Draft Minutes
Advanced Scientific Computing Advisory Committee Meeting
March 30-31, 2010, American Geophysical Union, Washington, D.C.

ASCAC members present:

F. Ronald Bailey	James J. Hack
Marsha J. Berger	Anthony J. G. Hey
Jacqueline H. Chen	Thomas A. Manteuffel
Jack Dongarra	Linda Petzold
Roscoe C. Giles, Chair	William Tang

ASCAC members absent:

Susan L. Graham	Vivek Sarkar
Larry L. Smarr	Robert G. Voigt
John W. Negele	Victoria A. White

Also participating:

Melea F. Baker, Office of Advanced Scientific Computing Research, Office of Science, USDOE
William Brinkman, Director, Office of Science, USDOE
David Brown, Deputy Associate Director for Science and Technology, Computation Directorate, Lawrence Livermore National Laboratory
Christine Chalk, ASCAC Designated Federal Officer, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Vincent Dattoria, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Ian Foster, Director, Computation Institute, Argonne National Laboratory
Salman Habib, Nuclear and Particle Physics, Astrophysics, and Cosmology Group; Theoretical Division; Los Alamos National Laboratory
Barbara Helland, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Daniel Hitchcock, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Steven Koonin, Under Secretary for Science, USDOE
Douglas Kothe, Director of Science, National Center for Computational Sciences, Oak Ridge National Laboratory
Harriet Kung, Associate Director of Science for Basic Energy Sciences, USDOE
Arun Majumdar, Director, Advanced Research Projects Agency–Energy, USDOE
Paul Messina, Retired Director, Center for Advanced Computing Research, California Institute of Technology
Frederick O’Hara, ASCAC Recording Secretary
Walter Polansky (via Internet), Office of Advanced Scientific Computing Research, Office of Science, USDOE
Kenneth Roche, Fundamental and Computational Science Directorate, Pacific Northwest National Laboratory
Robert Rosner, Director, Argonne National Laboratory

Rachel Smith, Oak Ridge Institute for Science and Education
Michael R. Strayer, Associate Director, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Andrew White, Deputy Associate Laboratory Director; Theory, Simulation, and Computation Directorate; Los Alamos National Laboratory
Julia White, INCITE Manager at DOE Leadership Computing Facilities, Oak Ridge National Laboratory

About 70 others were in attendance in the course of the two-day meeting.

Tuesday, March 30, 2010
Morning Session

Before the meeting, new members of the Panel were sworn in by members of the DOE Human Resources staff, and the Panel was briefed on ethics issues by a member of the DOE General Counsel's Office. During the meeting, the floor was opened to public comment after each speaker.

Chairman **Roscoe Giles** called the meeting to order at 8:01 a.m. **Rachel Smith** made safety and convenience announcements. Giles reviewed the agenda and introduced **William Brinkman** to describe the FY11 budget request to Congress for DOE's Office of Science (SC).

In the FY11 budget request to Congress, SC got about a 4.4% increase. Congressional direction may be added; it is less this year than last. Advanced Scientific Computing Research (ASCR) got an 8.1% increase (with extra money for the exascale exercise), Basic Energy Sciences (BES) got a 12.1% increase , Biological and Environmental Research (BER) got a 3.8% increase (the percentage would have been larger had the pork in last year's budget not been counted), Fusion Energy Sciences (FES) got a decrease of 10.8% (because of last fall's difficulties, which have now been worked out with a new schedule; the amount will restored next year), High Energy Physics (HEP) got an increase of 2.3%, Workforce Development for Teachers and Scientists got an increase of 72.2% [because of the Graduate Research Fellowship Program established last year with American Recovery and Reinvestment Act (ARRA) funds], and science laboratories infrastructure got a decrease of 1.3% (rebuilding was done last year with ARRA funds, so this value appears less but is actually a significant increase). There are now 26,000 people using SC's user facilities each year across the country. The light sources are the biggest class of these facilities with one-third of the users.

The quality of the team brought in by this administration to manage the scientific enterprise is impressive.

Some organizational structures are being built around some areas of research, called hubs. They will provide focus and leadership for progress in these areas. A new one is on batteries and energy storage. Many trillions of dollars are going to be needed to deal with climate change. In DOE, there are pockets of expertise on climate science; these pockets will be pulled together in a more integrated fashion. In computing, the possibility of exascale computing is being investigated; the challenges (power, architecture, failure rates, etc.) are huge. A particular emphasis is on multiscale modeling of combustion and advanced engine systems. All scientific user facilities are funded. The SC Graduate Fellowship Program and the SC Early Career Research Program (ECRP) are funded.

SC has one of the three hubs in FY10: Fuels from Sunlight. Another hub, Modeling and Simulation of Advanced Nuclear Reactors, has strong ties to ASCR. A new FY11 SC/BES Hub for Batteries and Energy Storage
will address design of advanced materials architectures, control of charge transfer and transport, development of probes of the chemistry and physics of energy
storage, and development of multiscale computational models.

The Energy Frontier Research Centers (EFRCs) now number 46, representing 103 institutions in 36 states plus the District of Columbia.

SC leads the World in supercomputing capabilities. As Sec. Chu has said, "supercomputing modeling and simulation are changing the face of science and sharpening America's competitive edge." Leadership computing is making progress at the petascale in turbulence, nuclear energy, energy storage, fusion energy, biofuels, and nanoscience. Modelers claim that they can simulate the shape of trucks and save about 3 miles per gallon. Multiscale Simulation of Internal Combustion Engines is being used to develop more efficient and cleaner engines. Models include 150 types of combusting molecules (previously, there was only one type).

The light sources of the world have been doing structural biology for many years. The 2009 Nobel Prize work used all four BES light sources. This was the third Nobel Prize for this type of work. A database of 10,000 proteins has now been amassed to elucidate how ribosome translates the genetic instructions encoded by DNA into chains of amino acids that make up proteins.

The Linac Coherent Light Source (LCLS) has been turned on at the Stanford Linear Accelerator Center (SLAC); it is the world's first hard X-ray light source, providing an unprecedented combination of high spatial and temporal resolution for the investigation of atomic-scale structure and processes. Within 2 hours of turn-on, it was producing scientific data. It allows single-molecule protein crystallography in which diffraction patterns are observable just before the laser beam destroys the crystal. This opens an advanced area of science.

There are three Bioenergy Research Centers (BRCs): the Joint BioEnergy Institute (JBEI), Great Lakes Bioenergy Research Center (GLBRC), and BioEnergy Science Center (BESC). They are focused on converting cellulose to ethanol through genomics and microbiotics. At the JBEI, they have converted *E. coli* to change sugar to fuel in one step.

The genomic revolution is sequencing the 1.1-billion-base-pair soybean genome, publishing the *Genomic Encyclopedia of Bacteria and Archaea* [DOE/Joint Genome Institute (JGI)], and investigating viable microbes in toxic subsurface environments. It took 13 years to sequence 3 billion base pairs; it is now possible to do a trillion base pairs each year. This capability will lead to affordable whole human genome analysis.

In Nuclear Physics (NP), the Relativistic Heavy Ion Collider (RHIC) collides heavy ions at near light speed and has discovered a violation of parity in the strong interaction. The Continuous Electron Beam Accelerator Facility (CEBAF) is studying quark-antiquark fluctuations. Investments in the Facility for Rare Isotope Beams (FRIB) probe the properties of rare nuclear isotopes. NP has taken on the production of medical isotopes and is doing a good job.

In High Energy Physics (HEP), the United States has developed components for the LHC at CERN {Conseil Européen pour la Recherche Nucléaire [now European Organization for Nuclear Research (Organisation Européenne pour la Recherche Nucléaire)]} and hosts centers for data analyses. The LHC has turned on again and is building up its integrated luminosity. (The Tevatron is investigating the same area.) HEP captures the unique opportunities of neutrino science at the Tevatron, Sudan, and (in the future) the Deep Underground Scientific and Engineering Laboratory (DUSEL). Critical decision zero (CD-0) for this Long-Baseline Neutrino Experiment (LBNE) was approved in January 2010; the FY12 budget request provides for a preliminary engineering design for neutrino research.

$16 million will be available in FY11 to fund about 60 additional Early Career Research Program awards at universities and DOE national laboratories. There will be 150 additional grants this year for the SC Graduate Fellowships, for which there are 3000 applications.

He said that he was looking forward to seeing what ASCAC has to say about exascale computing.

Hey asked what had come of the ARRA funding. Brinkman replied that SC got $1.6 billion; it is all obligated. The Graduate Fellowship and Early Career programs have come out of ARRA.

Tang pointed out that there were possible alliances with the National Science Foundation (NSF) on training and fellowships. Brinkman said that, while that is attractive in general, he preferred to keep DOE’s separate. SC’s fellows will be offered experience at the national laboratories during the summer.

Giles noted that ASCR is called on to support facilities with impact across all of SC. Brinkman replied that ASCR is not paying for everything. The Innovative and Novel Computational Impact on Theory and Experiment (INCITE) is paid for in full. Scientific Discovery Through Advanced Computing (SciDAC) has been funded in part by ASCR.

Michael Strayer was asked to present the viewpoint of the ASCR Office.

A major accomplishment of ARRA funding was the deployment of the world’s most powerful computer at Oak Ridge National Laboratory (ORNL), which was carried out seamlessly and flawlessly in stages. It has been doing science for some time now.

Hopper Phase-1 of the National Energy Research Scientific Computing Center (NERSC) has been accepted. The early user period ran from November 2009 to February 2010. More than 10 million hours have been delivered to DOE users. When Phase-2 is completed, the machine will have just over a petaflop of capacity, and capacity at NERSC will have been increased by a factor of 100.

Current ASCR funding opportunity announcements include an Applied Math solicitation on Advancing Uncertainty Quantification (UQ) in Modeling, Simulation, and Analysis of Complex Systems. It is for $3 million per year for 3 years to fund two to six awards. There are three Computer Science solicitations. The first is for R&D in X-Stack Software Research that supports extreme-scale scientific computing, from operating systems to development environments; it is for $10 million per year for 3 years to fund four to five awards. The second is for Advanced Architectures and Critical Technologies for Exascale Computing; it is for $5 million per year for 3 years to fund four to five awards. And the third is for Scientific Data Management and Analysis at the Extreme Scale; it is for $5 million per year for 3 years for 10 to 15 awards. In Next-Generation

Networks, there is a solicitation for High-Capacity Optical Networking and Deeply Integrated Middleware Services for Distributed Petascale Science to energize the advanced networking/middleware projects and to take the Energy Sciences Network (ESnet) to terabit-per-second optical networks.

The ASCR FY11 President's Budget Request represents an 8% increase. The largest increases are in facilities for lease payment. In research, exascale computing areas will continue to be our focus. . The FY11 budget seeks to maintain the FY10 increases in Applied Mathematics and Computer Science to prepare for challenges of future architectures, huge datasets, and multidisciplinary science and to build on computational-partnership teams on transforming critical DOE applications to be ready for running at scale on multicore computers. This activity will be robust for science and for applied science. The budget will fulfill obligations at computing facilities (leases) and to the Defense Advanced Research Projects Agency (DARPA) High-Productivity Computing Systems program. It will begin site preparation for the ALCF-2 upgrade (of the Argonne Leadership-Class Facility) to be installed in 2012-2013, will support NERSC-6 (the Hopper machine) operations, and will support ESnet deployment of 100-Gbps technologies.

New ASCR staff are Sonia Sachs in Computer Science; Richard Carlson in Collaboratories and Middleware, part of the ESnet buildup; and Betsy Riley as a program manager for the Argonne Leadership Computing Facility in the Facilities Division. George Seweryniak is retiring March 31, 2010. Lali Chatterjee became Director of the SC Office of Communications and Public Affairs. The position of Director for the Facilities Division has been posted and closes May 3, 2010. This person will oversee the entry into exascale science.

ASCR Early Career Research Program (ECRP) awards have been made across the application areas of applied math, computer science, computational science applications, and network environment research. In this last cycle, 464 letters of intent and 369 full applications were received and 354 proposals were ultimately reviewed with 60 coming from national laboratories and 294 from universities. Ultimately, seven Early Career awardees were selected:

- Youssef Marzouk, Massachusetts Institute of Technology, in applied math: uncertainty, stochastic, and complex systems
- Anil Vullikanti, Virginia Polytechnic Institute and State University, in applied math: data and discrete system analysis
- Patrick Chiang, Oregon State University, in computer science: hardware and power management
- Michelle Strout, Colorado State University, in computer science: programming languages, models, environments, and compilers
- Grigory Bronevestky, Lawrence Livermore National Laboratory, in computer science: operating systems and fault tolerance
- Kalyan Perumalla, Oak Ridge National Laboratory, in computer science: programming languages, models, environments, and compilers
- Christiane Jablonowski, University of Michigan, in computational science: climate

All of the very bright applicants need to be encouraged to stay in areas of need of the ASCR research portfolio.

The number of applications for the Computational Science Graduate Fellowship increased from 157 in 2002 to 531 in 2010, showing the pressures on the program.

Manteuffel pointed out that the leases for the Leadership-Class Facilities (LCFs) are expiring or paid up and asked, what would happen now. Strayer replied that the “bad year” for the leases was FY11. Some tails will have to be paid off in FY12. A balance will be maintained between R&D and facilities. Research will increase each year through FY14. Strategic planning has been done, and the program managers are developing a strategic plan. There will be a robust buildout.

Dongarra asked what programs were affected by the Applied Mathematics decline from FY09 to FY10. Strayer answered that these numbers are the result of a complex iterative process. Rolloffs of existing programs in FY10 caused the $700,000 decrease and will continue to do so. This is the typical random walk of federal funding. The average over the years will be the important trend.

Hey asked how many ECRPs there would be. Helland replied that there were funds for six candidates; an extra one was chosen because there were so many strong candidates; the program will continue at this level in the future.

Berger asked for a more detailed explanation of the X-stack research. Hitchcock said that this is to do the research to underpin the software to be developed for exascale computing. It has to be done early so the software can be written. Working with the hardware people is essential at this level. Strayer added that one cannot support a hardware buildout with the current software knowledge. Working needs to be done on the software to see what ideas have merit.

Chen asked whether computational partnerships (like SciDAC) can be expanded outside SC. Strayer answered that the Office was talking with and has held workshops with several of the applied offices to see how ASCR might partner with them. The discoveries are encouraging.

Manteuffel asked for more information on the strategic plan. Strayer replied that there had been a meeting of program managers from which has come a skeletal idea of a strategic plan. It will have a 5-year forward look and may or may not be ready for the next ASCAC meeting.

A break was declared at 10:29 a.m. The meeting was called back into session at 10:48 a.m.; **Douglas Kothe** and **Kenneth Roche** were asked to present an update on recent activities and outcomes on the ASCR Joule metric for computational effectiveness.

The Joule metric is to improve computational science capabilities, defined as the average annual percentage increase in the computational effectiveness (either by simulating the same problem in less time or simulating a larger problem in the same time) of a subset of application codes. For each application, a benchmark is run at the beginning and end of the fiscal year, and the report discusses the differences. In the strong scaling mode of enhancement, one can simulate the same problem in less time; and in the weak scaling mode, one can complete a larger problem in the same time. One can also consider other variations, such as those in the problem, algorithm, and machine to achieve the fixed-time assertion.

From the machine perspective, complexity is driven by the floating-point computations, and efficiency improvement is achieved by reducing the number of instructions read.

In 2004, a standardized method to measure the efficiency was needed. A partnership with programmers and the machine centers was established. Climate, fusion, and combustion research have dominated the process. Two programs, Q2 and Q4, were benchmarked.

For FY09, Joule metric results were obtained for the four applications: RAPTOR, CAM, XGC1, and Visit. Performance gains were what was (aggressively) expected in terms of time to completion. This year, a new tool was available.

The Community Atmospheric General Circulation Model (CAM) was run in uncoupled mode. It is characterized by two computational phases, dynamics and physics, in latitude, longitude, and elevation. The year-end report shows a speed-up by a factor of 2 on the same hardware with changes in the code with OpenMP.

XGC1 is part of SciDAC and is used to understand and predict plasma transport and to profile the “edge pedestal” around the separatrix, a weak-scale problem along the time dimension. Time steps were increased from 4000 to 16,000. This was the first attempt to study the nonlocal H-mode coupling physics between the edge and core turbulence in a realistic DIII-D tokamak geometry. It was done successfully by increasing the time steps by a factor of 4. The focus of this code is solving a gyrokinetic Poisson equation, and solutions have to be interpolated back to particle positions to time evolve according to equations of motion. An improvement in message-passing interface (MPI) communication resulted in faster Poisson solution. One-sided message passing (OMP) parallelism was implemented, allowing the use of a quarter as many MPI processes.

Visit was used to do two problems: isosurface extraction and volume rendering. Isosurface extraction extracts the 3-D points in a volume with a specific value and connects them with a continuous surface. Volume render produces an image from a scalar field in a 3-D data set. One can now scale very well in the volume renderer because a problem that was previously unrecognized was fixed.

RAPTOR is a large-eddy simulation of turbulent, reactive, multiphase flows. It is fully modular, self-contained, and written in ANSI [American National Standards Institute] standard Fortran 90. It uses an interesting integration scheme to understand and apply Reynolds number scaling in combustion modeling, which is crucial if simulation is to affect engine design. This is a 15-year-old code, but performance was increased by a factor of 2 by improving halo exchanges and by knocking down MPI wait times.

The FY10 metric applications are TD-SLDA, POP, Denovo, and LS3DF. Things are on track for each application.

It was found that the metrics team needs to work with science teams in quantifying a bottom-line number for science metrics. It needs to be done carefully. One needs to be cautious about targeting full-system Q4 problems on systems that are new and/or being upgraded. In performance data collection, a verified standard approach for collecting performance is needed. Documenting the methods used to gain performance has begun. Joule application teams are brought together as one project team by holding weekly meetings/telecons and with an e-mail reflector. Results and documentation are shared, which drives accountability, competition, and emulation among the code teams. Ample leadership computing resources are provided that are not implicit in any pre-existing allocation (only 10% of the time burned). Off-code-team experts are brought in as part of the Joule effort.

Berger noted that one thing done was turning off the IEEE arithmetic and asked why precision was cut down. Kothe replied that some codes are okay with single precision at certain points. One can validate this by comparing results. One can speed up performance by going to simple precision. In general, though, there is no good answer. Hack noted that, before, there was 32-bit arithmetic. There are tests that introduce random errors and analyze precision statistically. Dongarra said that he did not see any assessment of the effort that went into the improvements. Kothe replied that there were not any data kept on manhours consumed. One to three man-months were used in carrying out the benchmarks. One finds some areas of improvement, but sometimes one does not know where to look. Roche added that the 10% number came from estimates of time spent by people on programming teams. Productivity is bounded by the tools available.

Petzold asked if the goal was speeding up codes. Roche replied that the goals were to move software toward scalability and to satisfy the metric language. Giles noted that this is part of the Office of Management and Budget (OMB) performance metric for ASCR. Hitchcock stated that one has to show that one is helping people do science. One can measure how productive machines are. One can also see how performance increases with the programming underlying applications. How much time it takes for the program to do this has not been measured. Kothe said that this Committee chooses the applications. Performance has improved over the years, and the performance increases can be leveraged in other programs. Petzold observed that, from a science perspective, each new architecture produces new barriers to the users and asked if efforts were being made to share lessons learned. Kothe replied, yes. It varies with the critical path of the application.

Julia White was asked to report on the management of the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program for which the 2008 Committee of Visitors (COV) recommendations were found invaluable in managing the program.

INCITE provides awards to academic, government, and industry organizations worldwide. Beginning in 2010, INCITE is jointly run by the ALCF and the Oak Ridge Leadership-Class Facility (OLCF) and managed by Julia White.

The impact of science and technology is the key driver of INCITE. It allows introduction of innovation and new physics into climate models. Advances in nuclear reactor modeling are aimed at making safe, clean nuclear energy available globally. High-temperature superconductor design requires revolutionary simulations for developing improved energy transmission. GE is looking at next-generation energy and propulsion.

LCF allocation programs have more than 2.7 billion processor hours available. INCITE allocates 60% of those hours for projects of high impact in science, engineering, and computer science that require leadership systems. Proposals are considered year-round and can be awarded up to millions of hours. There is also the ASCR Leadership Computing Challenge (ALCC), which is allocated by the ASCR Office.

Now in its seventh year, the INCITE program has seen explosive growth resulting in more than 1.6 billion processor hours being allocated in 2010. INCITE supports a wide range of disciplines. There is no designated number of hours for a particular science area. The emphasis in INCITE looks at all recommendations, focusing on potential for scientific or technological impact and then readiness. It has experienced tremendous success; it is oversubscribed by a factor of 2.5. The number of projects increased rapidly from 2 to 20 from 2004 to 2010 and is now flattening out. The average number of processor hours has increased significantly, from 1.6 million to 23.2 million over the same period. The size of the extreme uses has also increased.

INCITE 2010 awards went to 35 new projects and 34 renewal projects, with 41% of the new submittals and 83% of renewals being awarded. The distribution reflects a broad range of disciplines. Renewals are held to an achievement threshold. The 2010 project demographics by principal investigator (PI) affiliation are industry, 4%; DOE, 38%; university, 54%; and government (non-DOE), 4%. There were three projects with foreign-based PIs.

The 2011 allocations are now being planned, with a call for proposals from April 14 to June 30 and allocation awards through December 2011. The computational readiness review is conducted by center staff who are expert in these systems. Both centers review each proposal. The criteria for new proposals look at appropriateness for requested resources, appropriateness of the computational approach, and the technical readiness. The criteria for renewals are meeting the technical and computational milestones and being on track to meet future milestones.

The second step is the panel review by domain experts drawn from institutions worldwide who are looking for the scientific and/or technical merit, appropriateness of proposal method, team qualifications, and reasonableness of requested resources. For renewals, they look for change in scope, whether the awardees met technical/scientific milestones, whether they are on track to meet future milestones, and the impact relative to other proposals under consideration.

Based on the 2010 INCITE experience, a feedback step to authors was introduced for computational-review results. Program transparency was enhanced by providing reviewer comments to authors at the end of the review process (a response to a 2008 COV recommendation). A “blue ribbon” peer review panel has been created. A review of renewals has been initiated, another response to a 2008 COV recommendation. The ALCF and OLCF management teams and the INCITE manager have been engaged in the selection of proposals.

The 2010 INCITE panel reviewers were asked for their feedback on whether the INCITE proposals discussed in the panel represent some of the most cutting-edge computational work in the field, whether the proposals were comprehensive and of appropriate length given the award amount requested, whether the science panel was sufficiently diverse to assess the range of research topics being considered, and whether having access to the center’s computational readiness reports was valuable for assessment of the proposals. The responses were overwhelmingly positive.

The 2011 INCITE process will

- Post the reviewer questions with the call for proposals to increase the transparency of the assessment process for proposal authors
- Build the review panel earlier in the year
- Identify potential opportunities for making referrals from INCITE to the Energy Research Computing Allocations Process (ERCAP)
- Reassess and potentially redesign the web-based form for proposal submittal
- Build relationships with other centers whose users have science challenges that would benefit from the scale of the INCITE resources

Chen noted that industry was a small percentage of INCITE participants and that some metrics were different for their proposals than for proposals from other classes of participants. J. White replied that reviews last year said there should be a separate set of evaluative questions for industry, asking for competitiveness impacts. These criteria would be evaluated by external reviewers, also.

Bailey asked about the results of the proposals and the computations done and how they were distributed. J. White responded that the centers work closely with the participants and issue reports on scientific impact. A peer assessment of science impact is also sought.

Berger asked what percentage of the project uses “the full machine.” J. White said that she did not have that number available but could provide it later.

Daniel Hitchcock was asked to review the ARRA investments in ASCR.

Magellan is ASCR’s experiment in cloud computing. It is a $32 million project at NERSC and the ALCF consisting of a 100 TF/s compute-cloud testbed. The first hardware is now available. The question is how to run mid-range science and schedule usage on the cloud. We are using Eucalyptus at ALCF and Linux at NERSC. This is not part of ASCR’s mission, but ASCR is the best suited to address the question. The NERSC Magellan cloud hardware has 720 nodes, and 5760 cores in nine scalable units producing 61.9 TF and networking between sites will be looked at, as well.

The key is flexible and dynamic scheduling of resources, an interesting optimization problem.

We want to accelerate the commercialization of 100-Gbps networking technologies by deploying a national-scale prototype network that will span for distinct geographic regions, connecting the three major ASCR computing facilities and the New York multiagency peering point, providing transatlantic connectivity at 100 Gbps. Another goal is to complement the prototype 100-Gbps network with a testbed providing an experimental network research environment at sufficient scale to usefully test experimental approaches to next-generation networks and applications. The project plan is available.

The Advanced Network Initiative (ANI) is running in parallel with ESnet. The goal is a 100-Gbps prototype network preliminary baseline design.

The Oak Ridge LCF was upgraded to 6-core CPUs and worked with users to produce scalable, high-performance applications for the petascale.

Scientific progress has been made in turbulence, nuclear energy, energy storage, fusion energy, biofuels, and nanoscience.

Advanced architecture is a small project in two pieces: IBM’s Productive, Easy-To-Use, Reliable Computing Systems (PERCS) and the University of California at Berkeley’s Research Accelerator for Multiple Processors (RAMP). Its goal is to develop a complete initial definition on architectural features and performance levels for a system that will meet the needs of a least one science application that requires extreme-scale computing while using energy efficiency.

SciDAC-e is funding seven applied-mathematics research efforts in support of DOE electricity grid efforts. All awards have been made, and work has begun. Postdocs have been hired to provide expert-user support for Extra Energy-related projects at the leadership computing and NERSC facilities.

The floor was opened to public comment. There was none.

Manteuffel noted that the Joule Project has had a remarkable impact with just two postdocs and asked if there were any plans to scale it up. Hitchcock said that the results are leveraged by applying the improvements to other programs, especially those with large numbers of users. The results are published to allow others to adopt these improvements.

Messina stated that the support people see changes coming out of Joule (and other efforts), leveraging the results tremendously. Roche seconded that comment. Tang noted that it is not just faster codes but more substantial science.

A break for lunch was declared at 12:17 p.m.

Tuesday, March 30, 2010
Afternoon Session

The meeting was called back into session at 1:30 p.m. **Andrew White** was asked to report on the DOE Laboratory Exascale Working Group.

In June 2009, the eight national laboratories were asked to look at science applications; systems software and programming models; systems acquisition, deployment, and operations; and hardware technology R&D. The plan targets exascale platform deliveries in 2018 and a robust simulation environment and science and mission applications by 2020. Co-design and co-development of hardware, software, programming models, and applications requires intermediate platforms in 2015.

Three town hall meetings and 11 grand challenge workshops were held. Meetings were held with nine high-performance-computing vendors and then with universities, national laboratories, and vendors in 2009. An International Exascale Software Project is looking at the required software environment.

There will be a lot of competition for world leadership in the 21st century. China's computer capability is increasing rapidly, China is granting about the same number of PhDs in science and math as is the United States, and China is devoting a much greater portion of gross domestic product (GDP) to computing R&D than the United States is.

A community effort is needed to achieve the exascale. DOE and private industry (and partnerships) will play a role, also. Co-design (hardware and software) is necessary. Meeting attendees focused on the technological needs, co-design, and DOE mission needs.

Between 1992 and 2006, there was a factor of 1000 increase in computer performance, most based on increases in clock speed. Now, parallelism and concurrency must drive the increase in performance.

System power will be a problem. The processor is where most innovation will happen in response to all of these pressures. The operating system is not ready. The strategy and implementation times will be much shorter.

The target "swim lanes" are 200 petaflops/sec in 2015 and 1 exaflop in 2018. An exascale machine might operate at 1 TF or 10 TF. The mean time between failures (MTBF) will likely be 1 day.

The high-level system design may be similar to that of the petascale system of today, Windows Message-Passing Interface (WMPI), but checkpoint restart is likely too expensive to do. The node is the lag. It will have to have a heterogeneous functionality, a deep memory hierarchy, and a new programming model. All this will be stacked in a 3-D chipset. The intellectual property will be shifted to the CPU from the memory.

Power is the number-one issue. Memory bandwidth costs power. It will cost 100s of millions of dollars to reduce memory power usage to what would be needed for a 20-MW machine. The amount of memory will not scale with the increase in computer (CPU) performance. (CPU design is moving faster than memory design.) The industry will have to make the MTBF per component 1/10th of what it is now. One can put some storage memory on the node, and this will take pressure off the memory stack. There are "guard bands" on temperature, voltage, etc. Silent errors (where a value does not get set properly because something did not happen on time) will happen more and more frequently. System software will have to change with so many processors and threads. The work will be broken up between processors.

Two approaches to the programming model are inter-node and intra-node. A unified approach will probably be designed simultaneously. Some decisions will have to be made about the programming mode early on.

Co-design expands the feasible solution space to allow better decisions. Co-design of applications and technology will need to be optimized for power, performance, price, prediction, and/or productivity. Tools will need to be developed for a virtual machine to probe the issues of co-design.

The workshops focused on four DOE missions: climate, nuclear reactor design, combustion, and national nuclear security.

The National Ignition Facility (NIF) is looking at break even this year. Wake-field accelerators are projected for the end of the decade. These and other breakthrough scientific discoveries and facilities require exascale applications and resources

Climate change is the greatest problem being faced by the world today. The Earth system model uses coupled land, atmosphere, ocean, and sea-ice models. This model represents multiphysics interactions. Exascale computers would have impact on climate assessments. We need models that get the physics in greater detail. Climate needs to be predictable on decadal time scales. Ensembles of programs will be used to produce this predictability.

In energy, the focus will be on nuclear energy, combustion, transport (one-quarter of the CO_2 production), fusion and renewable energy. The range of scales is 20 orders of magnitude in time and size. It would be good to get the Nuclear Regulatory Commission (NRC) to recognize UQ.

The workshops looked at combustion as a DOE initiative.

The National Nuclear Security Administration annually certifies the stockpile, directs stockpile work, and conducts life-extension programs.

Three crosscutting issues came up in the workshops: UQ; multiscale multiphysics modeling; and statistics of rare events, all of which require exascale computing.

Errors enter not only through hardware but also in the way computations are done. The question faced is what programs can be gotten up and running at the exascale in 2018.

Bailey asked whether, in the co-design approach, it had been considered that one programming model might not be appropriate under all circumstances. A. White replied that that has been considered.

Tang stated that UQ validation would require dedicated facilities. A. White replied that, on the application side, they have to be bound to a mission. Also, in a larger context, the real power of co-design is adding functional value. It would be advantageous to get the experimental community involved.

Manteuffel asked how one can get the rest of it and whether we know we are going to be able to get there. A. White said that this is a transformation in computer technology. That is what exascale is about. There will be transformative changes in applications, too.

Berger asked if he had said that the user will have to manage the cache. A. White answered that, on Roadrunner, instructions and data cache have to be managed. Data movement will be expensive, so one will want to explicitly manage cache on the core. One should know where it is and where to put it.

Bailey noted that, in the past, high-performance computing had a lot of influence on the industry and asked whether the exascale will influence where the industry is going. A. White responded that investments will have to be made in memory bandwidth. Simpler processors will be installed, leading to more threads.

Chen asked what experience there was with importing MPI. A. White said, that was assembly language, the programming of which was not as difficult as it was thought it would be. But it was less efficient.

Petzold asked where this next generation of programmers will come from. A. White replied that it is everybody's problem so there is a lot more focus on that problem.

Hey asked what breakthroughs would be made with the exaflop. A. White responded that, in climate, it is cloud-resolving climate models. With weapons, we cannot do the basic science to get the material science right. In ASCI it was going to 3-D codes.

Foster, noting that White had mentioned a whole set of things that could be done with exascale computing, asked whether a scripting language will be needed. A. White answered, yes.

Messina said that there are a few workshop reports on the web. There is not a strong case for the exascale in all, but there is in most.

David Brown was asked to discuss the cross-cutting technology for computing at the exascale.

A workshop was held on the cross-cutting technologies required to deliver exascale computational science by 2018. Co-design is the interaction of system architecture, system software and tools, math models and algorithms, programming models, and scientific applications at the exascale. The workshop objectives were to outline the R&D needed for co-design of the exascale computational science environment, to identify opportunities for "disruptive" computational approaches for future scientific discovery that might be employed, and to produce a first cut at the characteristics of a hardware/software system roadmap that will meet science application needs during the next decade. This was the first workshop that brought applied mathematicians and application developers into the discussion.

Co-design will be essential for exascale scientific discovery by 2018. Tightly coupled multidisciplinary partnerships will ensure delivery of science applications on exascale platforms. The transition to the exascale will be as disruptive as was the transition from vector computers. Appropriate investments will be required. The Advanced Simulation and Computing Initiative (ASCI) spent only 20% of its funding on hardware. Significant investments are needed in computer science and mathematics research and in redesigning and rewriting of application codes.

The cross-cut workshop was a co-design “practice session.” It brought disparate computational science research communities together to understand exascale challenges. Breakout sessions successfully overcame communication barriers, and all the participants left the meeting with a deeper understanding of each other’s communities. Even the skeptics said that co-design could really work. The designers of extreme-scale hardware must obtain a detailed understanding of the scientific challenges. A multidisciplinary computational science culture has blossomed during the past 15 years. Organizational changes will be essential to meet the 2018 target.

Breakout sessions addressed three themes:

1. Mathematical models and algorithms, including (1) the impact of application needs and architectural developments on mathematical models, algorithms, and programming models and (2) the impact of application, mathematical models, and algorithm needs on architectural development
2. System software, including the functionality required at the exascale and tasks that will need to be addressed somewhere else than in the system software
3. Programming models and environment, including the models and environments needed and the abstractions and tools needed by applications and algorithms

Six “math” areas were used to provide context for the discussions: partial differential equations (PDEs) I, PDEs II, UQ/stochastic, discrete math, data/visualization, and solvers/optimization. Final sessions synthesized results for each theme.

The exascale is not just the petascale times a thousand. Traditionally, PDE-based applications have expected increases of a factor of 10 in resolution with each factor of 1000 increase in compute capability, but not this time: 1000 times the memory will not be available, the processors will not be 10 times faster, and the machines will not be able to move as much data on or off each processor. The introduction of massive parallelism at the node level is a significant new challenge, and that means that MPI is only part of the solution. However, exascale computing is an opportunity for more fidelity (incorporating more physics instead of increased resolution) and greater understanding (developing UQ to establish confidence levels in computed results).

UQ is likely to permeate the exascale. UQ is the end-to-end study of the accuracy and reliability of scientific inferences. Large-ensemble calculations will have dynamic-resource-allocation requirements significantly different than those of traditional applications. Traditional space-shared, batch-scheduled usage is unlikely to be effective for UQ or new multiphysics codes. The client/server model for UQ requires a different failure model. Significant code redesign is likely.

Understanding the characteristics of PDE solvers is important for co-design. Domain decomposition with nearest-neighbor communication patterns will be needed. Elliptic solvers have smaller, nonlocal communication patterns. The network will need low latency, high bandwidth, and high message rate for point-to-point and collective communication operations. And it will be highly desirable that the physical topology of the machine matches the communication patterns.

Memory hierarchy becomes deeper and more complex at the exascale. The code will take a more active part in memory management. There are inadequate tools and interfaces to hint, manage, and control memory for run-time systems. Fine-grain, node-level parallelism in PDE solvers could exploit a hierarchical two-level machine/programming model. One wants system software that can exploit spatial/temporal locality hints from the application code about what to do. Cache is energy-expensive; alternative fast local memory access approaches are needed for performance and energy savings. Low-cost thread creation/destruction is essential for performance.

At the exascale, one will want adaptive run-time systems that could address dynamic load balancing, dynamic power allocation, the ability to reconfigure around faults, and dynamic resource requirements. Applications must care more about fault tolerance and resilience. Checkpoint-restart will not scale with current storage systems. Co-design will be required to develop a standard fault-management application programming interface (API). Application-specific fault recovery is likely. Local recovery from faults should be considered.

Discrete math applications will challenge exascale machines. They move a lot of data around with a few FLOPS [floating-point operations per second]. Adaptive runtime will be important. Energy efficiencies could be achieved with dynamic power allocation between FLOPS or data movement. The machine will need to be able to efficiently handle irregular data. Co-design opportunities include discrete-event simulators for exascale architectures and the use of graphic algorithms for task scheduling on nodes and across nodes.

Familiar system support issues will be even more challenging at the exascale. File-system scalability and robustness will continue to be the weakest link at the exascale. Hierarchical debugging tools will be needed. Performance tools are not keeping up with the largest machines, resulting in an understanding gap as the exascale is approached because existing performance tools do not address heterogeneous architectures; research is needed to develop a vertically integrated performance-analysis tool for exascale applications.

There is optimism that there will be compilers to support heterogeneous and multicore processors by leveraging recent advances in compiler technology. There will be new opportunities in power management, small memory capacities, resiliency, and interoperability.

The preliminary panel findings are grouped into three categories:

1. Algorithm R&D is needed to support new architectures.
2. R&D is needed for programming models to support exascale computing.
3. R&D is needed for system software at the exascale

In algorithms R&D, it will be required to recast critical applied-math algorithms to reflect the impact of the anticipated macro-architecture evolution. Data-analysis algorithms need to be adapted for the exascale. Numerical-analysis questions associated with the move from bulk-synchronous to multitask approaches need to be addressed. “Mini-applications” and simulations of emerging architectures need to be developed.

Critical applied-math algorithms include, for PDEs, new PDE discretizations reflecting a shift from FLOP to memory-constrained hardware and new algorithms with more compute and less communication. For UQ, there is an opportunity to redesign codes with UQ built in and to move statistics inside loops. For solvers and optimization methods, there is a need for solvers that have reduced global communication, that leverage low-latency on-chip all-gather, and that have new sparse eigensolver formulations and fast Fourier transforms. For novel algorithms, reduced-precision arithmetic algorithms that store less, but maintain accuracy, will be needed.

Data-analysis algorithms could leverage increased node-local NVRAM [nonvolatile random access memory] availability, using analysis algorithms for streaming data and leveraging global address space. The best place to do analysis may be in situ as part of the simulation code, during postprocessing on the exascale platform, or during postprocessing on a dedicated analysis platform. Research will have to be conducted on the development of common data structures or data-access patterns to enable reusable data-analysis software.

Numerical-analysis issues will need to be addressed like asking what is known about accuracy, stability of multiphysics and multiscale coupling; high-order operator splitting methods; and accuracy and stability of methods that apply operators more asynchronously.

The role of simulation in co-design is to develop "mini-applications" that capture the essential elements of large scientific applications and to develop the simulation tools for emerging architectures.

New programming models will have many important requirements:

- To investigate new exascale programming paradigms
- To develop APIs for dynamic resource management
- To support memory management at the exascale
- To study scalable approaches for input/output (I/O)
- To supply interoperability tools to support the transition to the new environment
- To provide language support for programming environments at the exascale
- To give programming-model support for latency management and fault tolerance/resilience
- To provide APIs for power management

Investigating and developing new exascale programming paradigms will be needed, such as hybrid programming models, effective abstractions that expose loop-level and data-level parallelism, improved abstract machine models, programming model support for multiple networks on the same machine, and a new programming model with an opportunity to change how computational science is done.

Memory management, I/O, interoperability, and language support will require programming model support for data-structure linearization and language support for uncertainty-carrying variables.

Latency management will be needed to overlap computing, analysis, communication, and I/O. Programming models will need to be power-aware. For resilience, there will need to be programming-model support for fault management, fault-tolerant MPI collectives, and API for checkpointing.

Topics that came up in system software were tools to support node-level parallelism, system support for dynamic resource allocation, system-software support for memory access, performance/resource measurement and analysis tools, system tools to support fault management, and system support for exascale I/O.

Node-level parallelism will require small, light-weight messages; light-weight, fine-grained, and flexible synchronization; low-cost thread create/destroy; fast all-reduce; tools to manage communication patterns; tools to support moving away from bulk-synchronous parallelism; and tools to support maintaining the local state when objects migrate between processors.

For performance/resource management and analysis, one might use data-mining methods to help develop performance measurement tools, performance tools for heterogeneous environments, and system calls to query relative costs of various operations.

Finally, one might have system tools for fault management, such as tools to support fault-tolerance management; research in debugging at scale; research in the fault-tolerance implications of UQ; and a taxonomy of faults to support advanced fault handling. If a smaller system (e.g., 10%) is used for data analysis, resilience will be less of a problem because it will not be possible to move large amounts of data off the main compute platform.

The workshop concluded that co-design is essential for exascale scientific discovery by 2018. It will need close multidisciplinary partnerships to ensure the delivery of science applications on exascale platforms. All partners must commit to significant changes in both hardware and software design. New programming paradigms will be required, with the emphasis shifting to physics fidelity and UQ. Appropriate investments will be required.

Petzold asked what the next killer app was for exascale computing. Brown responded that that was not the purpose of the workshop.

Hack asked if there were a dollar amount to do all this. Brown answered that ASC cost $611 million per year.

Messina stated that the goal was to involve applied math from the very beginning. Discrete math was included as a small user, but it may be large. The community needs to give DOE some topics with which to build a program if a program comes about.

A break was declared at 3:46 p.m. The meeting was called back into session at 4:03 p.m. and **Robert Rosner** was asked to report on the ASCAC Subcommittee on Exascale Modeling and Simulation.

The Subcommittee decided not to go discipline-by-discipline; that is what previous exascale workshop reports had done already. It decided to use the workshop reports as inputs to its report and to focus on what could be done that would change a discipline or an industry. Bigger may not be better; it may not lead to understanding. The Subcommittee decided to identify the roadblocks common to all applications, from programming models to scalable algorithms to unstable architectures, asking what needed to be worked on in order to succeed in getting to the exascale and how does what is now known about the hardware affect the prospects. It decided to think broadly.

The Interim report has an Introduction, considers the big questions that can be answered by exascale computing, and identifies the big obstacles and what needs to be done to resolve them. The Subcommittee still has to cover biology, health sciences, and disciplines that focus on extremely large data sets.

Experiments are expensive, and time to market is shortened by modeling and simulation. These considerations alone justify exascale computing. Experience has shown that fundamental products can be optimized, designed, and manufactured with no developmental testing because of modeling and simulation.

Case studies show that HPC is essential because some problems cannot be solved any other way. Modeling and simulation produce quantitative prediction. They are game-changing; either there was a scientific and/or technical discovery or they were faster making exploration feasible.

The workshops made the case for specific exascale applications. They have examined the role of extreme-scale computing in many areas. A real case for the exascale producing added value has been established in a handful of areas. Compelling societal, economic, and scientific benefits are associated with a few areas. In the end, these are things that lead to societal benefits, industrial benefits, or scientific discovery.

Exascale computing could enable the following advances in climate: understand water and carbon cycles; in combustion: a 30% efficiency gain in combustion engines; and in nuclear reactors: high-burnup (low-waste), proliferation-resistant designs. Some of these solutions have an element of timeliness to them.

Exascale computing can enhance U.S. leadership in key technology areas like information technologies, materials science and engineering, and systems-level engineering.

The following disciplines are exascale-ready: climate, combustion, reactor design, weapons, and science. In climate change, the transformational exascale capability is robust climate models for early warning, adaptation, and mitigation. In combustion, it is a 30% increase in engine efficiency while meeting stringent emissions standards. In nuclear-reactor modeling, it is high-fidelity, robust, and well-validated thermal hydraulics, neutronics and structure modeling tools; fully coupled thermal hydraulics, neutronics, and structure analyses; and predictive materials analysis tools for fuels, cladding, reactor vessel welds, all to undergird increased fuel utilization, power uprates, and reactor life extensions. In core-collapse supernovae, it is physically realistic 3-D supernova models. In stockpile stewardship, it is to make confidence in the national nuclear weapons stockpile high in the face of no future nuclear weapons tests, sharply reduced numbers of on-duty weapons, and the certainty that there is no known alternative to stockpile certification via simulations. Advances via exascale computing may occur in carbon capture and storage, the national electric grid, gas turbines, and chemistry.

There is considerable urgency in global security, meeting the mandates of the Clean Energy and Security Act of 2009, and American competitiveness and domestic jobs.

Each generation of architecture exposes new challenges and opportunities. When those challenges are overcome, there is a flow-down through the rest of the industry. There are obstacles/barriers to getting to the exascale. The consensus is that the requisite programming models and environments do not yet exist but require early investment. Candidate programming models are available. Model and algorithm development is needed to both support new architectures and take advantage of them.

It has always been true that machines do not do the same thing every time. That uncertainty has rarely been accounted for. There are inherent errors in the design of the machine.

Export control will matter but will conflict with international collaboration. Advances needed for extreme-scale computing require innovation and persistent development efforts by top-level research groups around the world. Coordination and open sharing of results among these efforts is critical to engender benefits to DOE. DOE must facilitate these interactions by appropriate interpretation of export control laws.

The workforce experienced in the requisite software underpinnings does not exist at the scale needed for the transition to the exascale; expertise is needed to develop codes and use codes. At the entry to college, we are not getting the right people. While we are waiting for the exascale, we should start ongoing workforce development now.

Giles noted that the Committee needed to transmit this preliminary report on the following day. The Committee needed to feed information back to the Subcommittee on what it has done so far. Rosner said that the Subcommittee will have one more meeting and finish the report.

Hack commented that the conclusions are what the Committee expected and asked if the premise was that exascale has to be justified with engineering problems not science problems. Rosner replied, no. It is just not clear that climate models can be calibrated to solve the problem. Years ago the community looked at seven codes to see if they could calculate a solution for which experimental data were available. One was way off. The other 6 got the same answer but not in line with experimental data. The question is whether model intercomparisons can ever lead to a certain solution.

Hack asked what "exascale-ready" was. Rosner answered: Can you make significant advances that fundamentally change the field?

Bailey asked what the strategy was to accomplish exascale computing, as called for in the charge letter. Rosner stated that the first three talks addressed that issue, and he did not want to repeat them. Bailey asked whether this was not a standalone report. Giles answered that the report presents co-design and other strategies but not with a timeline. Rosner said that what will be done depends on funding and feasibility. Bailey urged that the final plan have a strategy. Rosner said that the Subcommittee wanted to ensure that there are applications to justify exascale computing and also to list the hardware and software capabilities. Giles suggested that he outline the strategy. It is needed.

Bailey added that an advocacy stand would strengthen this report, and it could point to the success of petascale computing. He noted that Alex Lazalere at the previous meeting had shown that the modeling of nuclear fuel had allowed current nuclear power plants to produce more power. People will accept that the physics can be improved in future projects because it has been done in the past. Rosner noted that the efficacy of wire-wrap in fuel bundles is true, but it has not been implemented yet.

Berger said that the examples are from basic research, but there is not any discussion of where that fits in to the exascale program. Rosner responded that there are some problems that will come up again and again; the algorithmic theory is not there. New types of calculations are needed, even independent of the exascale. Giles noted that this would be in the broad strategy for the exascale. Part of the strategy would be a component of investment in basic algorithms for which there is a potential connection to solving exascale problems, but where the connections did not have to be demonstrated in advance as directly.

Tang stated that the report has to convince the experimental community that this is not a sandbox exercise. Emphasis has to be put on examples from past experiences. This whole area of validation brings one into the real world, so the report should tell what is going to be done in that area. Rosner replied that the scientific community is facing a number of cases where what is being confronted are questions that cannot be addressed experimentally. But in some there is a sociological problem. Computational people need experimental data on subcomponent behavior to validate and inform application codes.

Manteuffel stated that the Graduate Fellowship Program is nowhere near the effort needed to solve this problem. The report should stress this need.

Hey suggested that a list of what should be done first, second, etc. would be helpful in the report.

Chen urged making the whole story for funding; DOE is not the only party that needs this capability. Rosner agreed that partners are needed to fund these efforts.

Bailey suggested that large-parameter problems may be low-hanging fruit to be picked.

Habib said that one problem is our own success. There already is middleware. An appreciation for creating code needs to be created. Also, in the exascale scenario, code will have to be matched to the machine, which may not be permanent. One will need new tools to build code with. Some type of block-code approach might succeed. Retrieval of archived data might be a problem, also.

Rosner said that academics may compromise performance so they can run their codes on different architectures. Building a model with architectural flexibility is a huge concern. Berger objected that that goes against the advice to get efficiency by marrying the code to a machine.

Messina said that, with co-design, if the hardware people are allowed to design the machine, they will not give the programmers any memory. The algorithm people have to be able to push back. They need to influence the architecture. Portable code is great, but some people will tailor their code to an architecture. There is no one size fits all. A strategy is needed to make co-design work.

Bailey said that the strategy should look at what was done in the past. The ASCI strategy emerged from a decade of experience. A. White noted that collocation helped a lot. Bailey cautioned that interaction with industry and universities has to be part of the strategy as well as involving NSF and DARPA.

The meeting was adjourned at 5:40 p.m. Giles asked that the Committee immediately help write a transmittal letter for the report.

Wednesday, March 31, 2010
Morning Session

The meeting was called to order at 8:37 a.m. **Steven Koonin** was asked to present an update on the Department's activities.

The administration has put some duties on DOE:

- Nuclear security (diminishing the risks of nuclear proliferation)
- Energy (catalyze growth of energy systems)
- Discovery science
- Enhance U.S. competitiveness and create jobs

Some of these are new. A strategic planning exercise has been going on for months preparing for FY12 budget request. The shared vision of this planning will be available soon.

The Department is trying to pull basic research and applications closer together. We have to open up the National Nuclear Security Administration (NNSA) in some ways to infuse science into it and to apply some of its crown jewels (e.g., simulation capabilities) to other portions of the Department. We are trying to connect the stovepipes. Computing is large in our plans.

Rosner's charts lay out a good example of the need for the exascale. DOE has been discussing with the private sector what the hardware and software might look like and how the technology might be developed. There may be a push toward exascale computing.

In energy-systems simulation, ASCR has done a good job in the development and deployment of advanced systems. Validation science has also developed. Many applications have benefitted: nuclear reactors, carbon capture and storage, and internal combustion engines. The private sector needs to be involved in these efforts. The Department has been integrating HPC into industry.

Berger picked up on his statement about better communication between stovepipes, and asked if there were any specific plans or suggestions. Koonin replied that the people in the Department all talk and work together at different science levels; hold joint workshops; and have budgets and personnel that stretch across stovepipes.

The private sector is goal focused and command driven. Universities have diffuse goals and are consensus driven. Government has the worst of both worlds: it is consensus driven and goal focused.

Giles asked Koonin for his thoughts on how the Department interacts with the community. Koonin responded that every program office has an advisory committee. These function well and are effective. In applied areas, it gets spottier: the community is heterogeneous; external advice is informal and episodic. The focus of the Secretary is to improve the quality of external input and peer reviews.

Manteuffel asked what was happening on job creation? Koonin said that the DOE budget has loan guarantee programs to catalyze firms in energy sector (e.g., the nuclear loan guarantee and the Nissan Leaf guarantee). The Small Business Innovative Research (SBIR) and Small Business Technology Transfer (STTR) programs catalyze small business. How we get it all together is still a work in progress. Innovation is glacial in speed.

Manteuffel asked about the scientist pipeline? Koonin stated that one can change universities by orthodontia: apply long-term pressure. Many university presidents come to DOE to ask what they can do. Focus on energy problems in how you teach students. There are really interesting science problems in almost any DOE field.

Tang asked about alliances with NSF. Koonin noted that DOE talks with NSF extensively. The agencies have different but congruent missions in engineering education, HPC, climate modeling, and DUSEL. The agencies are in dialogue. Another agency DOE works with extensively is the U.S. Department of Agriculture (USDA) on biomass.

Chen noted that European countries have framework programs and asked if there was a move in that direction in the United States. Koonin said that it is a tactical call: Do you want broad contact between stovepipes or discrete topics? Both are needed. Certain topics could have an impact in 5 years.

Hey pointed out the increasing importance of data-driven experiments. An open-science-data agenda is important. Koonin replied that, on the openness of data, the United States should have as much as it can get, especially in climate. He did not know DOE's policy on data sharing but will find out. The Office of Science and Technology Policy (OSTP) has published something on data sharing, also. DOE deals with big data sets in HEP; in DOE, energy-system modeling and the Energy Information Administration

(EIA) would benefit from more data sharing. Those outside DOE would have a larger interest in data sharing, and DOE should talk with them. The technology of the exascale will probably force more sharing.

Bailey noted that one often refers to the valley of death and asked Koonin what his thoughts were on that problem. The relationship between the National Advisory Committee on Aeronautics (NACA) and government has been fruitful. Such a relationship on nuclear energy might also be helpful. Koonin responded that those relationships carried over into space enterprises, but not elsewhere. There are deep philosophical divides in Washington about how much government should be picking winners and losers. Loan guarantees are important, but should government be a venture capitalist? It is a good subject for debate.

Dongarra asked if basic research in DOE were going to go toward the exascale to the exclusion of other research directions. Koonin answered by asking, do we need basic R&D in the exascale? Giles answered, yes, in algorithms, architecture, etc. One has to do it well ahead of deployment. Koonin noted that it was a balance issue. One should not swing all the way in one direction.

Hey said that there was a lot of technical discussion on the exascale, but cost is not mentioned. He had heard $2 billion for hardware and reports of $10 billion for software. He asked if the government were going to front this money. Koonin said that he was skeptical of the numbers. Those numbers are an order of magnitude larger than what he had heard. A funding profile is needed.

Giles opened the floor to discussion of the exascale report. He distributed a rough draft of a cover letter in which additional information that the Subcommittee will consider is reviewed. This will obviate having to have the Subcommittee rewrite the preliminary report before acceptance by the Committee. He asked for suggested additions to the cover letter. [The final cover letter is included as an appendix to these minutes.]

Hack commented that there is not great value in distinguishing between exascale ready and almost exascale ready. Giles replied that the concern was that using a broader brush might reduce impact. Tang said that he was sympathetic to Hack’s comment. These arguments must demonstrate quite credibly that these applications will have significant societal impact.

Chen said that the report needs to highlight more holistic approaches and elements that might not be funded by the exascale initiative but would occur to support or respond to it.

Bailey said that the report should point to where all this is going and how it is going to get into the hands of the users. Transfer of technology and tools from DOE to industry is very important in energy systems. Part of the strategy should be a statement of the endgame.

Rosner replied that the Subcommittee could meet Hack’s objective by being silent about the application area where there is no consensus.

Hey noted that interactions with industry are important in co-design.

Manteuffel asked if this Committee would interact with the Subcommittee anymore. Giles responded that the Committee will review the final report. The Subcommittee could have an open meeting that the Committee members could attend. There are three ASCAC members on the Subcommittee. Rosner added that anyone who is interested should feel free to attend the Subcommittee meetings. The rewrite has to be done by June. Giles stated that the Committee might consider having a June meeting.

Strayer stated that he was impressed with the work of the Subcommittee. It has struggled with the basic issue. He was delighted with the engagement of the Committee with the Subcommittee. He thanked the Subcommittee for it's hard work.

Harriet Kung was asked to review the role of advanced computing in the Office of Basic Energy Sciences (BES).

The mission of the BES program is to support fundamental research to understand, predict, and ultimately control matter and energy at the electronic, atomic, and molecular levels. The priorities of BES are to discover and design new materials and molecular assemblies with novel structures, functions, and properties (high-temperature superconducting materials, buckyballs, etc.); to conceptualize, calculate, and predict processes underlying physical and chemical transformations, especially those that display emergent, coherent phenomena; to probe, understand, and control the interactions of phonons, photons, electrons, and ions with matter; to conceive, plan, design, construct, and operate scientific user facilities to probe fundamental electronic and atomic properties of materials at extreme limits of time and space; and to foster integration of the basic research conducted in the program with research in NNSA and the DOE technology programs.

The office is managed through three divisions: Materials Sciences and Engineering; Scientific User Facilities; and Chemical Sciences, Geosciences, and Biosciences. It interfaces with BER and ASCR.

The Materials Sciences and Engineering Division is made up of three teams: Condensed Matter and Materials Physics; Scattering and Instrumentation Sciences; and Materials Discovery, Design, and Synthesis. The Chemical Sciences, Geosciences, and Biosciences Division also has three teams: Photo- and Biochemistry, Chemical Transformations, and Fundamental Interactions.

BES uses several modalities of funding. It started with support for single-investigator and small-group projects. It found that the field self-assembled into cross-disciplinary teams, so it funded the EFRCs, which are $2- to 5-million-per-year research centers (for five years), established to accelerate the scientific breakthroughs needed to create advanced energy technologies. They are still in the scientific-discovery domain. Now, BES is also establishing Energy Innovation Hubs to integrate science push and technology pull and to shift expertise to address the most pressing issues. They are funded at $20 million+ per year for fuels from sunlight and electrical batteries and storage.

BES scientific user facilities are parallel to the ASCR user facilities. Stronger cooperation is needed to deal with data reduction and analysis. BES supports five light sources, three neutron sources, three electron-beam sources, and five nanoscale-science research centers. These facilities have computing capacity built into them. The synchrotron light sources support 8000 users annually.

A unified theme of BES is pushing the ultra-small and ultra-fast. In the past decades, it has dropped to the nanoscale in size and to the femtoscale in time.

In doing so, BESAC has issued many planning documents. One has to establish a framework of grand challenges to pull together and inspire the scientific community. So BESAC was charged to identify such grand challenges. These documents have made the transition from the 20th Century science of observation to the 21st Century science of control, producing designer catalysts and superconducting materials.

In addition to grand-challenge science, BES has systematically reached out to the applied-science side in hydrogen, solar, superconductivity, solid-state lighting, and advanced nuclear energy systems. Twelve workshops later, it sees the future as a transition from energy as a commodity to energy as a renewable resource. It sees its task as moving grand challenges to discovery and use-inspired basic research so they can be moved on to applied research and to technology maturation and deployment. The need for better tools, especially computational tools, has been a constant realization. BES looks to a continued and expanded partnership with ASCR.

The EFRCs came about as a new funding modality to support multidisciplinary energy research. It now funds 46 centers with $777 million per year. Most are led by universities, some by national laboratories, and one by industry. They addressed many energy-science topics, such as understanding radiation resistance in materials, an old field in which many new insights are being gained with new tools.

A number of SciDAC awards have been made jointly by BES and ASCR.

The Science Base for Multiscale Simulation of Internal Combustion Engines is a new initiative in FY11. One can get large technical gains by understanding the underlying basic science. This technology is facing very rapid changes (e.g., in fuel chemistry as the shift is made from petroleum to advanced renewable fuels).

BES held a workshop on basic research needs for clean and efficient combustion of 21st century transportation fuels. It resulted in a new BES activity to consider models that span vast scale ranges, improved understanding of fundamental physical and chemical properties, and engine simulation. This activity is pulling together basic research with applied technology and industry.

Bailey asked how one would get car companies to use these codes. Kung replied that the companies are being engaged in the workshops to see what they can use and then directing BES's research toward those goals.

Giles asked how one integrates computing into experimental research. Kung answered that that is the most exciting aspect of BES's partnership with ASCR. Data gathering needs to be integrated with instrument design as the user facility is designed. Who leads and who follows will vary with the field. Experiments provide validity to data and modeling.

Tang stated that this is a tremendous opportunity. These are world-leading facilities. Validation and verification is important to ASCR, and stronger links with BES offer excellent opportunities.

Bailey said that an excellent way to interact with industry is to share large-scale facilities with them. It is a draw to industry to pay attention to what DOE is doing. Kung noted a recent study that pointed to the user facilities as loci of bringing researchers together with industry. BES needs reviewers with the capability to judge these advanced-technology proposals.

Chen asked if there are linkages between core researchers and the hubs. Kung answered that these look very duplicated, but the core is basic science, and the hubs are more applied technology. The different performers are brought together in meetings to promote further exchange between fundamental science and industry. Other mechanisms will be sought to improve that interaction.

A break was declared at 10:23 a.m. The meeting was called back in session at 11:45 a.m. to hear **Arun Majumdar** describe the Advanced Research Projects Agency–Energy (ARPA-E) initiative. ARPA-E was founded in response to the recommendations of the *Gathering Storm* report of the National Academy of Sciences (NAS).

DARPA was launched in 1958 in response to Sputnik. The United States is undergoing a Sputnik moment right now in energy research, greenhouse gas emissions, and U.S. technological leadership.

The Earth has a huge population whose energy use is very low now but is growing rapidly. The average CO_2 production per capita for the world is 5 tons per year; the U.S. production is 20 tons per year. In China, they are trying to take a low-intensity energy trajectory.

There have been several such challenges before, such as the 1898 call upon science to save the world from an impending starvation. Then there were guano wars for fertilizer to boost agricultural production. In 1908, Fritz Haber discovered a catalyst that would combine atmospheric nitrogen with hydrogen to form ammonia; and in 1913, Carl Bosch developed a process to mass produce ammonia and made fertilizers. After that, population expanded rapidly.

Today, the U.S. is falling behind in the clean-energy race. In worldwide shipments of solar cells, the U.S. market share has been declining since 1995 and is now less than 10%. The United States' research infrastructure is amazing; its great strength is its intellectual capabilities.

Game changers during the 20th Century were artificial fertilizers, the green revolution, polio vaccination, antibiotics, airplanes, electrification, nuclear energy, transistors, integrated circuits, fiber-optic communication, wireless communication, and the Internet. Imagine all of this happening in the next 20 years:

- Solar electricity generation at a cost lower than that of fossil fuels ($1/W fully installed)
- Carbon capture and use at a net cost lower than its market price
- Low-cost desalination of water
- Real-time optimization, security, and storage for the grid

ARPA-E was authorized in 2007 as part of the America Competes Act. Its first budget was included in the American Recovery and Reinvestment Act of 2009 (ARRA).

In the first round of funding, 3700 papers were received, 312 full applications were encouraged, full applications were panel reviewed, and 37 projects were funded for 2 to 3 years. The contracting process of DOE was changed to process awards in 3.5 months.

ARPA-E is looking for high impact on ARPA-E mission areas; disruptive, innovative technical approaches; strong impact of ARPA-E funding relative to the private sector; and best-in-class people and teams.

A few ideas from the first round of funding:

- Artificial cellulose breakdown is expensive, so Agrivida is taking known genes and putting them in the plant in a blocked status to be triggered upon harvest
- A breakthrough high-efficiency mixer/ejector wind turbine from FloDesign Wind Turbine Corp.
- Grid-level electricity storage from the Massachusetts Institute of Technology

The second round of awards is considering three technologies. The first technology is batteries for electrical energy storage for transportation. Lithium-ion batteries have a high cost and a low energy density. Targets set include a cell-level energy density of 400 W-h/kg and a cost of $250/kWh. Metal–air batteries and Li–S batteries may reach these targets but are hard to exploit. The second technology is innovative materials and processes for advanced carbon-capture technologies. The hope is to lower the cost of capturing CO_2 from the current $70 or $100 per ton to less that $30 per ton. Enzymes may be usable to accomplish this reduction. The third technology is electrofuels to produce “gasoline” from CO_2 and hydrogen with off-peak electricity.

The third round of funding is considering transformational approaches to energy storage that enable grid-scale deployment at very low cost (about $100/kWh); cutting building cooling energy consumption and greenhouse-gas emissions by 25 to 40%; advancements in power-electronics materials coupled with advanced circuit architectures and scalable manufacturing processes to result in low-cost, higher-performance power electronics across many applications.

The ARPA-E organization is lean, nimble, collaborative, and flat. It reports directly to the Secretary and coordinates closely with SC and the applied-technology offices. Its objectives are to break down stovepipes, encourage debate and partnership between technology pushers and pullers, and provide thoughtful leadership to create new programs. It has the Panel of Senior Technical Advisors (PASTA), a coordinating council of leaders from across DOE.

The United States needs to build on its strengths, which are

- The best R&D infrastructure in the world;
- The best innovation ecosystem in business and entrepreneurship; and
- Highly energized youth, ready to deeply engage.

The ARPA-E Fellows Program has been created to bring best and brightest scientists, engineers, and technical entrepreneurs into ARPA-E and create a think tank. It is also recruiting program directors, active researchers who have one foot in science and engineering and the other in technology development and business.

In 10 years, it is hoped to see increased domestic and global sales and U.S. market share; avoided greenhouse-gas emissions; reduced oil imports; the creation of a new technology/business or new industry ecosystem; jobs; and a besting of current projections and trajectories. In 3 to 5 years, it is hoped to see follow-on investment post ARPA-E award, an increase in the enterprise value of companies, companies being created, an initiation of new technology-business ecosystems, accelerated market entry of new products and product sales, patents filed and licensed, papers published in top journals, world-record-setting performances, and the identification of mechanisms for scaling innovations.

The diversity of science is being addressed through a coordinated approach, including EFRCs and Energy Innovation Hubs along with ARPA-E. This diversity and multiplicity of approaches is needed (with competition). Historically, change is slow, energy is a ubiquitous commodity, and investments and systems can last a long time; but for air conditioners and batteries, change can occur more rapidly.

The ARPA-E Energy Innovation Summit was held in early March to bring together scientist/engineers from academia, national laboratories, and industry; investors; small/large industry senior management; policy groups; and Congress to discuss: How do we foster and identify game changers? Is it random or is there a system? How do we go from lab to market with disruptive energy technologies that challenge business as usual?

How do we scale innovations in the United States? How do we accelerate the pace? How do we balance global competitiveness and partnerships? How do we ensure national security through energy technologies? And how can DOE play a role in energy innovation? There were more than 1700 attendees from 49 states and 15 foreign countries. Twelve national laboratories and dozens of universities participated. Financial deals were worked out on the floor. Next year's summit is now being planned.

Giles asked Majumdar if he had a sense of how to be risk tolerant over time and not be conservatized. Majumdar responded that one has to build a culture of taking risks; ARPA-E is writing a strategic plan. It also depends on who is running the organization.

Bailey noted that, in DARPA, the Department of Defense (DoD) is the customer. It picks winners and losers. He asked how ARPA-E would avoid such criticism. Majumdar replied that one needs to put up targets. What is being seen are multiple approaches being supported, and the private sector is being allowed to develop the best technology. The energy sector is an open economy, and DARPA's is a closed economy. With ARPA-E, there are many customers, including DoD. One strategy that is being discussed is using the purchasing power of the U.S. government to support the marketing of leapfrog technologies.

Manteuffel asked how ASCR can contribute. Majumdar replied that he was coordinating very closely with all DOE offices. ARPA-E needs a lot of tools to design energy-efficient systems. Materials issues exist in many technologies. How about a genome for materials? To build a pilot plant is expensive, so systems integration by computer models could be extremely useful. Manteuffel suggested that part of the award could be computer time allocation from other parts of DOE.

Linda Petzold was asked to report on the progress of the applied mathematics COV. The visit has not occurred yet. The COV membership has been set. May 12 and 13 are the dates of the visit. Giles asked if the COV report could be ready in time for a June ASCAC meeting. Petzold replied that it could if it were held in the latter part of June.

Giles opened a discussion on the report of the Subcommittee on Exascale Modeling and Simulation. Rosner reported that he had added a slide on combustion and that two other slides are being deleted or modified. Giles said that that seems to satisfy all the recommendations of the Committee. He asked for a vote on accepting the report and transmitting it to Brinkman and Koonin. The vote was unanimously positive.

Bailey voiced support for a meeting before the regularly scheduled August meeting.

Manteuffel asked if the Subcommittee meeting will be announced to all Committee members. Rosner replied, definitely. Giles said that it could be immediately before the next ASCAC meeting.

The floor was opened for public comment. There being none, the meeting was adjourned at 1:47 p.m.

Respectfully submitted,
Frederick M. O'Hara, Jr.
Recording Secretary
April 27, 2010