

Minutes for the

Advanced Scientific Computing Advisory Committee Meeting

October 25-26, 2001, Crowne Plaza Hotel, Washington, D.C.

ASCAC members present:

John W. D. Connolly, Vice Chair	Gregory J. McRae
Jill P. Dahlburg	Juan C. Meza
Roscoe C. Giles	Karen R. Sollins
Helene E. Kulsrud	Stephen Wolff
William A. Lester, Jr.	Margaret H. Wright, Chair

ASCAC members absent:

Ellen B. Stechel	Warren Washington
------------------	-------------------

Also participating:

David Bader, Acting Assistant Director for Scientific Simulation, Office of Science, USDOE

Melea Baker, Office of Advanced Scientific Computing Research, USDOE

Michael Colvin, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory

James Corones, President, Krell Institute

James Decker, Acting Director, Office of Science, USDOE

Stephen Eckstrand, Acting Assistant Director, Office of Science, USDOE

Gary M. Johnson, Associate Director for Special Programs, Krell Institute

Paul Messina, Director, Center for Advanced Computing Research, California Institute of Technology

Frederick O'Hara, ASCAC Recording Secretary

Edward Oliver, Acting Director, Office of Advanced Scientific Computing Research, USDOE

Robert Ryne, Accelerator and Fusion Research Division, Lawrence Berkeley National Laboratory

George Seweryniak, Program Manager, MICS, OASCR, USDOE

Horst Simon, Director, National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory

Rick Stevens, Director, Mathematics and Computer Science Division, Argonne National Laboratory

Andrew White, Acting Director, Computer and Computational Sciences, Los Alamos National Laboratory

Andreene Witt, Oak Ridge Institute of Science and Education

Thomas Zacharia, Director, Center for Computational Sciences, Oak Ridge National Laboratory

About 40 others also were in attendance.

Chairwoman **Margaret Wright** opened the meeting at 8:50 a.m. with introductory remarks. She had visited with Director of the Office of Science and Technology Policy Marburger and Under Secretary of Energy Card. Marburger said that one cannot see the long-term effects of September 11. Before 9/11: balance in science; after 9/11: counterterrorism and balance in science. Card said he needs the help of this Advisory Committee to get out the story and impact of advanced computing. He also stressed the need for long-term support for science.

Each member of the Committee then introduced himself/herself. Wright introduced **James Decker**, Acting Director, Office of Science, to review the changed management environment at DOE.

The consequences of the events of 9/11 are having profound effects: attention to daily security issues in DOE mean increased costs throughout the DOE complex. The war on terrorism is shifting federal budget and, hence, federal R&D priorities. The administration is determined not to let the deficit get out of control. At the same time, with the anthrax scare in the congressional offices, work on the FY 2002 budget has halted, and the government is operating on a continuing resolution until October 31, 2001.

The Secretary remarked at the quarterly leadership meeting, October 16, 2001, on the DOE mission and priorities and laid out his vision for the Department for the first time:

"... our overarching mission is national security ... our energy and science programs should be judged by whether they advance this nation's energy, and hence, national security"

The National Nuclear Security Agency (NNSA) priorities are particularly obvious: guarantee the safety and reliability of the nuclear stockpile and produce a plan to address and resolve the threat of weapons of mass destruction.

Energy priorities also reflect the events of 9/11:

1. Protect the nation's critical infrastructure for energy production and delivery.
2. Implement the National Energy Plan (NEP).
3. Direct R&D budgets to ideas and innovations relatively immature in development, and ensure greater application of mature technologies.

The Department's priorities for environmental management are:

1. Review the entire Environmental Management Program, producing a plan to accelerate cleanup and to close sites where there is no longer a national-security mission. This process is now going on; the mandate is to do it cheaper and faster.
2. Complete the process of determining the suitability of Yucca Mountain for permanent spent-fuel storage.

The Secretary outlined DOE's priorities for science: "The national laboratories are rightly viewed as a national treasure. But the national laboratories and our science programs are not a treasure to be raided regardless of mission, scope, or budget. They are too important to be squandered. So, from now on, I will expect us to implement a major change in how we do business. That change means that our science programs and national laboratory work should directly relate to and support the missions outlined above. Programs and projects that fall outside those missions will not receive my support for funding without a clarity of mission and compelling circumstances." The Secretary announced a strategic mission review to be completed by January 31. It is not yet clear how this review will be carried out. The Secretary has also made clear that DOE senior management has made performance, planning, and accountability a priority: "I also expect measurable performance objectives and accountability. Where performance does not measure up, I have made clear to my entire leadership team that changes will be made." Funding decisions are now to consider how well programs are managed. To implement this new management environment, Deputy Secretary Blake has initiated operational program reviews.

A great emphasis will be placed on developing strategic plans that link to budgets and also link to performance by managers. This emphasis on strategic plans is worrisome to those in basic science and research, but we are used to planning within projects and will work through this new emphasis.

The Office of Science (SC) is just starting its strategic mission review, which is consistent with the Government Performance review Act (GPRA). SC's old strategic plan did not link to the budget well. The FY 2003 budget is where we are really supposed to integrate strategic plans and performance management.

Scientific Discovery Through Advanced Computing (SciDAC) is seen to be an important program with great potential. DOE is devoted to it, although it will be difficult to manage because of its distributed funding, its cross-cutting nature, and its many institutions involved. "Project orientation" is needed, with clear goals and deliverables, especially for applications. SC is counting on Dave Bader (the Acting Assistant Director for SciDAC) to pull this together. We need to go back to the original plan and see how we are doing.

Giles asked him to elaborate on the concept of balance. Decker said he was referring to the physical sciences and life sciences; 40% of federal funding goes to the physical sciences.

Sollins asked to what extent computing gets funding outside those two categories. Decker responded that a great emphasis is placed on advanced computing that is tied to program activities. Sollins asked what amount of SC's budget that was. Stephen Eckstrand replied that it was \$3 million for SciCAC. In the fusion program there was \$24 million for computational theoretical research, of which about 30% is relatively high-power research. About \$10 million is in computational science out of about \$240 million. Sollins said that she was trying to get at computing research, not computing for research. Oliver responded, about 10%. Wright commented that that is the point that (Under Secretary) Card was making. It is important for us to make budget makers and policy makers understand the need for research in advanced computing, not as a tool for other programs. We need to put together a statement that defines the importance of advanced computing.

Meza asked Decker to update the Committee on the effort to enlist new science managers in the Department? Decker said that SC is making some progress, but it does not know how the 2002 budget will turn out and how program-direction funds will fare.

Connolly asked how counterterrorism will be treated within DOE. Decker replied that the emphasis will probably be largely in NNSA.

Wright noted that, in funding high-tech research, you will frequently fail. Should failure be a metric of success in funding high-tech research? Are there any meaningful metrics for high-risk research? Decker replied that some feel there is too much incrementalism in program management. The turnover rate in projects that are funded needs to be investigated. SC is instituting a Committee of Visitors (COV) through the Basic Energy Sciences Advisory Committee (BESAC), and that COV will try to see how

turnover rate affects or reflects project success. It is hoped that sensible criteria will come out of this effort.

Sollins commented that a more subtle question is: If someone does research, will some use for it come out tomorrow? Sometimes it takes decades for research to become fruitful as something useful and applied. To look for useful applications on an annual basis is to court disaster. Dahlburg stated that one fair measure is how many Nobel laureates came out of the research. DOE was once mission-oriented in its research; then it pulled back and now is going once again toward mission. Decker responded, that is correct. The Department is constantly correcting its position to respond to the wishes of Congress.

McRae commented that Decker's initial charge to ASCAC was rather narrow and has since been expanded. He asked Decker what kinds of themes he wanted to come out of this Committee. Decker replied that, although the Office will have some specific issues, mostly it wants the Committee to look at SC's activities broadly. People in the Department are often focussed on the smaller portions of the budget and how they fit together. The Committee should focus on two areas:

Issues faced by the OASCR organization, such as facilities and other issues that SC would like it to help with and

All the computing activities in SC because those activities are spread across many program offices and, therefore, have only a very small presence in any one program office, like High-Energy Nuclear Physics (HENP) or Basic Energy Sciences (BES); as a result, there is not a strong focus on computing by the other advisory committees, seeking advice from them on this topic is not very effective, and a lot of crosscutting issues do not get addressed.

April Burke [of the Society for Industrial and Applied Mathematics (SIAM)] commented that there are two ways to look at it: (1) One would be to throw up our hands and say that the bureaucracy is going to be petty and squeeze all the innovation out of science. (2) The other way is to ask what the goals of the criteria should be and to frame the criteria in such a way that innovation gets built in.

Meza pointed out that in May, Decker had given the Committee a charge to look at biotech and asked if Decker wanted the Committee to enlarge the scope of that study in light of the events since 9/11. Decker responded that that was a good point, but that he would have to go back and look at that possibility. Meza quoted from the charge letter and noted that it focused on the computational side of biotech. Messina interjected that the Committee might want to take a few hours to consider how SC could contribute to the counterbioterrorist effort.

Wright opened the floor to comments.

Stevens strongly encouraged the Committee to look at the risk problem. The laboratories are supposed to look at the high-risk issues and problems. He hoped the new orientation in Washington since 9/11 would emphasize the long-term as well as the short-term efforts. Efforts need to be guided by a combination of

high-level policy, metrics, a clear definition of the program, and the avoidance of being risk-averse. We need to remember why we went into science in the first place: to make the world a better place. This Committee is in an excellent position to brainstorm on this.

McRae said that the issue of metrics comes up when people have problems discerning value in making investment decisions and asked Stevens how he would communicate to the community at large the value of advanced computing. Stevens replied that the computing community needs to steal a playbook from the National Aeronautics and Space Administration (NASA) and create a public-relations engine that showcases the technology produced. Mission is good, but excitement is better; DOE laboratories do not use television, the Web, and science writers enough.

Messina commented that many people believe that computing research happens on a scale of months rather than decades and that industry does most of the work. Also, some specific projects may fail, but ideas survive. Sollins agreed and pointed out that this is also true in other areas, such as mathematics. A complicating factor is that no one knows what you are talking about. Government funders have to live up to measures themselves, but short-term measures cannot be applied to science successfully. Only later on will the results of this research turn into tools for DOE and the scientific community. Wright added that they will not help everyone; they will help individual groups within science.

Giles expressed frustration at the absence of ways of understanding the role of risk in the scientific endeavor; that the idea seems to be to start from scratch with each change of administration. Dahlburg noted that missions get defined and redefined. Missions allow you to figure out how to achieve the mission and what to pursue along the way. The bad side is that nothing outside the mission gets funded. One has to balance the mission against basic research. Sollins pointed out that if computing is improved, a lot of issues will benefit across all the programs. Wright interjected that the scientific community does not shrink from saying this should be funded and that should not, but it objects from having to rank things on a scale of 1 to 10.

Giles pointed out that there needs to be a relationship between the length of a mission and the length of the careers of the people employed to achieve that mission. Connolly said that that raised a difficulty; there are two types of mission: one produces a new product useful to society, and the second advances fundamental knowledge. A scientist will believe in the second type of mission; everyone believes in the first.

Wright observed that, at some point, the Committee will need to take on this issue of long-term risk and metrics in some form and take swift and meaningful action on the issue.

Sollins pointed out that another issue that has come up is the need to make an investment in a group that may be needed in the future. When an organization shifts toward a mission orientation, that investment is not made, and support is turned on and off as the mission dictates. Wright said that one cannot just write a check every year; one has to have some metrics. She declared a break at 10:07 a.m.

The Committee was called back into session at 10:45 a.m. with the introduction of **Jill Dahlburg** to begin the discussion about the DOE scientific-computing facilities, which is the topic of a charge letter to the Committee. Dahlburg quoted the questions that the Committee has been asked to seek answers to:

- What is the overall quality of these facilities relative to the best-in-class in the United States and internationally?
- How do these facilities relate and contribute to Departmental mission needs?
- How might the role of these facilities evolve to serve the missions of the Office of Science over the next three to five years?

She turned the podium over to **Rick Stevens** to speak about the high-performance computing (HPC) facilities. The National Science Foundation (NSF) awarded \$53 million from FY01 through FY03 to several institutions to deploy a production Grid environment. It will have a staged deployment based on service priorities. The first priority is a linked set of working IA-64 based clusters, which will be immediately usable by the current NSF Partnership for Advanced Computing Infrastructure (PACI) user base and will support current high-end applications and provide a platform for building standard cluster and data-management software.

The technology choices are based on the facts that the top 20 PACI users compute on Linux clusters and the majority of applications will be data intensive. This design will map onto the DOE use. Major tasks include building a management structure across multiple sites, engaging major application teams, constructing a high-bandwidth national network, integrating terascale hardware and software, establishing distributed TeraGrid operations, and deploying and hardening grid software (which will require grid testbeds).

He showed the planned layout of the SDSC-NCSA Linux TeraGrid that will use off-the-shelf technology. It will be the fastest research network in the world when it comes up later this year. There will be three levels of middleware: (1) the support of basic services, (2) core grid services, and (3) advanced services. Middleware will use the Globus infrastructure, which was described in an earlier ASCAC meeting. This grid middleware (toolkits for building grids) will provide public-key infrastructure (PKI) based security infrastructure, distributed directory services, reservation services, metascheduling and coscheduling, quality-of-service interfaces, grid policy and brokering services, common I/O and data-transport services, and meta-accounting and allocation services.

Three classes of users are expected to embrace the Grid: existing supercomputer users, new-capability users, and data-intensive and remote-instrument users that will use linked archives, instruments, visualization, and computation.

The scientific-computing community uses three approaches to developing and deploying Grid infrastructure: top down (which usually entails a small number of people and the introduction of hardware), bottom up (usually Web-based computing), and user-community based (which would allow

DOE's widespread facilities to be integrated). Grid strategies that are appropriate for DOE-SC would probably have a three-layered structure of Grid resources:

1. Large-scale Grid power plants (1 to 10 of these),
2. Data and instrument interface servers (about 100), and
3. Principal-investigator- (PI) and small-laboratory-based resources (1,000 to 10,000 of these: workstations and small clusters, laboratory data systems, databases, etc.).

All of these would be tied together in an integrated environment.

Currently, DOE is building a lot of its computing facilities on clusters, but clusters will not get us to petaflops. They are built on Web-server technology, not optimized for scientific computing. Going to 64-bit systems provides good price performance, but the machines are designed for the commercial world.

The limits to cluster-based systems for HPC include memory bandwidth, communications fabric/CPU/memory integration, and system packaging density. If we look inside the PlayStation, the architecture is outside of the scientific computing realm but could get us to petaflops for \$50 million and to teraflop machines for \$50,000. IBM, Sony, and Toshiba are putting \$400 million towards a teraflop processor (1000 times the performance of a PlayStation 2). If the same level of integration was used in clusters, their scalability problems would be solved. The VIRAM-1 (vector intelligent random-access memory) Integrated Processor/Memory being developed at the University of California at Berkeley puts a 256-bit media processor [vector] and 14 Mbytes of dynamic random-access memory (DRAM) on a chip to perform 2.5 to 3.2 billion operations per second. From this, one could construct a very-large-scale (about a 50-teraflop) machine. In the next couple of years, we should be able to determine what an architecture would look like that would exploit these high-integration-design tools. A community effort could result in the development of \$50 million petaflop systems and \$50,000 teraflop systems.

How does this link back to the Grid? Circa 2006 to 2010, DOE computing facilities may have 1 to 10 petaflop computers at the power-plant level, 20 to 100 100-teraflop servers at the data server level, and 1,000 to 10,000 1-teraflop lab systems. The ideal bandwidth of the networking system will probably include 10% of the bisection bandwidth per system for peer-to-peer communication; terabit wide-area-network (WAN) backbones and backplanes will be needed.

In summary, Grid-based computing is well matched to DOE's distributed facilities and missions needs. Grids do not replace the need for large-scale computers. They will increase the high-end demand via portals and increase data-intensive computing and high-performance networking. Software environments will link desktops to high-end platforms (at petaflop rates). However, grids require new ways to allocate and manage computing and data resources, and they need a broader view of resources and resource allocations. Grids and the technology needed for petaflop facilities make sense together; technologies for petaflops will power future grids at all levels.

He offered several recommendations:

- DOE should aggressively pursue development of Grid technologies and the deployment of Grid-based infrastructure.
- DOE should facilitate Grid application communities that are relevant to its mission areas [security, energy, climate, high-energy physics (HEP), etc.].
- DOE should participate in national and international coordination of Grid development and deployment.
- DOE should support development of computing platform technologies that will enable future Grid engines, including new approaches to petaflops and associated affordable teraflops.

Dahlburg introduced **Robert Ryne** to speak on the needs of a large-system user.

The national investment in particle physics is enormous (many billions of dollars) and is used for research in biology, materials science, chemistry, and other applications. The contributions of accelerators have a significant economic impact on and great benefit to society, such as medical isotope production, electron microscopy, accelerator mass spectrometry, medical irradiation therapy, ion implantation, and beam lithography. They have also been proposed to solve major societal problems, such as the transmutation of radioactive waste, accelerator-driven energy production, and hydrodynamic imaging.

The accelerator community uses high-performance computing to tackle a wide range of problems, falling into three areas:

1. Designing complicated electromagnetic structures,
2. Modeling high-intensity beam dynamics, and
3. Exploring beams under extreme conditions.

One might ask, why do we need terascale computing for the next-generation accelerator design? They answer the need for high-accuracy requirements (such as the design of 3-D electromagnetic components) and for large-scale requirements (such as designing 3-D electromagnetic components, modeling 3-D intense beam dynamics, and modeling 3-D advanced accelerator concepts) by including more physics.

Noteworthy computations that have been performed at the National Energy Research Scientific Computing Center (NERSC) include:

- The first successful 3-D eigenmode calculation of the ~50-cell Next Linear Collider (NLC) structure, including the accelerating cells and damping manifold;
- The first self-consistent 3-D Fokker-Planck simulation;

- Simulations that identified the possible heating mechanism in the PEP-II B-factory interaction region;
- Simulations in support of the Stanford Linear Accelerator Center (SLAC) E-157 plasma wakefield accelerator experiment;
- Simulations in support of the Spallation Neutron Source (SNS) linac design effort; and
- Simulations in support of the CERN/SPL (Conseil Européen pour la Recherche Nucléaire/ Superconducting Proton Linac) design effort (similar to the U.S. neutrino factory design effort).

Meza asked what the computation complexity of these problems was. Ryne responded, three dimensions, a number of degrees of freedom in the low to 100 millions, and parallel-particle tracking. Dahlburg asked what size grid was needed. Ryne replied that, for a low-resolution design, a grid of 128^3 is a good starting point; then up to 1000^3 , putting one cell on each machine. We would love to have access to 1000 processors on a regular basis.

The 3-D first-principles Fokker-Planck simulation requires an analog of thousands of space-charge calculations per step. According to the *Journal of Computational Physics* (Vol. 138, 1997), "... it would be completely impractical (in terms of number of particles, computation time, and statistical fluctuations) to actually compute [the Rosenbluth potentials] as multiple integrals." This project demonstrates that these computers make it possible to do calculations previously thought impossible because of both size and speed.

Connolly asked why someone would make such a statement just 4 years ago. Ryne replied that they must not have been aware of these machines' being developed.

NERSC's codes are being used to identify localized modes to understand beam heating in the center beam pipe of the PEP-II IR Beamline Complex. In addition, they are being used not only to design future machines but also to understand fundamental science, such as the investigation of laser/plasma-based acceleration, which can produce gradients of about 100GeV/m.

NERSC provides outstanding support; its user services has been cited as being reliable, timely, and friendly, offering great phone support, excellent web pages, and ease of adding new users. Its hardware is world class, and it offers science-enabling resources include Seaborg, High-Performance Storage System (HPSS), ESnet, and auxiliary systems. It is a very stable environment. Algorithm and software development is available, and it has outstanding support staff for event coordination, training, and publications.

NERSC does an excellent job of serving the needs of both a large user community and a small number of power users/projects. NERSC has made the right choices in hardware. The staff has been very successful at procuring systems that are near the leading edge, yet with the stability and reliability to serve its large community of users. NERSC provides enhanced support to selected projects, such as the SciDAC modeling project to which they provided server support, Concurrent Versioning System (CVS), software requests, and technical support.

What can the HEP community do for the nation and society? Accelerators are crucial to scientific discoveries in high-energy physics, nuclear physics, materials science, and biological science. Design has a huge cost impact in accelerators. In the Superconducting Super Collider (SSC), a 1-cm increase in the aperture of a component because of a lack of confidence in design resulted in a \$1 billion cost increase. (This calculation would have been trivial to perform.) In large-scale HEP simulations, advanced computing can produce cost savings. Improved accelerator design for the Next Linear Collider (NLC) is estimated to have saved \$100 million.

Where are we going with these simulation activities? We are modelling the beams for many facilities, such as the Fermi National Accelerator Laboratory (FNAL) booster, Brookhaven National Laboratory (BNL) Alternating Gradient Synchrotron (AGS), Los Alamos National Laboratory (LANL) Proton Storage Ring (PSR) beam loss, and SLAC E-162. HPC will play a major role in present accelerators, maximizing the investment by optimizing performance, expanding operational envelopes, and increasing availability and reliability. In next-generation accelerators, it will bring about better designs, feasibility studies, facilitate important design decisions, and completion on schedule. In accelerator science and technology in general, it will help develop new methods and explore beams under extreme conditions.

Dahlburg then presented **Dalton Schnack's** presentation for him; it was a mission-driven perspective on the status and needs for DOE computing. He is the project leader for the Non-Ideal MHD with Rotation Open Discussion (NIMROD) Project, a large-scale, multidisciplinary, multi-institutional computational project with the Office of Fusion Energy Science (OFES).

The mission of the OFES program is to advance plasma science, fusion science, and fusion technology to develop the knowledge base needed for an economically and environmentally attractive fusion energy source. Fusion plasmas present an extreme scientific challenge because, although the fundamental laws of physics are known, the collective nonlinear dynamic behavior is very complicated, the dynamics are characterized by extreme separation of space and time scales and extreme anisotropy, and the science problems are as challenging as fluid turbulence. The scientific uncertainties can have a large impact on program costs, and relevant experiments are very expensive. Therefore, realistic simulation with large-scale computing is cost-effective.

Realistic simulation is important because an advanced fusion experiment would cost \$2 to 10 billion, the power density is proportional to the square of the maximum pressure, the uncertainties in nonlinear physics would account for about half the cost of the experiment, and predictive fluid simulation with realistic parameters could provide a high leverage to remove these uncertainties.

Centralized computing started with fusion, a decentralized program. In the 1970s, it was recognized that the required simulation capability could only be obtained with a centralized computer center and high-speed, long-distance networking. Centralized computing eliminated institutional duplication of effort and institutional monopolies, and it facilitated collaborations. The model has subsequently been adopted by other programs and other agencies and has evolved into the present DOE computing environment.

NERSC has produced a world that could not have been imagined. Problems whose solution is essential to the OFES mission, such as turbulent transport and long-time-scale fluid dynamics, require the largest, fastest supercomputers imaginable. The majority of fusion computing problems, however, do not require the most advanced supercomputers. Important fusion problems that do not require supercomputing include calculating the static balance between magnetic and pressure forces in an experimental geometry, determining whether that equilibrium is stable to small displacements, and calculating the diffusion of mass and energy in response to microscopic processes. People can run these latter problems on their PCs.

The important fusion problems are time-dependent and therefore 4-dimensional. They require many thousands of time steps. Often, a large, ill-conditioned linear-algebra problem must be solved at each time step. Parallelization cannot be applied to time because it is prohibited by causality or because the problems are not optimally suited to massive parallelization.

Fusion computing is affected by (1) the ascendancy of massively parallel computer architecture over serial machines and the associated conversion costs, (2) the uncertainty in the configuration of future computer architectures and the anticipated costs of additional future code conversions, (3) the high cost and long timeline of code development by individuals and the general inaccessibility of the resulting codes to the fusion community, and the increasing programmatic importance of physics problems that require extensions of the capability of existing computational tools.

Fusion computing needs to

- lower the future cost of application development, maintenance, and use;
- implement computational and programming methods that isolate architecture dependencies so that future codes will be more efficiently adapted to evolving new architectures;
- make application codes more readily available to, and usable by, the overall fusion science community; and
- develop advanced simulation codes to address physics problems of immediate programmatic interest.

OFES responded to these needs with project-based computing in which code is developed by interdisciplinary multi-institutional teams; is publicly available; and is extensible, modular, portable, and maintainable. What are needed in this new environment are centralized facilities that not only provide high-quality CPU cycles but also facilitate project function.

One example is the NIMROD Project. Its vision is to provide the magnetic fusion research community with a useful and widely available computational tool that can be used to obtain a predictive understanding of the nonlinear dynamics of the most modern and expensive fusion experiments. The NIMROD Project represents a new paradigm for fusion computing because:

- It seeks to solve the nonlinear physical problems facing the fusion program within a user-

friendly, widely available, self-contained structure that allows interaction with commonly available analysis tools and experimental databases.

- It aims not only to provide this advanced computational capability but also to change fundamentally the way that capability is developed.

The typical NIMROD calculation is a time-dependent, nonlinear evolution of a slowly growing, nonideal magnetohydrodynamic (MHD) instability at large (10^6) Reynolds' number. Increasing the number of processing elements (PEs) has little effect because of causality; this performance is only marginally acceptable.

To attack such problems requires large data "farms" for archiving data (many Gbytes/run), fast networks for transferring data between central and local facilities, and mathematical software libraries. It also needs (1) the largest, fastest systems available; (2) optimized performance for long runs (which in turn requires specialized systems for time-dependent problems, longer residence time for jobs, and the need to revisit the massively parallel paradigm); (3) maximized memory, and (4) the offloading of smaller jobs to midrange systems.

Most fusion computing capacity does not require high-end computing, but there is a lack of generally available midrange computing systems because they are too expensive or specialized and they need someone expert in their operation. As a result, these jobs are clogging the queues on present DOE supercomputers

What are needed are specialized architectures optimized for time-dependent problems requiring many time steps; the limitation of access to the system for only the largest, programmatically important problems; a centrally available midrange computing facility for offloading the majority of fusion applications that does not require the most highly advanced computing capabilities, eliminating institutional monopolies; and storage, networking, communication, and graphics to facilitate code and data comparisons and to enable remote collaborations.

Messina asked what the investment in hardware vs infrastructure is in the NIMROD system. Eckstrand replied that, as time progresses, the percentage changes. To begin with, the infrastructure costs were about 50% of the project. McRae commented that the infrastructure seems common to a number of areas (biology, chemistry, etc.) and asked if they should be standardized. Eckstrand acknowledged that, although they are quite specialized applications, the techniques and infrastructure could be made common. Messina commented that "specialized" means not off-the-shelf.

Ryne said that, given the great demand for the highest-end machines, data-parallel jobs could run just as well on something else and should not be clogging up the usage. Dahlburg replied that that is one of the big questions to be answered.

Connolly noted that the NSF has a hierarchy of machines and asked if DOE had that type hierarchy. Dahlburg responded that finding the answer to that question is the main job of the Subcommittee.

Messina replied that DOE really does not have that type of hierarchy.

A break for lunch was declared at 12:20 p.m. The meeting was called back into session at 1:55 p.m. Dahlburg introduced **Stephen Wolff** to speak about networks, specifically ESnet, in the assessment of facilities.

A review of the ESnet program was held in September. At first blush, ESnet is a small ISP. New applications could add 1 GB/sec or 300 TB/month. In February 2001, it carried 45 TB. It is experiencing a 100% growth each year. It is governed by and is extraordinarily responsive to its users. The danger is that the net has to meet demand even if the governance structure is not well constituted to deal with rapid growth. That governance is risk averse and not well constituted for a strategic view. Rather, it supports current connectivity stability (the status quo). On the plus side, it has relieved the user sites of the need to provide networking expertise and user services. Other things on the horizon: The community will boost its usage and there will be new users (SciDAC and biotechnology). In terms of performance, ESnet's connectivity is adequate for most current users. It has good management tools and user services; it is a lean, cost-effective operation.

ESnet possesses no central capability for networking research. It needs to take on research tasks. The establishment of the ESnet Research Support Committee (ESRSC) is a good step, but it needs wider scope. Currently, ESnet is hard to defend as a commodity ISP.

Industry trends are informative: Of the 39 million miles of fiber installed in the continental United States, only 20 to 35% of it is lighted, and only 2% is in use (carrying traffic). The principal costs involved in installation are trenching and terminating. Today, the economic parameters are unclear because of the fiber glut. Because fiber is now so cheap, people are willing to put money into optics for getting the most out of using that fiber. On the technical side, all national networks are mandating IPN6 as the mandatory protocol. Collaboration and the adoption of a Grid paradigm are emerging as drivers. One approach (Canada's CA*net 4) is predicated on a commodity market in lambdas and a surplus of fiber. It postulates a transition from the network as a set of services to a set of owned objects, or "object-oriented networking." This structure produces end-user control, a full mesh among administrative domains, and the avoidance of a backbone network. In such a system, links are owned or leased. An example is the Distributed Terascale Facility (DTF) or Teragrid, currently under development with \$53 million in NSF funding. It is a backplane first, and a network second. DTF is a neat project; because of its clientele and source of technology, it could have been a DOE project rather than an NSF project. There are many other things it could be doing but is not, and therefore it misses the mark.

Connolly asked if anyone proposed replacing ESnet with a commercial ISP. Wolff replied, no; the Subcommittee considered that it was better to point out the risks, consequences, and setting.

Bill Kramer (NERSC) asked about the financial structure. Wolff pointed out that ESnet is very lean. Stevens commented on DTF. Of \$53 million, \$4 million goes to the network, buying or leasing lambdas from Quest; all the network management is done by the network. Network staff is provided at four sites,

usually one person per site. They are also involved in software development.

Wright asked what the implications for DOE would be if ESnet was a "commodity." Wolff said that it would require more people, more money, and more capabilities. This is going on in the Netherlands, with Abilene, and in Canada. Sollins said that, if someone is trying to do research on a network that someone else is trying to use as a commodity, neither one will be happy. To do research on user statistics, however, you need users out there. DOE should be doing networking research, but at least pieces of that research should not be carried out on the ESnet core.

Meza asked if the Subcommittee considered ESnet's providing all the network management a drawback. Wolff replied that they considered it a strength of ESnet.

Messina asked when the world could expect to see an all-optics network. Wolff replied, in Japan in 2003. Messina went on to ask if the DOE community would like Esnet to be a good place to deploy and evaluate all-optical network hardware. Wolff said yes. Giles asked if mixed (commodity plus research) solutions are possible. Wolff replied, yes and that ESnet has been very good at that. A lot of that type of activity goes on, mostly in-house.

Jill Dahlburg summarized where the Subcommittee was in relation to answering the three questions put to the Committee in the charge letter by presenting the outline of the Subcommittee's report as it currently stands. The draft is to be completed by December 10, and the final report is to be delivered to Decker in January.

The Subcommittee started by developing a list of facility attributes in terms of raw infrastructure, quality of service, adaptability, and strategic directions (technology assessment). Giles if this was done in terms of a snapshot or calculated as rates of change. Dahlburg replied that a static assessment was made, and then a historical look was taken at the rates of change.

There are two strategic discussion points:

- Mission-directed research (from basic to highly applied) is the orientation of the Office of Advanced Scientific Computing (OASCR).
- High-end computing is the unique charge of OASCR.

It must be recognized that mission needs require computing from a small number of processors to the highest-end machines. The whole issue of allocating resources is large and complex. Ideally, high-end jobs should be run on high-end machines, and the high-end machines should be running at 90% capacity. The Subcommittee is proposing that individual applications ought to be using a significant capability of a machine for a significant part of the time.

Giles commented that that is good if you think of capacity as a one-dimensional attribute, but machines might have six dimensions, so capacity use might be spread out across those dimensions. Meza pointed

out that one could run 1000 jobs and "be using" a good deal of the machine's capacity. Dahlburg stated that the Subcommittee's definition is *one* job using the resources. A lot follows logically if you have a performance metric, and the Subcommittee suggests that system capacity (the number of processors, memory, bandwidth, or all three) should be the primary metric. This idea arises from costs of, for example, the bisectional bandwidth. If a system-capacity metric is suggested, one must think about a clear line of authority to make sure the metric is met. One will need allocation procedures and strategies, and the allocation process needs to be thought through globally and integrated across facilities. Which brings up the question, should there be an OASCR Allocations Committee? If OASCR is to get where it wants to go, allocation procedures and strategies will be needed. There is a need to allocate in context, against the resources available and also against the mission (i.e., work that must be performed).

An allocation process brings up several additional questions. Should a range of facilities be encouraged by the office? (The Subcommittee uses the generic term "focus centers" in its report.) Should the focus be on systems' interests, or should they be metric driven? The Subcommittee suggests considering focus centers based on computational similarity. If there are centralized allocations, most users will be geographically remote from the facility; evolutionary changes in the Internet might then entail the development of greater bandwidth and/or middleware.

The end result is a network like a nervous system, connecting mission applications (i.e., computational resources, archival data, and experimental facilities).

An intriguing question is whether we can use the Grid to make all the users on the network virtually present in a collaboration or at an experimental run. Messina pointed out that this would play well because it means getting more science done with fewer expenditures.

The Subcommittee's working recommendations are that strategic guidelines should include (1) mission-directed research (from basic to highly applied) as the orientation of OASCR, (2) high-end computing as the unique charge of OASCR, and (3) the need for strategic system-capacity performance metrics. These guidelines lead to a number of particulars: Should the problem be approached in terms of a balancing of the number of processors against mission goals? What role should quantitative statistics play? And what role should philosophy (selecting the questions to address and determining what issues can be addressed from a common framework) play? A general point to consider is how to harmonize all the elements in program planning: mission orientation, high-performance mandate, system-capacity metrics, and peer review.

Stevens asked if the Subcommittee considered how the supercomputing research facilities fit in. Dahlburg replied that they had been lumped into the generic "focus center."

McRae asked if the Subcommittee was identifying anything that is broken that needs to be fixed. Dahlburg said that nothing is broken. The Subcommittee is looking at options for the next 5 years.

A break was declared at 3:05 p.m. The meeting was called back into session at 3:37 p.m. Dahlburg

called for comments from the Committee on the Subcommittee report. Giles asked where the software came from. Stevens said that, for grids, much of the development is focussed on lower middleware and has already been produced with DOE, NSF, and Advanced Research Projects Agency (ARPA) money. SciDAC is developing toolkits.

Meza asked what the bottlenecks are for using the Grid and which types of applications will benefit and which will not. Stevens said that there are three classes of bottlenecks: (1) getting peak performance from current system design, where the bottleneck is in the memory subsystem; (2) the system and network itself, where the bottleneck is getting data between the two (what kind of traffic will be faced is not known); and (3) end-to-end performance because of software compatibilities. Each has to be addressed in a different way. The class of application that will benefit most is the group of data-intensive applications in communities that have already solved their model problem. Less clear is how traditional supercomputer users will benefit from the Grid.

Sollins asked if there is a value to DOE's doing things their way rather than the facilities' acting individually, spanning multiple networks, pointing out that there is no sense in duplicating grids and building networks. Stevens replied that, in HEP, the groups that got together had a large vision and their supporting agencies could do the heavy lifting.

Sollins commented that intellectual sharing (e.g., how to make large models) will happen more easily in shared facilities. Dahlburg recalled two strategic points mentioned in the presentation: high-end goal and mission-oriented, and pointed out that one characteristic common in shared facilities is that you have to wait in a queue. Sollins acknowledged that that is a real problem, say, for weather forecasters for whom the delivery of results is time-critical. The Committee should be talking about when sharing makes sense and when it does not. Stevens observed that, when you cross agency boundries, such sharing becomes difficult. Looking at the Internet, success and progress came when all the agencies adopted the same protocols. We do not have anything like that for the Grid. We need that type of cooperation before we can talk about sharing.

Giles said that the Committee needs to think about rates of change. A metric only makes sense if a facility's lifetime is longer than the careers of the facility's users. Someone has to account for how the use is growing. In real life, usage could drop off even as capabilities increase. In one way, the best metrics would be the percentage of mission accomplished per dollar expended.

Horst Simon (Division Director, NERSC) observed that just focussing on flops or something like that misses the mark; you have to look at the facility. The DOE science grid was not listed as a facility, but it will be an important element for the next several years. The big question is, given the NSF investment in the DTF, can a plan be developed to leverage the NSF investment to the benefit of DOE.

Karlsrude asked if "management structure" meant anything more than "people." Stevens replied that, in building a grid or network, you are touching the edges of institutions and you have problems of control, allocation of resources, funding, handling proposals, etc. You have to find common ground, which

produces friction; therefore, you need leadership. With a grid, you allocate resources from one institution to another. Slight changes in policy can make the simplest operation very difficult. Karlsruhe followed up by asking what unique hardware and software innovations would be needed for the Grid. Stevens pointed out that the Grid combines resources regardless of location. Every scientist can see a huge menu of capabilities that can be used to build powerful applications. The sophistication of the software needed to do this is going to be immense. Today, the lower-level protocols are being developed. Some of the upper-level SciDAC collaborative pods are trying to build prototype applications that sit on top of that general infrastructure. What is not happening is a broad attack on all DOE science missions. What can be imagined is an array of innovative software that would take advantage of advanced hardware, a lot of which does not exist; and what is not known is how much of that hardware and software will be needed to make the Grid usable. Sollins noted that the Grid must also gracefully incorporate new capabilities as they are developed. Messina added that there are bandwidth problems that must be overcome, also.

Wright noted that Congress is interested in why we have so many more facilities and such low peak usage, asking why we do not stop buying new hardware and put money into some more efficient software. Giles said that that goes back to the method of measurement. Flops can be counted easily, but that is not necessarily a good measure of a facility. Computing facilities should be looked upon as a replaceable commodity. The cost of hardware is not trivial, but to buy more may be more cost-effective than developing software. Dahlburg asked if what he was saying was, "use what you have to get your answers and do not worry about percentage of peak usage." Giles responded, yes, but we should consider the best way to reach those answers. Often the most cost-effective way is to buy a machine that is twice as fast than to invest in software that is twice as efficient.

McRae said that his personal metric is minimizing the time between identifying a problem and solving it.

Messina commented that all the available machines look alike; one is not much better than the others in efficiency.

McRae stated that he spends more of his time trying to map a problem onto an architecture than on running the problem on the computer.

Messina observed that the people at DOE's facilities are treasures that should be retained.

Ryne went back to Wright's comment about the question from Congress and said that the answer is that you are going to get the most science out of it. Dahlburg said that it might be better to leave a system 50% unused because that means it is being used for what it is intended. Wright noted that, several years ago, a General Accounting Office (GAO) report seized on the efficiency of the DOE computers. The Committee needs to make clear to such critics that that is not a meaningful method of assessing resource usage. The Subcommittee should come up with some way of making that clear. Stevens suggested the phone system as a good analogy: you always get an instantaneous dial tone because the system goes

unused a good deal of the time. Sollins cautioned that a simple modification of the model (the introduction of modems) brought the model down and it had to be radically changed.

Connolly asked if anyone was working on controlling network resources. Wolff responded, yes, but it is not a technical problem, it is an administration problem.

Roscoe asked about batch processing for visualization, etc. Ryne said that it varies from user to user. Some users can wait and receive the results from a job run whenever they get it. For others (e.g., weather forecasters), the information is time-critical, and they need their data back quickly. Either they pay more or something is done to give them priority. Recently, one community (climate researchers) wanted priority, and the NERSC staff worked it out. NERSC is now looking at offering so many node hours to particular communities.

Connolly asked what happens at machine saturation. Ryne responded that NERSC has a committee that deals with allocation. It has only dealt with one community so far; it may be different when the process is extended to five communities. Kramer stated that it is possible to create flexible service models. The Ultrascale Evaluation Report issued by Defense Progeams (DP), NSF, and the Department of Defense (DoD) covers a lot of the ground that is being addressed here.

Zaharia noted that the reason we have these facilities is to drive science. If we had 100% efficiency, we would just be a big ISP. A refresh of a facility's equipment enables new science that otherwise would have been unattainable. Dahlburg asked why, if you get new breakthroughs with new facilities, you do not you get new breakthroughs with doubled efficiency. Andy White (LANL) said that there is an odd, symbiotic relationship with improving codes resulting from new machines. Zaharia said that creating new facillities also brings together groups of people to work collectively on societal problems, creating new algorithms to address problems that would not be addressed otherwise.

Kramer pointed out that one has to be careful what one wants to measure. Not all facilities have to have the same metric.

George Seweryniak said that it was his understanding that ESnet could be done cheaper on the outside. Wolff replied that he did not know that that could be done; at present it is being done as cheaply as possible.

Wright introduced **Gregory McRae** to comment on the Overarching Issues Subcommittee's initial framework for discussion. The initial charge to that Subcommittee covered facilities and comparative biology, and its initial mission was to look at the advanced scientific computing of SC. The Subcommittee looked at the budgets of the different divisions, including the Office of Biological and Environmental Research (OBER):

Office	Budget (millions of dollars)	Computing expenditure (millions of dollars)	%
OASCR	300	160	53
BES	1000	10	1
OBER	450	30	7
OFES	250	10	4
HENP	1000	20	2
Total	3000	230	8

A few observations can be drawn from these numbers:

- HPC is a critical enabling technology for all DOE missions.
- Computing is more than hardware; it must include a problem-solving environment.

Operating systems

Algorithms and software

Networks and communication

Analysis, interpretation, and storage of data

Data assimilation and control of experiments

etc.

This Committee should gather statistics on where the money is going. There is a huge and growing diversity of applications in many areas with many commonalities; those commonalities could be exploited. A critical role is user support in the migration to a new computational-science paradigm.

Questions that came up in the course of the Subcommittee's discussions included:

- Where is the high-level vision for the role of computing within the agency?
- How does one make most effective use of the diverse skills and experience of DOE in emerging

missions (e.g., Genomes to Life and missions to nanoland)?

- DOE does a terrible job of demonstrating the value of the investment in computing to Congress, to the public, to other scientists, etc.

Many issues require going beyond OASCR. For example, in discussions with Decker, it was decided that ASCAC should be talking with other advisory committees of DOE and should be looking at the needs of five offices in SC, specifically the HPC needs, management issues, and funding requirements.

Reports that are due include one from this Subcommittee on Jan. 11, 2002, and one from the Technology Composite Committee, which is due Sept. 1, 2002.

McRae made a specific proposal to form a Subcommittee to (1) write a brief section for the January report that identifies big-picture issues and (2) begin to frame the issues and agenda for the meeting with the other advisory committees.

In preparation to begin these tasks, ASCAC needs to (1) discuss the identity of the big-picture questions; (2) prioritize those issues; (3) figure out and define what the Subcommittee's agenda, charter, objectives, and deliverables would be; and (4) identify some volunteers to help in this process.

Wright underscored Undersecretary Card's request for help in stating to Congress and the public the need for and value of high-level computing. Meza responded that a big issue might be the 8% of the budgets of five offices of SC that is devoted to HPC. Connolly stated that we should also point out that a good deal of investment supports other agencies. Meza asked how OASCR was to sustain this effort with its low current funding. Messina said that the cited budget figures were estimates, not real numbers. McRae stated that the Subcommittee needed hard data for these statistics. Coronos noted that the big step was the establishment of the High-Performance Computing Centers (HPCCs). McRae said that what needs to be done is to put these numbers together and compare those numbers with industry's and see if DOE is even close in the level of reasonable investment. Messina noted that the Jet Propulsion Laboratory (JPL) once measured the percentage of their budget that went to computing. The answer was 20%, and little went toward hardware. Sollins noted that computing is coming to the point where storing and managing data is very important, and paying for those services gets very short shrift.

McRae called for volunteers to be on the Subcommittee. He asked if the DOE people had any input to the Subcommittee's discussions. Dan Hitchcock said that what was missing from the discussions was people and training; the aging population in the DOE research community needs to be replaced. He also commented that a lot of data curatorship is not thought of as computing.

Wright called for public comment. There being none, she adjourned the meeting for the day at 5:22 p.m.

Friday, October 26, 2001

Margaret Wright called the meeting to order at 8:47 a.m. with administrative announcements and

introduced **Edward Oliver**, Associate Director of OASCR. He displayed the organization chart of the Office, which showed the three divisions: Mathematical, Information, and Computational Sciences; Technology Research; and Office of Scientific and Technical Information. (Another member may be added to ASCAC to reflect this last activity.) In addition, there is the Corporate R&D Portfolio Management Environment (PME) Project, which may soon become a division. Each of these activities touches all parts of SC.

The office's budget is essentially flat, with \$161.296 million appropriated in FY 2001 and \$163.050 million requested for FY 2002. The FY 2001 number excludes \$3.9 million that was transferred to the Small Business Innovative Research (SBIR) program and \$239 thousand that was transferred to the Small Business Technology Transfer (STTR) program. That money is split out as follows: facilities 37%, SciDAC 24% (a major increase, all in research and intellectual activity), and base research 34%; 2% is devoted to joint activities with BER, and 3% goes to SBIR/STTR.

He reviewed the personnel situation and the staff responsibilities, pointing out that many of the staff have responsibility in more than one area.

Accomplishments in FY 2001 include initiating the SciDAC research program, upgrading NERSC to 5 teraflops, initiating \$3 million in research projects on computational biology with OBER, and acquiring an IBM Power 4 (4 teraflops) for evaluation/scaling studies at Oak Ridge National Laboratory (ORNL). The plans for FY 2002 include ensuring the success of SciDAC, making an effort to strengthen the core basic research efforts, and initiating the procurement cycle for NERSC-4.

The Office also wants to initiate an early career researcher program, the goal of which would be to include exceptionally talented "early career" researchers in the Mathematical, Information, and Computational Sciences (MICS) Division base research program. This program would be a new base program element for FY 2002. The Office plans to issue requests for proposals in applied mathematics, computer science, and network research that will target individuals in tenure-track regular faculty positions at U.S. academic institutions (within 5 years after the completion of a PhD or postdoc position). The program would be phased in over 2 to 3 years, with about 30 awards to be made in FY 2002, each at \$100K/year for 3 years (with 10 new ones each year). Extra consideration may be given to applications in which part of the research is conducted at a DOE national laboratory (e.g., during the summer months).

Sollins suggested sending graduate students to the laboratories. Wright said that an early career program is fabulous because it gets new researchers involved and asked what the FY 2003 request looked like. Oliver replied that that was embargoed information. McRae commented that, in addition to young investigators, perhaps there should be an avenue for those making a midcareer change. Meza suggested that perhaps we should be encouraging joint grants with BER or FES for investigators going into cross-disciplinary studies.

Oliver then introduced **Stephen Eckstrand** to speak about the status of SciDAC and how it got there.

Eckstrand noted that, at the previous ASCAC meeting, the peer review process to award grants was just starting. That review process has now been finished, and research projects are under way.

The goal of SciDAC is to create an integrated program to

- Produce a new generation of scientific simulation codes that take full advantage of the extraordinary computing capabilities of terascale computers,
- Develop the mathematical and computing systems software to enable these scientific simulation codes to use computers effectively and efficiently, and
- Provide a collaborative software environment to enable geographically separated scientists to work together.

The SciDAC vision is that major increases in computing power will enable a new generation of simulation codes, based on reliable experimental and theoretical inputs, to lead the way to greatly increased scientific understanding, which in turn will lead to new theoretical and experimental discoveries.

Three program-management strategies are being considered:

- Coordinate planning and management of SciDAC activities across the Office of Science (through the efforts of the SciDAC director and a SciDAC Coordinating Committee),
- Establish integrated scientific challenge teams (applications scientists, applied mathematicians, and computer scientists) with clear deliverables, and
- Manage the resources wisely.

He reviewed the personnel on the SciDAC Coordinating Committee and then the program's history:

Develop SciDAC Strategic Plan	3/2000
Prepare budget requests	1999-2000
Prepare solicitations (6)	12/2000
Conduct merit reviews	4-5/2001
Make funding recommendations	6/2001
Prepare grant and contract awards	6-7/2001
Manage start of individual projects	7-9/2001

Prepare for PI meeting	10-12/2001 (cancelled because of 9/11)
------------------------	--

The initial awards focus on software: 24 projects funded at \$500,000 to \$3 million per year. For the first year, of the \$57 million total SciDAC funding, 55% was distributed to national laboratories, and 45% to nonlabs. Of the \$37 million MICS SciDAC funding, 53% was distributed to national laboratories, and 47% to nonlabs. After the review, 13 national laboratories were involved, many involved in six or seven projects, and 53 universities in 26 states and the Province of Quebec were involved, some in as many as seven projects. He reviewed the awardees and topics of the successful proposals and identified the four Computer Science Integrated Infrastructure Centers funded under the program: Scalable Systems Software for Terascale Computers, High-End Computer Systems Performance, Center for Component Technology for Terascale Simulation Software, and Scientific Data Management Enabling Technology Center. He said that the best news in the briefing was the high number of interactions between SciDAC and projects in other offices of DOE, which were shown in a table.

Wright asked if the money was going to people who were new workers in the field or to people already working in the field. Eckstrand said that it is a mix. Some projects are entirely new personnel. At the national laboratories, many of the proposers are already there, but this program is allowing them to hire new people. Wright asked how long this money is going to last. Eckstrand replied that SciDAC is not in danger of losing its funding; it involves half the offices of SC and would love to add more.

Meza asked what the metrics for SciDAC are and who will evaluate them. Eckstrand responded that each program office is defending its involvement in SciDAC separately; overarching concerns have not yet been looked for.

Sollins asked what the relationship was between what was asked for and what was awarded. Eckstrand said that people always ask for more money than they need; many projects were funded at 80 to 90% of what they asked for. Sollins noted that the networking proposals funded were for middleware and stated that funding was needed for models and computation. Eckstrand acknowledged that not a lot of money was put into software and the Grid and that needs to be fixed. Sollins commented that, if such proposals are not requested, they will not come.

Dahlburg asked if funding at 80 to 90% of the request was still considered very lean. Eckstrand responded that many of the programs are 2 and 3 times what SciDAC is providing in funding.

Wolff asked what the impacts of these projects are on ESnet. Eckstrand said that there was no funding for network usage, and the staff worried about that. The expectation is that the awardees would not be using the network much in the first year.

David Bader introduced himself as the Acting Assistant Director of Scientific Simulation and spoke on

the topic of scientific discovery through advanced computing. During the past 12 months, SciDAC has identified and selected building blocks for the program. During the next 12 months, it will adjust its scope to budget realities while retaining the original vision, fill in missing pieces, integrate the building blocks into a comprehensive program, and establish performance criteria to track progress. The rapid ramp-up originally conceived will not occur.

The focus of the program is an enabling of science. The original goals and strategies are to promote scientific discovery through advanced computing. A hardware infrastructure will be put in place by OASCR, and a software infrastructure will be developed by OASCR in cooperation with BES, BER, FES, and HENP, which in turn will produce the science. An example of such an approach is Northwest Chem, which was described at a previous ASCAC meeting. A major difference is that Northwest Chem was limited to one institution; SciDAC needs to involve multiple institutions. As in Northwest Chem, SciDAC's development teams support the full software life cycle, including deployment and use.

The hardware infrastructure includes:

- A flagship computing facility (NERSC),
- Topical computing facilities to provide the most effective and efficient computing resources for a selected set of scientific applications and to serve as a focal point for a scientific research community as it adapts to new computing technologies, and
- Experimental computing facilities to assess new computing technologies for scientific applications

The main question is, how can we meet these goals when there is not a lot of money for hardware?

The computational facilities consist of (1) ORNL's Center for Computational Sciences (CCS) as the primary SciDAC resource for testing concepts for topical facilities and for an experimental facility and (2) 20% of NERSC's resources, providing "red carpet" treatment for SciDAC users and flexibility in resource usage. One might ask why experimental facilities are needed. That is because vendors' core capabilities are driven by markets and are not adapted to high-end scientific computing. An organized approach is needed for evaluating new computing technologies and for interacting with computer designers as early as possible. One might also ask why topical facilities are needed. This topic is controversial, and he did not know why except that this model has been tried before and has succeeded.

As metrics are looked at, one must remember that the infrastructure must support the end scientific goals; SciDAC must be more than just some incremental improvements on "business as usual"; measurable, quantitative metrics must be able to be developed to track progress; and the success ultimately depends on project leaders' and program managers' sharing responsibility for effective integration.

Connolly asked if he saw biotechnology and materials coming into the program. Bader said, biotech yes, materials no. Connolly asked if topical computing is considered by OASCAR to be important for the

future. Oliver responded affirmatively.

Lester asked how he saw the metrics being set up. Bader replied that he had been in the position only 2 days and did not know the process. Eckstrand said that everyone was in agreement that SciDAC has to be more than an incremental advance; the DOE staff is working with the PIs to establish milestones and deliverables for their projects. Giles asked if the milestones reflect "not business as usual" in a measurable way. Bader replied that they had a detailed task list that must be met. Sollins asked how detailed these lists are. Bader responded that they are not month by month, but the PIs have to meet identified, coordinated accomplishments. Connolly asked if the scientific results are the deliverables. Bader responded, yes.

Meza asked if Headquarters is also going to do things differently and if they had enough staff to oversee all these changes. Bader said that the HQ activities are working together very well. Decker has said he is committed to making this work. Bader is hoping that this can be proven as a model for success.

Wolff noted that SciDAC will bring new users and resources to ESnet and asked how many SciDAC representatives are on the ESnet Steering Committee. Oliver responded, none. Wolff asked, out of the SciDAC funds, how much has been allocated to ESnet. Oliver replied, none, but it is a high priority for 2002. Wolff stated that they could guess how many network services those researchers are going to get. Wright suggested that Wolff and Sollins write up a suggestion, and Bader stated that such funding should be hard and fast.

A break was declared at 10:12 a.m. The meeting was called back into session at 10: a.m. with the introduction of **Juan Meza** to speak about the ASCAC Biotechnology Subcommittee. He reviewed the personnel on the Subcommittee and the charter received from James Decker. The Subcommittee met to focus the charge:

- What are the grand challenges in computing/systems biology?
- Which areas have the greatest overlap with OASCR and DOE?
- What are the open mathematics and computer science problems in these areas?

The Subcommittee's goal is to determine if there is a model for bringing new people into these fields. He summarized the Computational Biology Workshop for Genomes to Life (the Lander workshop), the Computation and Systems Biology Workshop, a Subcommittee meeting, and numerous other discussions.

The Lander meeting was exciting and identified rapidly developing fields where computer science and mathematics could be most useful. It is important to address "social issues"; one cannot parachute into another's field. One must ask if biologists are receptive to offers of help and to developing new researchers at the interface of the two disciplines.

Topics the Subcommittee must address are the need for a workshop addressing higher-level and long-

term research issues, coordination with the Genomes to Life workshop, coordination with other DOE offices, and coordination with other federal agencies.

He introduced **Gary Johnson** to discuss computational biology. Systems biology is a systems-analysis and engineering approach to biology to understand the workings of entire biological systems. It requires the integrated application of methods from modern biology, computational science, and information science and technology and requires advanced measurement and analytical technologies.

Systems biology provides biological solutions to DOE problems through an understanding of biological systems from the genome to the proteome to the cell and organism and microbial communities. DOE is the only agency that can integrate the physical, computational, and biological science expertise at a large scale and scope. DOE is in this because of its interest in waste cleanup and disposal, worker health and safety, and technologies for detecting and responding to biological terrorism. In the mid to long term, the program is expected to help make the country less dependent on foreign oil and to advance the practice of carbon sequestration, all linked to the DOE mission.

Specific research activities include a joint OBER-OASCR program on Genome to Life, a joint OASCR-OBER project on Advanced Modeling and Simulation of Biological Systems, and the OBER Microbial Cell Project.

The Genomes to Life Scientific Plan calls for an understanding of how genes, proteins, and cells work in intricate networks to form dynamic living systems, exquisitely responsive to their environments. It involves computational, theoretical, and experimental activities and collaboration. A systems approach needs to be taken to very complex questions. How to collaborate with industry must also be figured out.

The solicitation in Advanced Modeling and Simulation of Biological Systems stated that the goal of this program is to enable the use of terascale computers to explore fundamental biological processes and to predict the behavior of a broad range of protein interactions and molecular pathways in prokaryotic microbes of importance to DOE.

Nineteen proposals were received in the areas of protein folding and docking and cell modeling. Ten awards were made; first-year awards totaled about \$3 million; funding was at 50% of the request. The funds were split about 50/50 between the national laboratories and universities.

The solicitation for the Microbial Cell Project stated that the project is focused on fundamental research to understand those reactions, pathways, and regulatory networks that are involved in environmental processes of relevance to the DOE, specifically the bioremediation of metals and radionuclides; cellulose degradation; carbon sequestration; and the production, conversion, or conservation of energy.

A table of agency funding levels of computational or systems biology indicated that DOE has set some very challenging objectives for itself:

NIH	\$50 to 100 million
NSF	\$48 million
DARPA	\$15 to 18 million
DOE/OBER	\$9 million
DOE/OASCR	\$3 million
USDA	\$3 million

Connolly asked if there had been any update on the NSF Computational Biology Centers. Kramer said that one was planned, but no budget has been allocated. They have a lot of program announcements on computational biology that will be tied together later.

Genomes to Life held a workshop in August 2001 and one on Visions for the Future in September 2001. More detailed planning workshops are to be held:

- January 2002: Computational Infrastructure for the Genomes to Life Program
- February 2002: Computer Science for the GTL Program
- March 2002: Mathematics for the GTL Program

Research opportunities in computational biology include managing data plus data as a tool for scientific discovery; molecular dynamics, protein folding, and docking; pathways and regulatory networks; and developing needed data sets against which to validate the models. Biology is going from being a data poor and qualitative science to being a data rich quantitative and predictive science. The irony is that a lot of people went into biology because they did not like mathematics and computer science. *PubMed* citations that include "simulation" or "modeling" in the title or abstract show a log-scale increase during the past 6 years.

The capabilities needed to be a leader in the emerging field of systems biology include a strong experimental biology program, theory and simulation, and high-performance computing. What are needed are a plan for an R&D agenda with components in mathematics and statistics, computer science, informatics, and hardware and networking infrastructure. It should be focused on DOE mission opportunities to use biological data to enable scientific discovery, to determine the structural details of biological "parts," and to model whole cells and microbial communities.

Connolly noted that postdocs may be a good mechanism for getting people to move into the field. Johnson said that that is a good idea. There is a good-sized pool of computational scientists out there that would like to move over into biology. They do not need a complete biology education; mathematicians and biologists can meet somewhere in the middle.

Karlsrud observed that data mining is another important field and asked Johnson if he had talked with that type of person. He replied that they had started doing that.

Meza pointed out that the problem of what to do when errors are found in databases has to be dealt with. Corona observed that there is a lot of energy out there; young researchers are really interested in computational biology. Johnson stated that it is the killer application, but where the money is to come from is yet to be decided.

Dahlburg asked what the Subcommittee is planning after February 2002. Johnson said that he would like to see what happens in the earlier workshops and keep the options open. The three workshops between now and February are designed to look at the details of what were identified as needs in August. He would also like to interact with industry more and to work in concert with them. Giles noted that, on the academic side, there are a lot of programs in *biomathematics* and also pointed out that the industry question is complex because of protection of intellectual property. Johnson said that that DOE wants its research expenditures to benefit the public; how to work with the private sector needs to be figured out.

Meza asked how to get OASCR and OBER to work together. Houghton said that one way is to have solicitations that are broad enough to bring in people from both offices. Thomasson commented that OBER has the same question; it needs OASCR as much as OASCR needs OBER. Johnson commented that the people work together very well once they overcome the institutional barriers.

Wright asked whether the concern with bioterrorism might alter the budget priorities in the future. Oliver said that that is conceivable. Thomasson responded that many of the programs in several offices will be affected by that concern. Dahlburg commended that it would be a good thing for a white paper to come out of these offices to address that issue and how it is going to be important for the next 10 to 15 years. There is a rumor that basic science funding will be cut by 5% next year because of the exigencies of funding counterterrorism efforts. It seems that science funding should be *increased* for the same reason. Meza said that his Subcommittee had been trying to identify people with a broad strategic viewpoint (such as representatives from the Institute for Mathematics and Its Applications and from the Mathematical Sciences Research Institute) and was looking at a February or March workshop to address that question. They would welcome ideas from anyone else.

Giles said that he felt intellectually challenged in understanding the term, "complex biological systems" and asked if someone would shed some light on what that refers to. Researchers and computer modelers generally look at specialized areas. Modeling a broad scope of biology and information is an overwhelming concept. He asked if the Committee had something special to say about that or could discuss it at a workshop. Michael Colvin (LLNL) asked rhetorically what it means to understand a cell. A lot of creative thought needs to go into the interface between biology and computation. Wright stated that many of these issues did not get well discussed at the Lander workshop and could have been and should have been discussed.

She opened the floor to a discussion of the issues brought up by the Committee at this meeting and asked **Gregory McRae** to start that discussion. He summarized some initial ideas that the Subcommittee has put forward:

1. Computing is vital to meeting the DOE mission, but there is no high-level vision for the role of computing at the agency. The Subcommittee recommends the creation of an agency-wide vision statement, action plan, and management commitment to its implementation at the highest level within DOE.
2. DOE is invisible in its contributions to computing. The Subcommittee recommends the identification, quantification, and publicizing of areas where DOE's investments in computing have led to major impacts. There should be a concerted and visible effort to demonstrate value to the nation (e.g., genome sequencing).
3. Computation must be seen as more than just hardware; it includes software, communications, data storage, etc. The Subcommittee recommends recognizing and reflecting a balance in investments made in underlying enabling technologies that facilitate meeting the science missions of the agency.
4. Effective management and implementation of DOE SC programs involving computing is critically hampered by lack of personnel. The Subcommittee recommends identifying and hiring experienced program-management personnel.
5. Computing is a vital aspect of many exciting and challenging problem areas, such as the stability of the national power grid. The Subcommittee recommends identifying and understanding how DOE should participate in these activities either as a lead agency or as a supporting agency.

He outlined the additional inputs needed by the Subcommittee.

Sollins stated that the Committee must distinguish between infrastructure and computational resources provided for others and must separate those agendas and their funding. Giles and Sollins said that metrics are needed for the Office of Management and Budget and other audiences. The Subcommittee also needs to capture the value of research. It needs to determine how the need for metrics may affect the risk/reward balance and how it may have a negative effect on long-term research that entails risk. The Subcommittee will report in May what has actually happened in the past and maybe have a small workshop. It will forge links to the subcommittee activities of Dahlburg and McRae. It will look at research vs artifacts, proposal skews, and a redefinition of what is "success." An important point is that producing negative results is not valued in computing.

Dahlburg commented that one does not want to stifle research with metrics, but one does not want to jeopardize funding, either. An interesting method would be to apply different metrics to different levels of research: very basic, high-risk, and high-payoff research should have different metrics than applied research.

Wright asked what some of the tradeoffs were in funding long-term research. It is easy to say that everything should be balanced, but it not so easy to understand how that should be done.

Wolff pointed out that the NSF has an implicit understanding of how the budget is divided and when deliverables are due.

Oliver commented that ASCR likes to think that its efforts are supporting basic research, but you have to be aware of the political pressures and the perceptions of the public and the media about the need and desirability of performing applied research.

Wright summarized where the Committee stood: It will not meet again before the facilities report is due to Decker's office. The full Committee will review what that Subcommittee does before the report goes on to DOE. The Biotechnology Subcommittee will prepare a report. McRae's Subcommittee will also bring something to the May meeting. A person may be added to the Committee to provide insight and expertise on information services. She called for public comment.

William McCurdy, of Lawrence Berkeley National Laboratory and a member of the Basic Energy Sciences Advisory Committee, said that the role of this Committee in the coming months should clarify whether computing is part of the DOE mission. It is clear from the Secretary's speech that it is. DOE probably controls about one-fifth of the high-performance computing on Earth. As an outside committee, ASCAC can make the value of that computing capability clear during the upcoming reviews.

Oliver noted that these reviews are going to happen quickly, probably in January. What role this Committee can play is not clear, although the Under Secretary is interested in advisory committees' input. Wright reiterated that Card said he needs more information on computing and would welcome our briefing him on this topic.

Corones asked about the possibility of preparing a white paper for him.

Wright noted that it is a case of time. Decker said that SC also wanted more information on the role of computing. We know the message and can come in to brief these people as needed and as asked.

McCurdy said that, if you have an invitation from Card, that is great. He is trying to protect parts of the program.

Wright said that she would be as forthcoming as possible to put this Committee's interests before DOE management. There being no further comment, she adjourned the meeting at 12:22 p.m.

Respectfully submitted

Frederick M. O'Hara, Jr.

Recording Secretary

Nov. 27, 2001

Revised

Margaret Wright

Chairwoman

Dec. ???, 2001