino transport) and fluid instabilities (it must include neutrino transport, which depends on microphysics and standing accretion shock (SAS) instability]. Precision (microphysics and macrophysics) modeling is needed. Even if explosions are obtained in a model with a subset of the above ingredients, modeling efforts must push forward until all are included. Any one of these can qualitatively alter the outcome and conclusions.

The initiative is using a staged, systematic approach:

- Layer the microphysics,
- Layer the macrophysics, and
- Layer the dimensionality.

It expects significant progress this year, beginning to merge the states of the art in microphysics and macrophysics and developing the first 2D models with 2D, multifrequency neutrino transport. It has already achieved scientific discovery in no small part because of interdisciplinary collaboration. The SciDAC model is working. This is an awesome program, and it has to continue. These problems cannot be solved any other way.

But the program's budget is now constrained and cyber-limited. It is chewing up massively parallel processors rapidly. At 18% of peak on 1024 processors, it takes ½ year on a wall clock per ½ supernova second to run a 2D, multifrequency flux-limited diffusion simulation. A 2D Boltzmann transport run is a step up in computational intensity, and a 3D run is even more. A factor of at least 5 improvement in throughput (by new algorithms and new architecture) would be desirable. Investments are needed now in algorithm, code, infrastructure, and scientific development to achieve the science that is to be performed in 5 to 10 years. In addition, next-generation mission data will require interpretation. ASCR's budget language needs to be strong to support future science.

McRae asked how the participants in the initiative actually communicated. Mezzacappa replied that they call all the families together and sometimes the *Duce*. They hold two full-collaborative meetings each year and several sub-collaborative meetings. Representatives from the ORNL Integrated Software Infrastructure Centers (ISICs) are very helpful. All of these communications are supported by e-mail and telephone messages. McRae said that that is the traditional way of conducting science and asked about new ways of moving data around. Mezzacappa responded that the Access Grid has helped those who have it. McRae suggested that they might want to look at how industries do it (running enterprises around the world 24 hours a day). Mezzacappa said the researchers were making headway but need a lot of other interaction channels. McRae asked how they distinguish between numerical and physical instabilities. Mezzacappa answered that in 1D one has a stable system, but in 2D linear stability analysis is another layer.

Connolly noted that the ultimate verification of any simulation is experimental data and asked if any differences will be seen from what LIGO (the Laser Interferometer Gravitational Wave Observatory) will see. Mezzacappa replied that it was too early to say. Connolly asked if they had enough computer resources to reach a realistic representation. Mezzacappa replied, no.

Meza noted that Mezzacappa had said that they were cycle-limited and asked what computational requirements they would need. Mezzacappa responded that they were running through their allocations at 2D; at 3D, they will need an order of magnitude more.

Wright proposed a resolution: ASCAC strongly commends the Office of Science, especially ASCR, for conceiving, initiating, and implementing the SciDAC Program, which serves as an exemplar of 21st-century partnerships, connecting theoretical, experimental, and computational science and engineering. McRae moved to accept, and Sollins seconded. The motion passed unanimously. Wright asked the Committee if there were any further topics for discussion, None were raised. She opened the floor for public comment. There being none, she adjourned the meeting at 12:25 p.m.

Prepared by
Frederick M. O'Hara, Jr.
Recording Secretary
April 10, 2003

Submitted by
Margaret Wright
Chairwoman
May 30, 2003