Final Minutes
Advanced Scientific Computing Advisory Committee
March 5, 2013
American Geophysical Union, Washington, D.C.

ASCAC members present:

Marjory S. Blumenthal
Vincent Chan
Barbara M.P. Chapman (via telephone)
Jacqueline Chen
Roscoe C. Giles (Chair)
Susan L. Graham
Gwendolyn L. Huntoon
Juan Meza
John Negele (via telephone)
Linda R. Petzold (via telephone)
Vivek Sarkar
Victoria White (via telephone)
Dean N. Williams

ASCAC members absent:

Marsha Berger
Vinton G. Cerf
Jack J. Dongarra
Sharon C. Glotzer
Anthony Hey

Also participating:

Melea Baker, Administrative Specialist, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Linda Biven, Senior Science and Technology Advisor, Office of the Deputy Director for Science Programs, Office of Science, USDOE
Arthur Bland, Project Director. Oak Ridge Leadership Computing Facility, Oak Ridge National Laboratory
William Brinkman, Director, Office of Science, USDOE
Keri Cagle, Oak Ridge Institute of Science and Education
Richard Carlson, Distributed Network Environment Program, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Christine Chalk, ASCAC Designated Federal Officer, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Vincent Dattoria, Computer Scientist, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Barbara Helland, Acting Associate Director, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Mohammad Khaleel, Director, Computational Sciences and Mathematics Division, Pacific Northwest National Laboratory
Alexandra Landsberg, Applied Mathematics Program, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Lucy Nowell, Program Manager, Office of Advanced Scientific Computing Research, Office of Science, USDOE
Frederick O'Hara, ASCAC Recording Secretary
Brian Plesser, Attorney-Adviser, Office of the General Counsel, USDOE
Ceren Susut, Partnerships Program, Office of Advanced Scientific Computing Research, Office of Science, USDOE

Julia White, INCITE Manager, Oak Ridge National Laboratory
Victor Zavala, Mathematics and Computer Science Division, Argonne National Laboratory

About 40 others were in attendance.

Originally scheduled to last two days, the meeting was foreshortened to one day because of an impending storm that closed government offices and airports along the northeast coastline.

Before the meeting, the Committee members were briefed on ethics by **Brian Plesser** of the DOE Office of the General Counsel.

The meeting was called to order by **Roscoe Giles**, Chairman, at 8:30 a.m. **Christine Chalk**, Designated Federal Officer of ASCAC, announced that the meeting was being foreshortened to one day because of weather. **Keri Cagle** of the Oak Ridge Institute of Science and Education (ORISE) made safety and convenience announcements.

Barbara Helland was introduced to give an update on the activities of the Office of Advanced Scientific Computing Research (ASCR).

The Office is operating under a continuing resolution; it received 47% of its money for the year. Facilities got 50% of the money that they would need for the year. All mortgages were paid at the first of the year. The FY12 enacted budget (less "taxes") was $441 million; the annualized FY13 continuing resolution budget is $444 million; and the FY13 budget with sequestration is $418 million. As a result of the sequestration, there is no funding for a second request for proposals (RFP) for FastForward; there are delays in research funding to university grants, impacting as many as 60 graduate students; and there is no planning for power upgrades. The Office of Science (SC) amount for the sequestration is $245 million. This is a 5% reduction to the annualized continuing-resolution level of $4,904 million. With this funding, ASCR provides high-performance production computing for SC, leadership computing for open science, ESnet, and research and engineering (R&E) prototypes.

The leadership computing facilities (LCFs) are allocated to the open-science community. In the 2013 call for proposals, a request for information helped attract new projects. The call closed on June 27, 2012. About 14 billion core-hours were requested, nearly three times the amount requested in the previous year. The 143 proposals submitted represented an increase of nearly 20% from the previous year. Awards of about 5 billion core hours will be announced in November for CY13. Nearly 50% of the non-renewal proposals are by new principal investigators (PIs).

The Titan hybrid system at Oak Ridge National Laboratory (ORNL) has been installed and is undergoing acceptance testing. It has passed its functionality and performance tests. A stability problem has required graphic processing unit (GPU) board repair at Cray. Users are on the system, running on the central processing units (CPUs). GPUs will be available to users in mid-March. It is expected that the acceptance testing will be completed in April or May after all boards are repaired.

The Argonne Leadership Computing Facility has the Mira computer, an IBM Blue Gene/Q system with 49,152 nodes and 786,432 cores. It has 786 TB of memory and a peak flop rate of 10 PF. Its Linpack flop rate is 8.1 PF.

Users were quickly up and running on Edison at the National Energy Research Scientific Computing Center (NERSC), which also has the 1+ PF Hopper Cray XE6. The first phase of NERSC-7 has been delivered; it will be the first Cray system with Intel processors. Its new Aries interconnect and dragonfly topology will enhance data transfer, which is a performance bottleneck for many codes. A high amount of memory per core will enable emerging data-centric applications. Users are off and running on the four-rack Phase 1 High-Productivity Computing Systems (HPCS) demonstration system. Phase 2 will be complete later this year and will bring Edison up to about 2 PF.

NERSC started giving an award for high-impact scientific achievement in 2013. It went to Jeff Grossman for a new approach to water desalination. The NERSC award for innovative use of high-performance computing went to Peter Nugent and the Palomar Transient Factory team for the project Data Pipeline Transfers, Analyzes, Stores, and Disseminates Astronomical Observations.

ESnet, the world's first continental 100-GB network moved to full production status last November. For that work, ESnet will receive two awards from the federal government: the Fierce 15 Award and the InformationWeek Top 15 Innovators for 2012.

As time progresses, more resources will be needed. NERSC and ESnet gather requirements directly from scientists through program requirement reviews and science case studies. Review meetings establish a consensus on requirements, capabilities, and services. Scientists, program offices, and facilities have the same conversation. Historical trends, technology advances, etc. are also incorporated into the assessment to provide a solid, fact-based foundation for service and capability investments that address DOE mission goals by ensuring DOE science is effectively supported.

Since the previous ASCAC meeting, Barbara Helland was named the facilities manager, and she initiated facilities strategic planning to inform the development of an ASCR facilities 10-year plan. The computer facilities were constrained to 30 MW of power. Facilities presented their plans to each other, to headquarters, and to the ASCAC Facilities Subcommittee in January. Findings from this exercise were that the Office needs to coordinate with other SC offices, to prepare applications for future architectures, and to develop a common strategy for addressing SC data analysis and archival needs arising from simulations and experiments. The next step will be to produce a draft ASCR 10-year facilities plan.

The overall acquisition strategy for the next systems (i.e., those of hundreds of petaflops and beyond) will employ joint RFPs between National Nuclear Security Administration (NNSA) and SC laboratories in acquisition strategy, core design, and architecture. This process will bring machines in more quickly. Only one RFP will be needed, and only one machine needs to be designed. Non-recurring engineering investment coupled with acquisitions will offer opportunities for software or technology investments to provide additional features and opportunities for variations through separate contracts.

The path to the exascale was started in 2010 when three funding opportunity announcements (FOAs) were issued for advanced architectures, the X-stack, and scientific data management and analysis. In FY11, three exascale co-design centers were funded, and a request for information was issued on critical and platform technologies. In FY12, programming environments and extreme-scale algorithms were considered, and the FastForward R&E prototype was started. In FY13, the operating system and runtime issues were addressed, the DesignForward R&E prototype was started, and uncertainty quantification (UQ) was studied. It is hoped to pull the Exascale Initiative's funding forward from its current start in mid-2014.

Huntoon asked where the scientific vulnerabilities were. Helland replied that they were in delaying the funding of the Exascale Initiative.

Co-design was started in 2011. Each exascale co-design center presented a 3-hr-long deep dive into the center's physics problems, models, equations being solved, and overall workflow, including the development of "proxy apps," which help answer co-design questions, such as

- What is the balance between serial and parallel parts of algorithms?
- How important is cache coherence if it costs significant additional power?
- If the programming model is message-passing interface (MPI) + X, what is X?
- Are apps more amenable to many small cores or to heterogeneous nodes?
- How much data must be moved, both intra- and inter-node?
- What is the relative importance of very-low-latency small messages vs. higher-bandwidth, higher-latency operations?
- Are there algorithms that emphasize fine-grain parallelism (strong scaling)?

These proxy apps must be custom-made for each question.

The FastForward projects are jointly funded by DOE SC and NNSA (1) to initiate partnerships with multiple companies to accelerate the transition of innovative research of critical technologies, such as processor and memory architecture, into products needed for extreme-scale computing and (2) to fund technologies targeted for productization in the 5- to 10-year timeframe. The funded vendors (AMD, IBM, Intel, NVIDIA, and WhamCloud) had to put in up to 40%.

The DesignForward RFP will be issued by Lawrence Berkeley National Laboratory (LBNL) on behalf of Argonne National Laboratory (ANL), Los Alamos National Laboratory (LANL), LBNL, Lawrence Livermore National Laboratory (LLNL), ORNL, Pacific Northwest National Laboratory (PNNL) and Sandia National Laboratories (SNL). It will look for interconnect networks and system design and integration. The RFP has not yet been released and could be delayed by the sequestration.

The goals and objectives of the X-Stack Programming Environment Program are to address fundamental exascale software challenges, create complete solutions that are portable across multiple future machine generations, define a transition path for existing scientific applications, and initiate an exascale community prior to establishment of the long-term exascale program. The key program elements are managing parallelism and data movement, establishing interoperability across different high-performance computing languages and interfaces, managing locality, minimizing energy consumption, and developing runtime systems that are dynamic and adaptive. Funding started last spring.

The Exascale Research Conference was held in October 2012. There were in-depth discussions among PIs on technical issues associated with proxy applications that helped participants gain a better understanding of the issues surrounding exascale co-design and provided opportunities for collaboration, leveraging of work, and the identification of new research directions. The conference had 209 attendees who presented 60 posters, 43 technical talks, 42 position papers, 15 co-design deep-dive talks, and 13 plenary presentations.

Four PI meetings were also held this past year.

In Applied Math, an FOA was issued for RX-solvers, applications for basic research in resilient extreme-scale solvers.

A mathematics committee of visitors (COV) is under way. An initial meeting was held with ASCR program managers on Feb 19, 2013, to discuss the charge and logistics. The COV will review the FOAs on uncertainty quantification (FY10), the Mathematical Multifaceted Integrated Capability Centers (MMICCs) (FY12), and laboratory programs (FY11-FY12). Committee recruitment is under way. It is expected that the COV will meet in June 2013 and issue a final report by late August 2013.

FY13 is the fourth year of the SC Early Career Research Program. ASCR received 75 of about 770 Early Career Award applications submitted to SC and anticipates making FY13 awards to three university PIs and two national laboratory PIs at the end of April.

Two ASCR researchers were named 2012 Association for Computing Machinery (ACM) fellows: Kathy Yelik and Robert Schreiber. In addition, Paul Fischer was elected an American Association for the Advancement of Science (AAAS) fellow, John Bell was selected as a member of the National Academy of Sciences, Pavan Balaji was recognized as one of Crain's Chicago Business 40 Under 40, Linda Petzold was awarded the 2013 Society for Industrial and Applied Mathematics/ACM prize in computational science and engineering, and Marc Snir will be presented the 2013 Institute of Electrical and Electronic Engineers Technical Committee on Scalable Computing Award.

No appointments will be made until there is a secretary confirmed by the Senate. Interviews for the associate director (AD) position are taking place now. There are a couple of Mathematics and Computer Science positions open. In Facilities, Carolyn Lauzen is working in the Office as a AAAS Fellow, Dave Goodwin is the NERSC program manager, and Ravi Kapoor is the Oak Ridge LCF program manager.

Giles asked if there were any effect on the ability to attract and retain personnel because of the continuing resolution and sequestration. Helland replied, in the facilities, yes; industry is hiring away the personnel. People at the national laboratories have not gotten raises for 2 years.

Chan asked if discussions with other SC offices had affected the strategic plan. Helland replied that those discussions are going well, but a better way to assess data needs is needed; the data plan is still very fluid. There will be more meetings with other offices in April.

Giles observed that progress is being made on the exascale, but funding for a robust program is not there. Helland replied that the Senate staff needs concrete examples of how the exascale will affect jobs and college programs. Only anecdotes are available. The Office is working with the Council on Competitiveness and International Data Corp. (IDC) to get data on this topic.

Vivek Sarkar was asked to present the report from the Data Subcommittee.

Computing is at the juncture of two major challenges: big data and the exascale. The charge to ASCAC was to: "examine the potential synergies between the challenges of data-intensive science and exascale. The subcommittee should take into account the Department's mission needs, which define the Office of Science's role in data-intensive science vis-à-vis other agencies. The subcommittee should specifically address what investments are most likely to positively impact both our exascale goals and our data-intensive science."

There are several federal and commercial initiatives under way to address the challenges of big data. The scope of this study was restricted to the intersection of big-data challenges and exascale computing in the context of data-intensive science.

The charge did not state a time frame, so the time scale assumed was about 2022 (when exascale capability is expected). Nearer-term considerations also received attention because they will influence the migration path to exascale computing. The charge asked for identification of investments that are most likely to positively impact both data-intensive-science research goals and exascale-computing goals. Because facilities are the focus of another ASCAC study, this study focused on investments that can leverage synergies between data-intensive science and exascale computing rather than on facilities.

The context of the charge included several recent ASCR workshops have focused on exascale and data-intensive computing, and the Subcommittee leveraged that information.

A fourth paradigm of science has emerged beyond theory, experiments, and simulations. Data-driven science leads to the use of automated analysis in massive datasets to drive scientific discovery. Challenges arise from the increasing velocity, heterogeneity, and volume of data generation.

The overall trend is that most science domains will become data-intensive in the exascale timeframe. Costs of memory have been going up at the rate of about 20% per year, of processors at about 36%, of sequencers at about 60%, and of detectors at about 72%. Most science domains are going to become data intensive. A notable such domain is high-energy physics. The ATLAS [A Toroidal LHC ApparatuS] and CMS [Compact Muon Spectrometer] detectors produce analog data at rates equivalent to 1 PB/s, which is several orders of magnitude greater than one can collect now. The output rate after data reduction is 1 GB/s or about 10 PB/year. How does this fit into researchers' workflow? One benefit is that a homogeneous community of physicists' access read-only shared data with the Worldwide LHC [Large Hadron Collider] Computing Grid (WLCG). There are three data challenges in climate science: distributing live data streams and large-volume data quickly and efficiently, analyzing large-volume data in place for big-data analytics, and producing on-demand data products for heterogeneous communities. This community has built the Earth System Grid Federation (ESGF), which manages several petabytes of data.

The S3D combustion code, when projected to the exascale, will require 3 PB of memory and will produce 400 PB of raw data (1 PB every 30 min). Workflow challenges include co-design for simulation and in situ analyses.

In biology and genomics, the data-intensive challenges include biophysical simulations of cellular environments, cracking the "signaling code" of the genome across the tree of life, and reverse engineering the human brain. The KnowledgeBase (KBase) center currently manages about 2 PB of data for genomics research; the workflow is based on a service-based infrastructure. There are significant differences between the data characteristics in KBase and other domains (lots of integer data, random access, large intermediate data size during computations, and poor locality in cross-correlation).

In light sources [e.g., the Advanced Photon Source (APS) and Linac Coherent Light Source (LCLS)] the data production rates will be magnified in the future, and they have a dispersed community. The APS includes about 65 beam lines, with about 1 TB of data produced per day. Future light sources are expected to produce data at the rate of 1 TB/s. GridFTP and GlobusOnline services help some APS users with their workflow, but many others bring their own storage devices and perform manual analysis of their data. The LCLS provides users access to about 2.5 PB of storage via the LCLS portal, where data are stored for 2 years, and to an on-line cache of about 50 TB, where data are stored for 5 days. These volumes are expected to increase dramatically in the future. The researchers are limited by the workflow, not storage.

The trend is toward greater data rates. In an information-rich world, analysis is the limiting resource. There is a widening gap between input/output and computational rates will make in situ analysis and visualization a necessity for the exascale. Visualization will guide scientific discovery. Visualization is used to grasp structure also, impacting the workload of the exascale machine (doing analysis as well as simulation). The focus is how one derives value from information overload.

Data streaming and near-sensor computing must be used to deal with the problem of massive data rates. At the Spallation Neutron Source (SNS), the challenge is that the reduction and visualization of some of the large data sets take hours after the data have been collected. Collection is dealt with by the Accelerating Data Acquisition, Reduction, and Analysis (ADARA) Collaboration's streaming data system that provides in situ reduction of data as it is produced by the instrument. Challenges in in situ reduction are synergistic with data-movement challenges in exascale computing. For near-sensor computing, triggers detect events of interest to be recorded, filters reduce data as close to the instrument as possible, and the data are curated and archived for reprocessing and re-analysis.

Many requirements for big data and the exascale are intertwined. Big data produced by the data-driven paradigm will need to be analyzed by Big Compute (exascale or extreme-scale) systems. "Extreme-scale systems" refers to all classes of systems built with exascale technologies that deal with power usage, data movement, etc.

Big Compute will produce Big Data. Data-intensive simulations on exascale systems will produce data volumes comparable to the data produced by the largest science experiments. Data-driven and data-intensive approaches have evolved somewhat independently of each other. It is important for each to learn lessons from the other because their fates are intertwined.

The cross-cutting issues identified were dealing with the data lifecycle (retention, preservation, and sharing of data); software challenges; technology disruptions (e.g., new analysis algorithms); provenance, metadata, security, and privacy; and the expertise and skills gap.

The findings of the Subcommittee are:

1. There are opportunities for investments that can benefit both data-intensive science and exascale computing. Data-intensive science relies on the collection, analysis, and management of massive volumes of data, which will have to be performed by exascale systems or, more generally, by "extreme-scale" components of exascale systems. Extreme-scale components will include innovative memory hierarchies and data movement optimizations that will be essential for the analysis components of data-intensive science workflows. There is also synergy between algorithms for near-sensor computing in experimental facilities and algorithms for in situ analysis in simulations.

2. Integration of data analytics with exascale simulations represents a new kind of workflow that will impact both data-intensive science and exascale computing. Exascale simulations require in situ analysis and visualization, thus necessitating a new kind of workflow for scientists. In situ analysis will impact the workloads that high-end computers have traditionally been designed for. Tighter integration of simulation and analytics in the science workflow will impact co-design of these systems for future workloads and will require development of new classes of proxy applications to capture their combined characteristics.
3. There is an urgent need to simplify the workflow for data-intensive science.
4. There is a need to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing. This is critical because earlier workflow models allowed for a separation of concerns between computation and analytics that is no longer possible. This approach is not sustainable in data-intensive science, where the workflow for computation and analysis will have to be co-designed. There is a need for an increase in the number of computer and computational scientists trained in both exascale and data-intensive computing.

The recommendations from the Subcommittee are that

1. SC should give higher priority to investments that can benefit both data-intensive science and exascale computing to leverage their synergies. For science domains that need exascale simulations, commensurate investments in exascale computing capabilities and data infrastructure are necessary. In other domains, extreme-scale components of exascale systems are necessary for near-sensor computing and other tiers of data analysis. Innovations in algorithms to address fundamental challenges in concurrency, data movement, and resilience will benefit data analysis and computational techniques for both data-intensive science and exascale computing.
2. ASCR should give higher priority to investments that simplify the science workflow and improve the productivity of scientists involved in data-intensive and exascale computing. Greater attention must be paid to simplifying human-in-the-loop workflows for data-intensive science. A Virtual Data Facility (VDF) will provide a simpler portal for data services than current systems. Libraries of scalable data analytics and data-mining algorithms and software components should be developed for use in workflows. New classes of proxy applications should be created to capture the combined characteristics of simulation and analytics to feed into future design/co-design activities.
3. ASCR should adjust investments in programs (such as fellowships, career awards, and funding grants) to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing. There is a significant gap between the number of current computational and computer scientists trained in both exascale and data-intensive computing and the future needs for this combined expertise in support of DOE's science missions. ASCR investments (such as fellowships, career awards, and funding grants) should increase the pool of computer and computational scientists trained in both exascale and data-intensive computing.

Graham said that the Subcommittee had done a spectacular job. She commented that the word “higher” in the recommendations should be “high” and that security is the real issue in big data, not privacy. Sarkar agreed that the report could definitely emphasize security.

Blumenthal said that the report needs to provide a “big picture” as in Fig. 2.1, but it is difficult to tell what is going on that figure. Also, the discussion of workflow is distributed and diluted; workflow is very important; a discussion of its different dimensions might help. Privacy may be important in a situational context (i.e., as with metadata, especially in the biological sciences).

Chan noted that there are two types of big data: elephant (big-science) data and mouse (small-science) data. These two forms of big data have different challenges in managing and archiving the data. He asked if there had been any discussion of the possibility of leveraging the technology that had been developed in the commercial sector (e.g., by Google), and who should take the lead in developing the enabling technologies? Sarkar replied that, on the data side, there was a lot of discussion referencing Google, for example, when the volume of data that DOE and Google need to deal with was considered. When the discussion shifted to the intersection with exascale computing, the relevance of Google became less. What Google is doing in terms of data management is very relevant, but the scope of the study calls for more specialized requirements in regard to exascale computing. Chan suggested that there be a short paragraph in the narrative to make it clear what can be leveraged (e.g., in architecture) and where DOE needs to take the lead.

Giles commented that the discussion needed to be centered on what was needed to allow a vote on the report. He asked how this workflow and ASCR activities engage with work in the other offices of SC. Does the Subcommittee see changes or additional forms of support that are needed to involve that human interaction? Should there be co-design for big data or INCITE or are the right structures already in place? Sarkar noted that in the current structure, there are opportunities for interaction in computing and data support. More interaction is needed; the situation is complicated by the fact that all of the offices in SC have different performance metrics that they have to meet. Applied mathematicians, computer scientists, computational scientists, and other types of researchers can have a big influence on simplifying the workflow because science is dominated by analysis and analysis will need analytical algorithms, and scientists will need support and interaction to develop those statistical and analytical algorithms.

The floor was opened for public comment. David Brown (LBNL) commented that there is a need for the development of new analytical algorithms. The problems coming up are not addressable by those currently in use. Sarkar responded that there *are* algorithms being developed; they could be highlighted in the report.

Giles asked if there should be something in the report about the underlying research on analytics. Sarkar agreed that there could be more. Meza suggested that the word “research” could be added to the recommendations.

Giles asked for the consensus of the Committee on accepting the report subject to minor changes. The vote was unanimous (including the four members on the telephone) to accept the report and to thank the Subcommittee.

A break was declared at 10:11 a.m. the meeting was called back into session at 10:32 a.m.

Roscoe Giles presented the Facilities Subcommittee Report.

The charge came as part of a survey of facility needs conducted throughout SC. The survey asked each advisory committee for a "prioritization of scientific facilities to ensure optimal benefit from Federal investments. By September 30, 2013, formulate a 10-year prioritization of scientific facilities across the Office of Science based on (1) the ability of the facility to contribute to world-leading science, (2) the readiness of the facility for construction, and (3) an estimated construction and operations cost of the facility."

The process was for an AD to create a facility list, the relevant advisory committee to review the list and make any additions or subtractions, a report to be forwarded to the Director of SC with those recommendations and justifications, and the SC Director to prioritize the proposed new scientific user facilities and major upgrades across scientific disciplines according to an assessment of the scientific promise, the readiness of the facility to proceed to construction, and the cost of construction and operation. An ASCAC subcommittee held a strategic planning workshop on Jan. 30, 2013, with teleconferences and e-mail exchanges before and after.

The ASCR facilities "ecosystem" includes NERSC, the two LCFs, and ESnet. The plan covers upgrades to those facilities. A data facility (the virtual data facility or VDF) may be needed for large-scale data handling shared among the computing sites. The plan reflects a balanced roadmap for upgrading existing ASCR computing capabilities to meet the expected and emerging needs of DOE and the nation's scientists. The Subcommittee categorized the readiness in the near term to be A for NERSC, the LCFs, and ESnet and to be B for the VDF. It categorized the readiness in the longer-term to be B for NERSC and the LCFs, to be B+ for ESnet, and to be C for the VDF. It categorized the expected impact to be A of all four facilities. The longer-term categorizations were lower because there are technological challenges that need to be overcome. The VDF is needed, and that need will grow. In the future, exascale simulations, visualizations, and other services will be available to the whole ecosystem.

A model for the VDF shows a virtual machine serving the projects of all SC offices from a VDF common access layer that draws on individual VDFs at each LCF and NERSC. The challenges and goals of such a data service include providing seamless cross-site resilient access; scaling to data volume; providing real-time steering, processing, and storage; and covering multiple sites.

Graham noted that there was a desire for public transparency and availability of data. She asked if DOE were talking about making this huge amount of data publicly available. Giles replied, yes. There are two issues: who will gain access to it, and who will pay for it. There is not a good plan to address these issues, yet. Williams stated that long-term funding is needed for the VDF (e.g., from the other SC offices). Giles replied that it seems logical to put that financial responsibility on the user projects or facilities. The cost should not be put on ASCR. Blumenthal asked whether ESnet had a funding model that could be emulated. Giles admitted that he did not know.

Huntoon noted that the Subcommittee had discussed having analytics housed in the VDF and asked if the Subcommittee had considered whether the data might get too large to move. Giles said that it had not.

Chen asked whether remote, smaller projects would need to upgrade to higher capabilities to participate in the VDF. Dattoria said that the Office was doing some proofof-concept experiments to see how distance affects networking. Some of the interfaces are not there yet. The distance does not affect the connection of the Stanford Linear Accelerator Center (SLAC) with NERSC. More distant users are still one off.

Chan said that the issues of making raw data available and of the right of the data producer to publish first need to be worked out. Sarkar noted that there is a cultural aspect, and the cultures differ across communities and countries. Giles said that those arrangements grow up where they are. Some standards are needed, as are experiments in workflow design.

Khaleel said that the usage mode of big data should be considered as a factor; SC facilities are unique; the VDF will be unique, also.

Bland said that in regard to long-term storage and public access, everything cannot be stored forever. Peer-reviewed data should be retained. Access will be an SC policy issue. DOIs have looked at as tags for data sets. ESnet is an important part of the support capabilities, especially in security and login names. As much data as possible should reside at the facility that produced it.

Nowell noted that a policy that was recently released says that only data that has been peer reviewed will be publicly released. The Office of Science and Technology Policy (OSTP) says that DOE is to release data only within budget. Giles noted that the Subcommittee's advice does not need to follow the rules; it may not be accepted in the end. Nowell said that OSTP says that DOE is not necessarily the one who will make data available; the public sector may play some role.

Graham suggested reworking the conclusions table to make the impacts more visible.

Giles asked if ASCAC would approve the letter to SC. The sense of the Committee was unanimous to approve the letter, including the three members who were participating via telephone at the time (Petzold was not present).

Jacqueline Chen was asked to provide an update on the ExaCT Co-Design Center.

The goal of the Center for Exascale Simulation of Combustion in Turbulence (ExaCT) is to consider all aspects of the combustion simulation process, from formulation and basic algorithms to programming environments to hardware characteristics needed to enable combustion simulations on exascale architectures. The Center interacts with vendors to help define hardware requirements for combustion simulation at the exascale, it interacts with vendors and the computer-science community on requirements for the programming environment and software stack needed for combustion simulation, and it interacts with the applied-mathematics community on mathematical issues related to exascale combustion simulation.

Combustion is a surrogate for a much broader range of multi-physics computational science areas. The starting points are petascale codes with high finite-difference discretization, projection formulation, detailed kinetics and transport, and a hybrid parallel model with MPI plus OpenMP.

The expectation is that the exascale will require a new code base. The target computational capability at the exascale includes high-fidelity physics, support for both compressible and low-Mach-number formulations, block-structured adaptive-mesh refinement, higher-order spatial discretizations, higher-order temporal integration, support for embedded uncertainty quantification (UQ), and in situ analytics.

The goal is to discover the science that will allow engines to operate in the premixed charge compression ignition (PCCI) and homogeneous charge compression ignition

(HCCI) regimes. There, lower temperatures and pressures do not produce as much NO_x and soot as conventional diesel and internal-combustion engines do. In addition, operation in these regimes can increase fuel efficiency by up to 50%. These are chemical engines.

How to achieve the HCCI and PCCI regimes is a problem. It requires a tiny domain and grid size; the size of the state is about 1 PB; the high-water memory use is about 3 PB; the number of time steps is 1.2×10^6; the total run time is about 20 days; and the total amount of data needed for analysis is 1.0 exabyte.

The current petascale workflow model will not scale to the needed problem size. The I/O bandwidth constraint makes it infeasible to save all raw simulation data to persistent storage; the simulation and analysis must be integrated. A workflow needs to be co-designed that supports smart placement of analyses, visualization, and UQ, tracking large graphs and reducing checkpointing size with in situ analytics. Such an in situ design has the potential to reduce the amount of data by a factor of 1000.

The exascale machine would write data to disk for downstream analysis. Proxy apps could do some of the key tasks with a meta-skeletal workflow proxy app managing the process. There already are uniform-grid compressible-flow proxies, and proxy machines are under design: exaNode1 and 2 are many-core architectures, exaNode1 uses commodity network-interface-controller (NIC) and memory technology, exaNode2 uses custom on-board NIC and faster memory technology, and exaPIM provides processing near memory or processor in memory.

The co-design methodology considers that measurement alone is not sufficient; an analytic performance model needs to be developed; and performance needs to be validated with hardware simulators/measurements to confirm key predictions and to model what cannot be predicted analytically. A performance modeling tool chain has been modified to automatically predict performance for many input codes and software optimizations, predict performance under different architectural scenarios, operate much faster than hardware simulation and manual modeling, and include a cache model to capture the working-set reuse to produce a performance model and dependency graph optimization.

The compiler can characterize aspects of codes in a hardware-independent way and tailor-design them to co-design questions and goals. Some NNSA work involving Byfl compiler-driven dynamic software performance counters has been leveraged. With these proxies, some hardware/software co-design questions have been determined:

- What is the instruction mix, and should the chip area be used for vector units for special functions?
- How many registers are needed to capture scalar variables to avoid cache spills?
- How sensitive is the application to the memory bandwidth?
- What is the memory footprint of the application?
- How sensitive is the application to NIC characteristics in terms of latency and injection bandwidth?
- What is the best topology?

The basic characterization of the SMC [Sophos Mobile Control] proxy can be broken down into chemistry evaluation and dynamics evaluation in terms of the instruction mix and CPU time. Data can be reused, and data traffic can be decreased. Software and hardware performance improvements were estimated. Neither software optimizations alone nor hardware optimizations alone will get one to the exascale; both must be applied.

An SST [Structural Simulation Toolkit] macro was used to model contention on a realistic network. Results illustrate the importance of locality on performance; internode data placement needs to be topology aware because locality is critical to the code. A domain-specific language (DSL) is a language of reduced expressiveness targeted at developers in a specific, focused problem domain. It is a perfect fit for co-design. It is influenced and driven by both the domain and the architecture. It contributes directly as an enabling technology that insulates applications from the complexity of the architecture. LLVM [formerly Low Level Virtual Machine] intermediate representation for vendor tool chain interactions and supporting infrastructure is built on top of the memory-hierarchy-aware programming model and runtime.

To expose locality and express parallelism in the S3D code, tasks are coded in familiar sequential style, and Legion runtime uses region information to automatically extract parallelism and to map tasks with the same data in proximity to benefit from locality. This process results in the identification of hundreds of independent tasks by runtime, allows interleaving of up to hundreds of inter-node transfers to hide communication latencies, and has a code structure that is very similar to current Fortran code. This capability is currently being worked on at the Center.

At the exascale, the model of compute first, analyze later will be infeasible. Scientific data management and analysis (SDMA) challenges at the exascale include (1) the widening gap between compute power and available I/O rates will make it infeasible to save all necessary data for post processing; (2) analysis codes have markedly different characteristics compared to simulation codes, challenging current hardware and software stacks; and (3) understanding and modeling interaction and coordination behaviors of end-to-end workflows will be critical to co-design.

Proxy/skeletal applications have been developed for a representative set of data-analysis methods and to characterize machine-independent characteristics. The next step is to define and characterize relevant workflow architectures. Finally, one will have to integrate analytics and simulation proxies to understand end-to-end performance characteristics of the combustion workflow and express this with the meta-skeleton abstraction.

Extracting knowledge from simulation requires a range of analyses with different instruction mixes. Analytics instruction mixes cover a wide range of behaviors. Algorithms range from FLOP-free to having more FLOPS than the solver does. A wide range of behaviors is represented.

A rich design space of workflows is available at the exascale. In locating analysis/compute resources, does one use the same cores as the simulation, dedicated cores on the same node, dedicated nodes on the same machine, or dedicated nodes on an external resource? In synchronization and scheduling, does one execute synchronously with the simulation every *n*th simulation time step or execute asynchronously? How does one access data, and how persistent should it be? Should one use shared memory access via hand-off/copy, shared memory access via nonvolatile near-node storage, or data transfer to dedicated nodes or external resources?

An empirical study indicates that hybrid workflow architectures show promise for minimizing the impact to the simulation. In situ statistics and hybrid statistics produce a minimal impact, in situ visualization and hybrid visualization produce a significant impact, and hybrid topology and simulation produces a profound impact for non-scalable algorithms. The workflow can be optimized by doing some parts in situ, some in transit, and some off-line.

Models are a source of error in direct numerical simulation. Chemical kineticists and chemical engineers need to be informed about what chemical properties need to be pinned down more accurately for optimal utilization of a given fuel. Sources of uncertainty in the embedded chemistry model are reaction rates, missing reactions, and transport coefficients. Combustion intermittency needs to be characterized by space-time localized phenomena of interest, tractable for UQ. Adjoint equations need to be solved backward in time; the primal state is needed at all times.

Intrusive UQ poses a co-design challenge. The question is how to effectively manage the work and data flow. Storing the entire primal state is infeasible. Recomputation reduces storage costs. Computation can be focused to regions of interest. An adjoint UQ proxy app is available to explore trade-offs in the adjoint work and data flows.

The Center is heading toward developing more complete models for exascale combustion simulation, exploring hardware tradeoffs with vendors and computer-science collaborators, and pursuing a focused interaction with the programming-environment community to ensure that future programming models will support effective expression of the methodology needed for combustion simulation.

Giles asked how much the work was tied to the combustion problem and how much was generally applicable. Chen said that most of this work is not specific to combustion at all. Giles asked whether someone who is interested in applying this work to another process would go to Chen for assistance. Chen replied that people can contact the Center and talk with experts. A lot of leveraging can be gotten from this work.

Meza said that UQ methods are well understood and asked if any new developments were being seen. Chen replied that applying UQ to chaotic flows and systems is still problematical. The Center is working with the Scientific Discovery Through Advanced Computing (SciDAC) UQ effort. Meza said that this is a good example of why more research is needed.

Chapman said that she assumed that the Center was going to make these proxy apps available to people who used traditional languages. Chen said yes; they are available on the Center’s website, and the Center can custom-develop solutions.

A break for lunch was declared at 12:13 p.m.

Afternoon Session

The meeting was called back into session at 1:38 p.m. **Julia White** was asked to provide an update on the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program.

INCITE allocates resources at the ANL and ORNL LCFs. At ANL, those resources are the IBM Blue Gene/P Intrepid with a peak performance on 0.557 PF and the IBM Blue Gene/Q Mira with a peak performance of about 10 PF; and at ORNL, they are the Cray XK7 Titan with a peak performance of about 27 PF.

There are three primary ways to access an LCF's resources: 60% (4.7 billion core-hours) are allocated by INCITE, 30% are allocated by the ASCR leadership computing challenge program, and 10% are allocated at the discretion of the center's director.

INCITE's mission is to promote transformational advances in science and technology through large allocations of computer time and supporting resources. The program was initiated in 2004 and covers a broad range of research areas. The hours requested by proposers currently runs three times as many hours as are available (5 billion core hours). INCITE proposals must meet (1) a merit criterion (it must be a research campaign with the potential for significant domain and/or community impact); (2) a computational leadership criterion (it must involve computationally intensive runs that cannot be done anywhere else); and (3) an eligibility criterion (in which grant allocations are made regardless of funding source and nationality of proposer).

A twofold review process involves computational readiness review and peer review. Peer reviewers look at technical merit, originality, appropriateness of the proposed method, team qualifications, and reasonableness of requested resources. The annual timeline starts in April with the solicitation and proceeds through computational readiness review, panel peer review, selection, awards, and access processing (creating accounts).

For 2013, the call closed on June 27, 2012, with requests for about 15 billion core-hours. The number of proposals submitted increased nearly 20% over the previous year, and 61 projects were awarded, of which 20 were renewals. Of those awards, 43% went to U.S. academics, 44% to U.S. Government personnel, 5% to industrial researchers, and 8% to international PIs. The acceptance rates were 33% for new submittals and 100% for renewals. By system, 32 projects were funded on the Titan with an average size of 58 million core-hours (1.84 billion total), 27 on the Mira with an average size of 78 million core-hours (2.11 billion total), and 27 on the Intrepid with an average 27 million core-hours (0.721 billion total).

Of the peer reviewers, 50% are society fellows, agency awardees, laboratory fellows, National Academy members, or national society presidents; for continuity, 41% participated in the 2012 review process. Of the 83 science experts who participated in the review, 53% were U.S. academics, 22% were U.S. Government personnel, 7% were industrial researchers, and 18% were foreign national researchers. On average, the reviewers strongly believed that the INCITE proposals represent some of the most cutting-edge computational work in the field, the proposals were comprehensive and of appropriate length, and the science panel was sufficiently diverse to assess the range of research topics being considered.

The INCITE Awards Committee, composed of the LCF management teams, identifies top-ranked proposals by peer-review-panel rating and reports and by additional considerations, such as the desire to promote the use of HPC resources by underrepresented communities. Individual project allocations are determined to ensure that each awarded project has sufficient allocation to enable all or part of the proposed scientific or technical achievements, to maximize the scientific support provided to each INCITE project, and to allocate all of the available INCITE hours at each site. Of the 41 new projects awarded, 13 are led by new PIs.

In addition to the INCITE route, Director's Discretionary (DD) requests can be submitted anytime. The DD route may be used for porting, tuning, or scaling in preparation for an INCITE submittal. One must submit applications at least 2 months before the INCITE Call for Proposals closes.

It is expected that 5 billion core-hours will be awarded this year on the Titan and Mira, with an average award of more than 50 million core-hours. Individual awards will be up to several hundred million core-hours. INCITE is open to any science domain and seeks computationally intensive, large-scale research campaigns.

Graham asked what percentage of the renewal proposals was successful. J. White replied, 100% this year; not in previous years. Graham asked if the pre-review of renewals would lighten the reviewers' workload. J. White replied, yes, but then the grantees would have a track record of only 6 months. What is being looked at is staggering the renewal deadlines.

Huntoon asked if the Office kept track of actual usage. J. White said that the Center itself tracks the information.

Giles asked where unsuccessful proposers go. J. White said that it depends. Sometimes they just need to get more ready to use the machine. They might seek DD allocations to become more ready for submission and consideration. They could also apply for an allocation through the ASCR Leadership Computing Challenge, which has a 6-month disconnect from the INCITE program.

Chen asked what fraction of time goes to non-U.S. proposals. J. White replied, about 10%. Chen asked what reciprocity there was. J. White answered that some systems in Europe are making their policies consistent with INCITE's.

Victor Zavala was asked to review optimization under uncertainty.

An INCITE project at ANL is trying to optimize the operations of power grids, which can be described and modeled mathematically in terms of demand, renewable resources, and topology for the real-time market and for the day-ahead market. Robustness is embedded in grids through the maintenance of spinning reserves. The grid's time volatility indicates the inability to forecast source demand very well. The United States has prices for each node on its grid. One seeks spatial homogeneousness in the market. Where there is great disparity in prices from node to node, prices can go extremely high or go negative and start market manipulations. Wind-power adoption can produce changes in generation of several orders of magnitude in 30 minutes. Algorithms called solvers have been used to overcome inequality constraints. These solvers leverage linear algebra equations, so ANL has been exploring interior-point solvents.

Scalable stochastic optimization can be used to make a here-and-now decision. Typically, scenarios are sampled *a priori* from a given distribution. The problem induces an arrow-head structure in Karush-Kuhn-Tucker (KKT) matrices that can use linear algebra. In a large power grid, one has thousands of generators. The number and size of scenarios and first-state variables are large, and decomposition is based on a Schur complement. As a result, it is hard to get good preconditioners.

The Illinois power grid has 1900 busses, has 261 generators, and operates 24 hours a day. The problem space is huge. The weather's temperature is going to vary across the grid. One needs to sample random values that introduce uncertainty. $O(10^4$ to $10^5)$ scenarios are needed to cover this highly dimensioned spatio-temporal space (the wind fields). On the Intrepid, 6 billion variables are solved in less than an hour with 128,000 cores. These problems must be solved in a short time, or they are useless.

Simulations indicated that grid volatility could be reduced (in terms of spot price) through deployment and use of resources. In addition, social welfare (in terms of mean price) can be distributed more evenly.

A plot of the normalized cost as a function of the number of scenarios indicates that the normalized cost rises quickly with the first 1000 scenarios and approaches 1 after 8000 scenarios. Incorporating weather forecasting, one can forecast demand for wind power well one day out but not beyond two days. Forecasts with the Weather Research and Forecasting (WRF) Model are, in general, accurate with tight uncertainty bounds. However, excursions occur, and the probability distribution becomes inaccurate on the third day.

One can reduce complexity in the connections between stochastic communication and UQ with adaptive grids. When one gets weather data from a vendor, because of computational constraints, they will have run their model with coarse resolution over the U.S. domain; high resolution would require 50,000 processors. By targeting the UQ with adaptive grids, one can significantly decrease the number of processors needed to get the forecast information. By enabling the handling of ambiguity, one can relax the resolution. At what resolution to run the weather forecast is the question. In optimization under uncertainty, the first challenge is dimensionality (the semi-infinite nature), and the second challenge is ambiguity; one needs to sample a family of distributions. There are no good deterministic Newtonian methods for smaller sample sets. Therefore one samples the distribution, solves the problem, and tries to make important statements around resolution. There is an inefficient management of scenario or network redundancy. The bottlenecks are (1) a method for constructing steps from smaller sample sets; (2) the fact that progress and termination are deterministic, not probabilistic; and (3) the inefficient management of redundancies.

Scenario compression was devised as the only way to scale to a reasonable degree. Other methods (scenario clustering and elimination) were not effective. Experiments have been conducted with sparse multilevel preconditioning with scenario clustering. The findings were that clustering is two to three times more effective than elimination; compression rates of 70% are achievable; and multilevel preconditioning enables rates greater than 80%.

There are very few transmission lines that are congested at any one time, so one can aggregate all of the network that is not congested into a single node and use the same principles of analysis. We are now looking at coupled infrastructure systems like natural gas and electricity markets. As a result, the size of the network is getting very large.

Giles asked how much propagation delays matter. Zavala replied that a third market is based on volatilization. There is a coupling of dynamics to the market. An attempt is being made to try to understand the market interactions. One needs to incorporate dynamic constraints.

Meza asked if one can exploit the structure in the data. Zavala said that it is an interconnected problem; a multilevel analysis is being done.

A break was declared at 2:48 p.m. The meeting was called back into session at 2:56 p.m. **William Brinkman** was introduced to present an update on the activities at SC.

His testimony before the House earlier in the day focused on sequestration. Strong feedback is being obtained on what the effects of sequestration will be on the national laboratories. SC's science programs are being cut back while many other countries are expanding theirs. The United States could fall behind in science and technology competitiveness, including the next big computing machine. The world has changed in the past 30 years. Europe now publishes more scientific papers than the United States does. The United States used to publish 30,000 papers each year in *The Physical Review*; last year it was 22,000. The biggest mess is the International Thermonuclear Experimental Reactor (ITER), where the United States has committed financial support, but its fusion budget is not keeping up. SC is trying to sustain its current facilities but is coming up short in money. Everything is slowing down. The Next-Generation Light Source ($1 billion) is not getting funded soon, nor is the second target for the SNS in Oak Ridge. It is not a good time. Consequences of the sequestration will be very serious. Science will have to retrench. He hoped that everyone would work on his or her congressman. The United States has to think where it will position itself in global science. If DOE could reprogram money, it could ameliorate some problems; it is trying to get Congress to agree to such a possibility. Science as a community will have to speak up loudly. Scientists are becoming minority players on several fronts. Private foundations are making a big difference.

Giles asked if Brinkman had any comments on facility planning and what the next steps might be. Brinkman replied that a new facilities plan is desired by next fall. SC is also looking at the light sources, where new designs are emerging, like free-electron lasers. The question is being asked whether one can find a new way to explore the future.

Giles asked about the exascale, noting that a lot of investments in high-performance computing need to be leveraged, and the leadership needs to be sustained. Brinkman answered that an attempt was made to communicate to Congress that one cannot just turn off the spigot for a year or two.

Giles noted that ASCR does not have a permanent AD yet. Brinkman responded that the position needs to be designated by the new Secretary of Energy and approved by the Senate. The issue is being worked on.

Chan asked if Congress appreciated the importance of science and technology. Brinkman replied that some in Congress do, but they are not "Congress." There are a lot of congressmen who do not appreciate that importance. SC talks with those who love it; it should probably talk to those who hate it. Blumenthal asked him to elaborate on those latter. Brinkman responded that there are a lot who have reasons to oppose DOE's funding (for such reasons as climate change). There are others who just want to cut government spending; they do not care where those cuts occur.

Giles noted that one of the advantages of advisory committees is that they do not live within the same constraints that the federal government does. The advisory committee members are doing the best in their fields and trying to integrate that into technology. To see that effort frittered away by those who just do not care is frustrating. Brinkman rejoined that Giles should go to the Hill because he is more credible than Brinkman is.

Chen and Graham had to leave because of weather-related flight changes.

Alexandra Landsberg was asked to review the Mathematical Multifaceted Integrated Capability Centers (MMICCs) in the Applied Mathematics Program.

The goal of Applied Mathematics is to support the research and development of applied mathematics models, methods, and algorithms for understanding natural and engineered systems related to DOE's mission. It is very broad in nature. The long-term goals are to conduct mathematics research that will affect DOE mission efforts 5 to 10 years out. The program funds cross-cutting computational mathematics projects that address foundational, algorithmic, and extreme-scale mathematical challenges. It also supports exploratory research to bring in highly innovative science.

The program is trying to address the long-term mathematical challenges for one or more DOE grand challenges that require new, integrated, iterative processes across multiple mathematical disciplines. It was left up to the proposers to come up with the challenges. Another objective is to identify a set of interrelated mathematics research challenges that represent abstractions of the grand challenges. These abstractions would then be optimally addressed through a multifaceted, integrated approach.

Three MMICCs were awarded:

- Multifaceted Mathematics for Complex Energy Systems (M2ACS);
- Collaboratory of Mathematics for Mesoscopic Modeling of Materials (CM4); and
- DiaMonD: An Integrated Multifaceted Approach to Mathematics at the Interfaces of Data, Models, and Decisions.

These three centers have broad DOE mission relevance in dealing with complex energy systems; mesoscale modeling applicable to materials, chemistry, and biofuels; and multi-scale, multi-physics challenges related to subsurface flows and materials for energy storage. A large collaboration was started up with $9 million of funding in the first year, $4.8 billion to national laboratories and $4.2 million to universities.

These centers' projects are unique because they are trying to integrate mathematical subdisciplines, allowing applied-mathematics researchers to take a broader view of the problem as a whole and devising solution strategies that attack the problem in its entirety by building fundamental, multidisciplinary, mathematical capabilities and tools. This effort requires applied-mathematics researchers to work together in large, collaborative teams to more effectively address grand-challenge problems. One has to be careful to balance these challenges with DOE applications. Although motivated by DOE grand-challenge problems, these projects do not have the tight coordination and application-focus of SciDAC partnerships. Each center was asked to lay out a 5-year research roadmap for individual research goals and integration goals. There will be annual reviews by program managers and external review committees. Each center will be evaluated on its technical progress and on its effectiveness of integration. Outreach to the research community is being made to serve as entry points for external researchers.

M2ACS is taking a holistic view to develop a deep mathematical understanding and effective algorithms to remove current bottlenecks in the analysis, simulation, and optimization of complex energy systems. For integrated, novel mathematics research, one needs predictive modeling that accounts for uncertainty and errors; mathematics of decisions that allow hierarchical, data-driven, and real-time decision making; scalable algorithms for optimization and dynamic simulation; and integrative frameworks, leveraging model reduction and multi-scale analysis. Regular teleconferences are used to coordinate efforts among participants. They will provide highlights of how the mathematics has not been used before and how it will be brought to the operators. Each center has integration challenges. The first M2ACS challenge is the simulation of critical energy devices, such as high-voltage, direct-current substations, with the simulation running on microsecond timescales with a multi-time-scale model. A two-year research plan has been drawn up for this effort. The second challenge is probabilistic modeling for complex energy infrastructure. This effort will identify a multi-scale spatio-temporal probabilistic structure from high-resolution data; develop scalable algorithms for data analysis with a complex multi-scale correlation structure; and produce new closure approaches for probability-density-function evolution derived from the nonlinear differential-algebraic-system equations. These challenges are being faced by researchers at different institutions requiring integration of the research team to keep them working toward the solution.

The goal of CM4 is to develop mathematical foundations for the understanding and control of fundamental mechanisms in mesoscale processes. The long-term DOE impact will be in the developing of a hierarchy of mathematical models and numerical approaches to seamlessly model and simulate a system from the molecular to the macroscopic scales. The Center is to address a broad class of applications, such as control of chemical reactions in combustion and catalysis, reactive flow in complex liquids, and the formation of new energy-storage materials.

The DiaMonD Center is an integrated, multifaceted approach to mathematics at the interfaces of data, models, and decisions. Its goal is to address the mathematical challenges of end-to-end modeling and simulation for complex DOE problems by considering multi-scale, multi-physics methods, uncertainty qualification, optimization and design, inverse problems, and data fusion. Mathematics addresses a broad class of applications, such as (1) subsurface energy and environmental flows and (2) materials for energy storage and conversion. The Center's ultimate goal is to support decision making through integrated approaches to solving an inverse problem, solving an optimal design problem, solving a control problem, and quantifying uncertainties from data to model inference to prediction to optimal design and control.

Each of these centers has about 20 researchers plus 12 postdocs and 11 graduate research assistants. These centers' cost-cutting projects can affect any of the integrated projects and have impacts on DOE mission efforts through applied-mathematics research.

Williams asked how coordination with DOE issues worked. Landsberg replied that ASCR does not directly co-fund with other offices of SC. It coordinates with the other offices, and they coordinate with ASCR on their funding. This is very informal. There is a strong awareness of what is going on across DOE.

Chan asked what would happen after 5 years. Landsberg answered that there will be renewals after full competition.

Meza asked what the mechanism was for getting a big winner quickly integrated with other DOE missions. Landsberg replied that it is building up relationships over time. Partnerships have a role.

Ceren Susut was asked to talk on transforming GEANT4 for the future.

A year ago, Daniel Hitchcock requested a workshop on GEANT4 [GEometry ANd Tracking], which has nothing to do with the European network but rather relates to the LHC and the collection, analysis, and interpretation of the data collected by the ATLAS detector.

GEANT4 is a C++ toolkit that tracks particles through matter, breaking the particle motion into small segments, applying appropriate physical processes and probabilities at each segment. The successor to GEANT3, the GEANT4 Project began in 1994 with its first public release in 1998. GEANT4 is distinguished from other Monte Carlo particle transport codes by the comprehensive suite of physics processes and particle types treated, the complexity of geometrical descriptions that leads to realistic representations, and a collaborative open-source model leveraging international expertise.

GEANT4 is an international collaboration with about 100 scientists maintaining the code. It is a big code/collaboration. The data are approaching 100 PB, and High-Energy Physics (HEP) has more simulated than collected data.

GEANT4 is a sequential code. CPU capability has plateaued, and emerging architectures are being looked at. With this in mind, a joint ASCR/HEP workshop, Transforming GEANT4 for the Future, was held to review the status, successes, and limits of GEANT4; determine the challenges posed by emerging architectures; consider the opportunities in algorithms and optimization; ascertain the research needed for robust, sustainable code; create a foundation for information exchanged among ASCR and HEP investigators; understand and not duplicate international efforts; and explore transformative advances via HEP–ASCR collaboration.

The workshop had 50 participants from the HEP and ASCR communities and lasted 1.5 days. The workshop report was finalized in September 2012. It described an ASCR/HEP joint program to explore existing tools, strengthen U.S. efforts to refactor/re-architect for GEANT, plan and implement necessary validation and testing processes so the physics is not compromised, develop efficient I/O strategies, explore the possibility of using higher-level abstraction, and explore how to handle petabyte- to exabyte-scale data with much lower human effort.

A 2-year joint ASCR/HEP pilot study is under way. The ASCR effort is led by Bob Lucas. The object is to optimize within the GEANT4's current framework. The outcome is expected to allow refactoring demonstrations and prototypes

Meza noted that, when someone is faced with extremely serial code, an option is to start from scratch and asked whether that option had been considered. Susut replied, no.

Linda Biven was asked to present an update on the SC data policy.

The SC Statement on Digital Data Management (Policy on Digital Research Data) is currently in draft form, but its important points are firm. The Policy will apply to all applications for research funding but not to Small Business Innovative Research and Small Business Technology Transfer (SBIR/STTR) awards and not to applications for time on user facilities. The requirements take effect Oct 1, 2013, and are consistent with recent OSTP guidance on "Increasing Access to the Results of Federally Funded Research." Furthermore, it takes into account ASCAC's input from the reports of summer 2011 and other public comments.

The focus here is on *digital* research data. Data management reflects all stages of the data lifecycle, but the focus here is on data sharing and preservation. The stated requirements are for PIs and research institutions, but reviewers and program staff will also have new responsibilities. The approach has been to have a policy that is specific to SC needs and missions and to give SC programs maximum flexibility in tailoring implementation of the policy while providing a clear statement of goals and expectations from the Office, which are to be consistent with administration Guidance and take into account input from the community and public. There is no desire to overburden the research communities with a policy that is inconsistent with policies of other research funding agencies.

The principles of the policy are:

1. Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.
2. Sharing and preserving data are central to protecting the integrity of science.
3. Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data-management planning.

The requirements of the policy are:

1. All proposals submitted to SC for research funding are required to include a Data-Management Plan (DMP) of no more than 2 pages that describes how data generated through the course of the proposed research will be shared and preserved.
2. DMPs must provide a plan for making all research data displayed in publications resulting from the proposed research digitally accessible at the time of publication.
3. Researchers that plan to work at an SC user facility as part of the proposed research should consult the published data policy of that facility and reference it in the DMP. DMPs that explicitly or implicitly commit data-management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written approval from that facility.

Giles asked if these policies can be modified at the level of offices within SC. Biven replied that they can be augmented. Giles asked whether an additional burden was being placed on the PIs without giving additional resources or help. Biven said that that was correct.

Meza asked what happens when a center sets up a collaboration of different institutions with conflicting data-management policies. Which policy is followed? Biven replied that the data-management plan follows the proposal; the proposer sets the data management plan.

Williams asked whether there will have to be a certification of newness of data and a digital identifier for publications and illustrations. Biven said that there has to be *some way* to identify an illustration and to associate it with a publication.

Blumenthal asked whether libraries or universities need some statutory repository requirements. Biven said that that issue is covered by the OSTP policy. The National Science Foundation (NSF) did pioneering work on this, too. There is cooperation among agencies.

Giles asked if the data-management plan has to be a part of the proposal that is peer-reviewed. Givens said, yes; it will be a part of the proposal. Giles noted that the policy might therefore give a leg-up to proposals from institutions that have archival capability.

Huntoon asked if there were a time framework. Biven said that it was desired to leave that flexible.

Chan asked whether a laboratory's ability to archive and provide open access met the requirements of the policy. Biven said, yes, it does, in general.

V. White said that this seems to be a time when memoranda of understanding (MOUs) are difficult to set up. This policy may make it impossible to attain international MOUs. Biven said that international perspectives will be respected when data-management plans are reviewed. There may be data embargoes employed.

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

Richard Carlson was asked to review the SBIR program.

The SBIR program has a technology-centric focus to push basic research results to market; it places a strong emphasis on the commercialization potential of a project. A robust outreach effort plays an important role in making this program effective. There are two major technical topics:

- Advanced Network Technologies and Services
- Increasing Adoption of HPC Modeling and Simulation in the Advanced Manufacturing and Engineering Industries.

The topics have been broken into management tools for network operators, optical network support services, video collaboration services, big-data-aware middleware and networking, turnkey HPC solutions for manufacturing and engineering, HPC support tools and services, and hardening of R&D code for industry use.

Last year, 140 letters of intent were received, and about half of those were followed up with proposals: 28 in Topic 1, and 50 in Topic 2. Twenty proposals were funded for a total of $4.1 million: three STTR ($0.5 million), 16 SBIR ($2.4 million) and one fast-track ($1.2 million).

In the Phase II awards, there were 25 FY12 Phase I awards, and 24 new proposals were received. They are still under consideration.

In the HPC topic area, the codes are being converted into products that commercial firms can use; software as a service is an acceptable business model; building energy models for predictive control; and visualization of lighting to use internal and external lighting for energy conservation.

In conclusion, ASCR's SBIR program has been restructured to push commercialization of ASCR-funded research. This restructuring has produced positive results in the first year. ASCR will continue a robust outreach program to engage small businesses in developing new products, tools, and services.

Giles asked if Carlton wished that the program were bigger or smaller. Carlton replied that the Office could do a lot more in several areas. The program has given some good results in the past several years.

Williams asked what else would one like to see pushed forward from research to production. Carlton replied that most of those successful products had come out of other programs.

The floor was opened to public comment. There being none, the meeting was adjourned at 4:30 p.m.

Respectfully submitted,
Frederick M. O'Hara, Jr.
Recording Secretary
March 25, 2013