Meeting Minutes
ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE (ASCAC)
September 26-27, 2017
DoubleTree by Hilton Hotel Washington DC-Crystal City
300 Army Navy Drive, Arlington, VA 22202

ASCAC Members Present

Keren Bergman
Barbara Chapman
Jacqueline Chen
Silvia Crivelli
Jack Dongarra (teleconference)
Thom Dunning
Tim Germann
Susan Gregurick
Anthony Hey
Richard Lethin
Satoshi Matsouka
Juan Meza
John Negele (teleconference)
Daniel Reed (Chairperson)
Vivek Sarkar
Dean Williams

ASCAC Members Absent

Martin Berzins
Vinton Cerf
John Dolbow
Gwendolyn Huntoon
David Levermore
Linda Petzold
Krysta Svore

Also Participating

Steve Binkley, Acting Director, Office of Science (SC), Department of Energy (DOE)
Arthur "Buddy" Bland, Oak Ridge National Laboratory (ORNL)
Christine Chalk, ASCAC Designated Federal Officer, Program Manager, Oak Ridge Leadership Computing (ORLC), Advanced Scientific Computing Research (ASCR), SC, DOE
Phil Colella, Lawrence Berkeley National Laboratory (LBNL)
T. Reneau Conner, Oak Ridge Institute for Science and Energy (ORISE)/ Oak Ridge Associated Universities (ORAU)
Barbara Helland, Associate Director, ASCR, SC, DOE
Paul Messina, Argonne National Laboratory (ANL)
Paul Kent, ORNL
Rusty Lusk, ANL
Inder Monga, LBNL
Manny Oliver, SC, DOE
Prabhat, LBNL, National Energy Research Scientific Computing Center (NERSC)
Barry Smith, ANL
Deniese Terry, ORISE/ ORAU
Matthew Wolf, Georgia Institute of Technology
Kesheng (John) Wu, LBNL
Bob Voigt, Krell Institute

Attending

M. Ambrose (teleconference)
Lisa Arafune, Coalition for Academic Scientific Computation (CASC)
Francis Alexander, BNL
James Ang, SNL
Nathan Baker (teleconference)
Laura Biven, SC, DOE
Kathryn Boudwin, ORNL

Ben Brown, SC, DOE
Rick Brueckner
Richard Carlson, SC, DOE
Jonathan Carter, LBNL
J. Carruthers
Linda Casola, National Academy of Sciences (NAS)
Susan Coghlan, ANL
Leland Cogliani (teleconference)
Claire Cramer
Lori Deachen, Lawrence Livermore National Laboratory (LLNL)
Erik Draeger, LLNL
Al Geist, ORNL
Ray Grout, NREL
James Hack, ORNL
Steve Hammond, NREL
Bruce Hendrickson, LLNL
Graham Heyes, JLab
Jeff Hittinger, LLNL
James Hack, ORNL
Ben Kallen, Lewis-Burke Association, LLC
Doug Kothe, ORNL
Leland Logliani, Lewis-Burke Association, LLC
Alexander Lazalere, US Council on Competitiveness (USCOC)
Stephen Lee, LANL
Carolyn Lauzon, SC, DOE
Antoinette Macaluso (teleconference)
John Mandrekas (teleconference)
John May, LLNL
Sandra McLean, SC, DOE
Chris Miller, SC, DOE
Jeff Nichols, ORNL
Thomas Ndousse-Fetter, SC, DOE
Lucy Nowell, SC, DOE (teleconference)
Manny Oliver, SC, DOE
Michael Papka, ANL
Alex Romoff, Cassidy Associates
Claudette Rosado-Reyes, SC, DOE
Alison Ryan, RCSS
John Sarres, LANL
Michelle Schwalbe, NAS
John Shalf, LBNL
Karen Skubal (teleconference)
Punita Sinha (teleconference)
James Stewart, Sandia National Laboratory (SNL)
Pieter Swart, LANL
Ceren Susut, ASCR
Valerie Taylor, ANL
Kelli Tomasulo, Flagship Government Relations
Kershin Klege, BNL
Justin Whitt, ORNL
John Wieland (teleconference)
Stefan Wild, ANL
Ashlee Wilkins (teleconference)

Tuesday, September 26, 2017
Morning Session

OPENING REMARKS FROM THE COMMITTEE CHAIR

ASCAC **Chairperson Dan Reed** called the meeting to order at 8:34 a.m. and welcomed everyone to the meeting.

VIEW FROM GERMANTOWN

Barbara Helland, Associate Director of the Office of Science (SC) for ASCR, shared decisions made and actions taken for the FY18 President's Budget Request (PBR). Next-Generation Networking for Science (NGNS) collaboratories was moved into computational partnerships and networking moved to computer science (CS). ASCR received additional money for the Exascale Initiative (ECI) with a caveat that an exascale machine had to be delivered by

2021. Argonne Leadership Computing Facility (ALCF) will work with Intel and Cray to expand their 180 petaFLOP (PFLOP) machine to exascale and deliver it in 2021 and Oak Ridge Leadership Computing Facility (OLCF) will accelerate progress on Coral II. The current budget includes $100M for ALCF and ~$149M for OLCF. The Chief Financial Officer's decision that exascale would be a construction project created the ECI and the exascale project (ECP). ECP is solely focused on research and discovery (R&D) while ECI includes everything except R&D as well as the delivery of the exascale machine.

Engagement between facilities and ECP is now critical because the exascale system (non-recurring engineering (NRE) and testbeds) has been moved to the facilities. Common principles and shared values among the DOE facilities and the ECP were articulated on August 25, 2017. NERSC is also involved because of their pre-exascale machine and eventual post-exascale machine. Intel and Cray have worked with ALCF and had the first design review last week, the baseline review will occur in November 2017, followed by the NRE contract award.

House markups gave ECP $170M and recommended $100M for ALCF, $112M for OLCF, $92M for NERSC, and $65M to support infrastructure upgrades and operations for ESnet. Senate markups were $184M for exascale, $100M for ALCF, $150M for OLCF, $94M for NERSC, and $79M for ESnet. The Senate also recommended $10M for the Computational Science Graduate Fellowship (CSGF), $24M for research and evaluation prototypes, and $117M for mathematical, computational, and computer sciences.

Helland is now the Associate Director of ASCR, Christopher Miller is the new AAAS Fellow, Thomas Ndousse-Fetter has been moved into CS, and Rich Carlson into Computational Partnerships. Avanti Patra left ASCR in August and Laura Biven will temporarily help in the applied math area. Helland requested recommendations for Intergovernmental Personnel Act Assignments (IPA). Bill Harragin will become Acting Director of the Advanced Computing Technologies Division; the intention is to move him to permanent Director. A 120-day rotation is in place for the Directorship of Scientific Discovery through Advanced Computing (SciDAC) between Ceren Susut (Acting Director), Stephen Lee (previous Acting Director), and Robinson Pino. Christine Chalk will officially become the Acting Director of the Facilities Division and will rotate this position with Ben Brown.

In FY17 seven Early Career Awards (ECA) were given. The Associate Directors will now lead the ECA program, rotating that responsibility. Pavel Bochev (SNL) was awarded the Thomas J.R. Hughes medal by the Association for U.S. Computational Mechanics. ASCR held their first Applied Mathematics Principal Investigator (PI) meeting, since 2013, in September 2017.

Cori received CD4 approval on September 19, 2017 and NERSC is actively exploring deep learning (DL) on Cori Knights Landing (KNL). OLCF is beginning to deliver Summit. Interestingly, the entire infrastructure for Summit is above the machine, rather than under a raised floor. OLCF held an annual review in September 2017 and plans to begin early science on Summit in June 2018. OLCF is currently doing machine learning (ML) on Titan with the neutrino detection and fusion experiment using TensorFlow and message passing interface (MPI). ALCF has the A21 machine and a small KNL machine, Theta, which will give ALCF experience with Intel chips and Cray's software (SW) stack.

In June 2016, a funding opportunity announcement (FOA) was released for SciDAC-4. Ninety-six projects were submitted and 25 were recommended for co-funding. Across 15 of the 17 national labs there are 22 lab-led collaborations. In FY17, ASCR is contributing $19.3M and SciDAC institutes are contributing $12M for SciDAC-4. Two SciDAC-4 Institutes have been funded: Resource and Application Productivity through Computation, Information, and Data Science (RAPIDS) and Frameworks, Algorithms, and Scalable Technologies for Mathematics (FASTMath).

Helland shared highlights from the ASCR-Basic Energy Sciences (BES) SciDAC-3 on the understanding of chemistry and dynamics in lithium-ion batteries, and ASCR-Fusion Energy Sciences (FES) SciDAC-3's collaboration between Fusion-Edge Physics Simulation (EPSI) and SciDAC Institute for Sustained Performance, Energy, and Resilience (SUPER). In FY17 four renewal projects from computational partnerships were funded at $4M. The FY17 Mathematical Multifaceted Integrated Capability Centers (MIMICS) call for proposals yielded 14 proposals and resulted in one funded project MACSER: Multifaceted Mathematics for Rare, High Impact Events in Complex Energy and Environment Systems. The Applied Math program is supporting 10 renewal projects at $6.6M per year. CS funded six recompeted projects at $4.3M per year in Scientific Data Management, Analysis, and Visualization.

ASCR Associate Directors were asked to articulate a cohesive program of quantum computing across SC. Three Quantum Algorithm Team (QAT) proposals were funded at $4M per year at LBNL, ORNL, and SNL. In FY17 two quantum testbed pathfinders were awarded to LBNL and ORNL. With the remaining funding at NGNS, a small, fully funded project will explore the Science Internet-of-Things (S-IoT). ASCR also held lab reverse site visits, hosting lab representatives at Germantown who shared their activities with the program managers. The strategic vision for ASCR's research program focuses on emerging trends in increasing instrumentation, interconnection, automation, and acceleration.

In FY18 ASCR will follow BES's model on Basic Research Needs (BRN) Workshops, doing in-depth, deep dive workshops on Extreme Heterogeneity in January 2018 and Machine Learning in late January 2018; SBIR/ STTR recurring and innovative topics are currently planned.

Helland publicly thanked Paul Messina for his service and leadership on ECP. Messina is retiring in October 2017 and Doug Kothe will become the Director of the ECP.

Discussion

Susan Gregurick asked about the new initiative in quantum computing. **Helland** explained the initiative is looking at environmental sensors; more biology than ML.

Jacqueline Chen inquired about the connections between the applied energy programs and SciDAC-4. **Helland** said ASCR had a SciDAC partnership with Nuclear Energy in the applied programs and has met with Fossil Energy who are interested in working with oil and gas. The most representative partnership is with High Performance Computing (HPC) for Manufacturing; HPC for Materials was announced last week and will fund explorations of HPC use in industry.

Dan Reed asked how ASCAC could be most helpful to Helland. **Helland** said new voices need to be at the table to help define the vision for the future and provide ideas on new directions. Mostly Helland requested that ASCAC keep the communication moving and articulate support for science to Congress.

UPDATE ON COMMITTEE OF VISITORS (COV)

Susan Gregurick, ASCAC, briefly covered the upcoming meeting for the COV Review of Management Processes for the ASCR Research Portfolio to be held the last week in October 2017. The COV has space for one or two more colleagues to serve on the panel.

Discussion

None.

40th ANNIVERSARY OF DOE – COMPUTATIONAL SCIENCES GRADUATE FELLOWSHIP (CSGF)

Bob Voigt, Krell Institute, discussed the CSGF program's 40 year history. Voigt began by recognizing Jim Corones' contributions to CSGF stating that without his vision and passion the CSGF program would not have flourished as it did. Corones passed away in 2017.

CSGF provides a stipend for 4 years, full tuition and fees, professional development support, and practicum support. There is a CSGF annual program review and an annual renewal application. The student must submit a program to follow with alterations being reviewed by the Steering Committee. The Fellow owns the fellowship and it moves with the student. Financial support predominantly comes from ASCR but also from the Advanced Simulation and Computing (ASC) office within the National Nuclear Security Administration (NNSA).

DOE awarded the first CSGF management contract to ORAU in 1990. Twenty-two Fellows were selected in 1991, in 1992 20 more Fellows were selected, management of CSGF moved to the Ames Laboratory, and Corones became principal investigator (PI) on the grant. In 1994 Fred House became the program manager and was a champion for CSGF. In 1997, the Krell Institute was formed with Corones as the President and the management of CSGF shifted from Ames Laboratory to the Krell Institute, and in 1999, the ASC office started contributing funds to the CSGF.

The CSGF evolved to include a program of study with courses in science/ engineering, mathematics, computer science, and HPC or parallel computing, three essays, and a practicum. The practicum, which must be a "broadening experience" with exposure to technologies and ideas beyond the thesis topic, is 12 weeks at a DOE lab. To eliminate confusion, the lab mentor and the student's advisor have to sign off on the practicum. CSGF is open to undergraduates and 1st year graduate students.

The CSGF is constantly evolving to accommodate applications in computational science, biological sciences, and now ML and artificial intelligence (AI). Applications are reviewed in a down select method to the end-point when the selection committee discusses the applications face-to-face and makes recommendations for awards.

The 2017 class of Fellows is 45% male and 55% female. Despite exposure to the CSGF it is a struggle to attract minority students; however, the steering committee is committed to improving recruitment. CSGF Alumni are employed at DOE labs (14%), academia (28%), and in industry (35%). Alums in the labs are strong supporters of and very much involved in the program. Alums at universities now have CSGF students themselves. The alums have helped secure the fellowship by writing many letters of support. In 2018 there will be a new program track focused on Math and CS with the goal to make four awards.

Discussion

Keren Bergman asked for the application schedule. **Voigt** said the program is available in October/November with applications due in January. The review starts in February and will be completed in March.

Vivek Sarkar clarified who CSGF is open to. **Voigt** said CSGF is open to U.S. citizens and permanent residents.

Chen asked if there is a mechanism by which the practicum work could continue as part of the student's research at school. **Voigt** said the work can continue through a second practicum or an unpaid connection with the lab.

Thom Dunning asked if the individuals in the new Math/CS track will be required to take courses in engineering and science as well. **Voigt** indicated that students in the Math/CS track will have to include exposure to a science or engineering discipline in their program of study. **Helland** added that ASCR was very supportive of the Math/CS track.

ESNET 6 UPGRADE PLANNING

Inder Monga, LBNL, discussed ESnet today, Next-Generation requirements, and the ESnet6 project. ESnet is DOE's international science network and SC facility connecting most DOE labs. One unique factor is that ESnet connects to hundreds of other science networks around the world and to the Internet. The science networks collaborate widely with different research and education networks around the world. ESnet is engineered and optimized for big data science in terms of super high bandwidth, unique capabilities in traffic engineering, management, quality of service, and protocols, and supports protocols for big data management. The current generation, ESnet5, was constructed from equipment purchased in the 2009-2011 timeframe.

Critical Decision 0 (CD-0) for ESnet6 was approved and passed in December 2016. There are three mission needs for ESnet6: capacity, resiliency, and flexibility. The network traffic has increased, from ESnet's perspective, by 72% per year since 1990. Exponential traffic growth means capacity needs to be added cost effectively. The current equipment is aging and with age comes more issues, including cyber resiliency of the network. Techniques like ML and data analysis are changing so flexibility is needed at all levels of the network to support the next-generation requirements of science.

Six architectures for ESnet6 we originally investigated. These technologies fell within three big categories: 1) packet-optical integration, 2) traditional router-optical platform, and 3) SW-defined networking. Packet-optical integration is potentially scalable and cost-compelling from power and space efficiency and network perspectives, but resiliency became an issue. From a traditional router perspective resiliency is good but it is cost-challenged and unsustainable based on growth projections. A SW-defined network is highly flexible and cost-compelling but presents a non-trivial design complexity. The decision is to create a hybrid approach.

The preferred ESnet6 architecture is called a hollow core. The hollow core is low-cost, high bandwidth capacity that moves data from one edge to another edge. There is no routing in the core, so there is minimal packet processing in the core. Protection and restoration for resiliency is built into the hollow core. Flexibility is added at the service edge. Five guiding principles for ESnet6 are zero-touch network management; always have access to accurate fault stream; always be able to take corrective action; leverage cloud technologies for scaling; and clearly define truth source for configuration data.

The ESnet6 aspirational timeline is to achieve CD-1/3a in 2018, CD-2/3b by 2019, and CD-4 in 2022. ESnet6 is actively managing 28 project risks. ESnet6 is envisioned to be an innovative, enabling platform that ensures scientific progress is completely unconstrained by physical location.

Discussion

Tony Hey noted that two big data generators in DOE are Advanced Photon Source (APS) and Cryo-electron microscopy (cryo-EM). **Monga** said Cryo-EM is moving 400 gigabits per second and the Linac Coherent Light Source (LCLS) upgrade is representative of light sources. **Hey** asked if NERSC was working with the National Science Foundation (NSF) to connect university sites with science DMZ given the low cost of cryo-EM equipment. **Monga** said that NSF has adopted the science DMZ architecture. A cyberinfrastructure PI meeting will be held at NSF next week and ESnet will hold brown bags for the participants. ESnet helps share information so all can utilize the infrastructure.

Dean Williams asked how much effort it will take to support security and SW over many application areas. **Monga** said SW is a new skill set for NERSC and they are hiring SW engineers. A formal SW engineering group was formed in spring 2017 to design a process to build production SW. All of the SW has moved to GitHub and is being managed in a unified way. NERSC is defining the bare minimum feature set across the SW pillars and the SW abilities and building an open source community. NERSC is also trying to build application programming interfaces (API).

Richard Lethin asked if there is convergence in science workloads between labs and commercial efforts in open stack, open compute, or open source. **Monga** said that from a very basic perspective data still has to be moved from point of production to use. From a network perspective people will be using multiple sets of computing resources and this will continue to grow. Some innovations can be leveraged. NERSC is having a workshop on FAUCET, an open-source SDN (software-defined networking) controller, at LBNL with Google. NERSC is using technologies from commercial companies, but that requires NERSC to fill in gaps, do the integration and the support. NERSC is building the SW engineering group in order to more actively use these innovations as they come to the market place. **Lethin** asked about ESnet6 research benefitting U.S. commercial companies and the unified data model. **Monga** said the unified data model is an internal representation not a publicly standardized or adopted model. NERSC works closely with start-up companies, researchers, and testbeds, allowing the start-ups say if the technology has impact.

Sarkar inquired about the extent of the international collaborations. **Monga** said there are quarterly meetings with the experimental site owners, he attends the physics meetings, NERSC works with light sources and multiple science communities, and enters into dialogue. **Sarkar** specified he was interested in the investments which are synergistic with ESnet6. **Monga** said there are many personal relationships that provide information and influence. For example, Monga has monthly meetings with the CEOs of GÉANT and Internet2, both of whom know about European investments and university investments in networking. The network community has global network architecture and Global Lambda Integrated Facility (GLIF) for meetings to coordinate what is happening across each of the networks.

Reed asked about the biggest financial and organizational stresses NERSC faces. **Monga** said there is tremendous support from the sciences and ASCR for ESnet and NERSC is adding staff to adapt and manage the current facility. There is constant prioritization and planning for the future. **Reed** clarified that data transfer networking is the circulatory system of R&D meaning a 10-year-old supercomputer would not be called state-of-the-art. Reed asked if something similar was needed for ESnet6. **Monga** thought that was an open question.

DISCUSSION ASCAC

ASCAC Chair **Dan Reed** introduced the opportunity for ASCAC members to hold an open discussion about any presentations from the morning session or issues of interest.

Keren Bergman mentioned the value of CSGF and the relative lack of support given there are only 20 slots for 400 applicants. **Christine Chalk** pointed out that there was an effort to consolidate all the STEM programs into NSF. However, the Secretary of Energy called on the Office of Science and Technology Policy (OSTP) and Office of Management and Budget (OMB) to save the CSGF. While the $10M is a tiny amount of money that was hard won. ASCR continues to talk to NSF and work on the workforce issue, but it is not a DOE mission. **Sarkar** asked if it was possible to track how much benefit ASCR gets from NSF fellowships. **Chalk** said a study could be conducted about the benefits to encourage NSF to improve the focus on the mission agencies. **Barbara Chapman** added that for the new areas and CSGF funding is very small compared to the scope of potential activities and emerging technologies.

Williams commented that the 10-year cycle for ESnet equipment is not on the same par with supercomputers and perhaps that should be changed.

Reed said that the politics of the graduate fellowship are complex, but it all comes down to money. The U.S. support, as percentage of gross domestic product (GDP), for basic research and scholarship is at a 40-year low (American Association for the Advancement of Science (AAAS) data).

Dunning suggested adding presentations from programs who are heavily involved in cyberinfrastructure for ASCAC to get a better understanding of their future needs. **Chalk** said ASCAC has been trying to schedule Assistant Directors from across SC to come to the meetings. **Dunning** noted this would have two benefits, one is ASCAC learns about what the future program needs and the Assistant Directors learn, in the process of preparing, what is needed in their research programs.

Chen added that it would be great to have a strong connection between the early science ALCF/ OLCF machines and ECP application projects.

Reed adjourned the meeting for lunch at 11:42 p.m. and reconvened the meeting at 1:11 p.m.

Tuesday, September 26, 2017
Afternoon Session

DOE SMALL BUSINESS INNOVATION RESEARCH (SBIR)

Manny Oliver, SC, discussed SBIR generally, operational changes, and assessment findings of the DOE SBIR programs. SBIR began in 1982 and is funded at 3.2% of agency's (>$100M) extramural R&D budget; STTR began in 1992 and receives 0.45% of each agency's (>$1B) extramural R&D budget. Eleven agencies have SBIR programs. Phase I is for feasibility and proof of concept, Phase II is for prototypes and process development, and Phase III is for commercialization. There are 12 DOE offices that participate in SBIR/STTR programs: four focus on energy technologies, six focus on science and engineering, and two focus on nuclear security.

The Phase 0 program, an operational change in the DOE SBIR/STTR program, was established in 2015 to address one of the four program goals to encourage involvement by socioeconomically disadvantaged and women-owned small businesses. Phase 0 services focus on proposal preparation and review; there are four core services and one optional service with up to $5K support.

Phase 0 participation averaged ~5% for the past 3 years. The percent of Phase I applications from underrepresented (UR) groups is trending upward, partially due to educational efforts.

Approximately 50-75% of UR applications resulted in Phase I awards. The most noticeable gap is from minority-owned small businesses. Initial assessment of the Phase 0 program yielded the following findings: increased recruitment activities are needed, more work is needed to compare Phase 0 groups with a comparable peer group, Phase 0 assistance has educated applicants about women-owned and minority-owned designations; and SBIR administrative funding is essential for providing Phase 0 services and outreach to UR groups.

The application and award process was streamlined to allow more time for applications to be submitted and to reduce time for the awards to be made. Sequential Phase II awards provide an additional 24 months to complete Phase II projects. Phase IIA awards provide additional time needed to complete a prototype project. Phase IIB awards provide additional R&D funding to help transition from prototype to commercialization. Phase IIA awards have increased since 2014 while Phase IIB awards have remained steady.

In a December 2016 report by the National Academies of Science (NAS) SBIR priorities at DOE showed a positive overall impact having met three of the four legislative objectives. More work is necessary to encourage participation from women-owned and minority-owned small businesses. Some of the key findings from the 2016 report were that DOE SBIR/STTR had substantially improved its programs since 2008. DOE adopted a number of other initiatives and pilot programs, collectively improving the program. DOE SBIR/STTR is also seeking ways to improve its data collection and tracking. Nearly half of the respondents to the NAS' 2014 survey reported sales, 78% received additional investment funding in the technology, and 71% reported SBIR/STTR funding was vital to their project proceeding.

The NAS study found that UR group targets had not been met, women-owned firms were only 9% and minority-owned companies accounted for only 7% of Phase I awards. Working with the Census Bureau's Small Business Ownership study the women-owned target should be 13-15% and minority-owned 10-13% range. The NAS found that DOE SBIR/STTR was meeting mission needs and fostering innovative companies. STTR is also meeting its program objectives. However, the NAS found that the national labs are challenging to work with in part due to the administrative burdens being fairly significant. NAS had 21 recommendations in five areas and the SBIR/STTR Programs Office is working to address these recommendations.

Discussion

Silvia Crivelli asked Oliver to expand on outreach and recruitment activities with UR and women-owned companies. **Oliver** said SBIR/STTR has been working with the Small Business Administration on outreach to UR groups. Program managers meet one-on-one with the small businesses who are interested. The most useful element for recruitment into Phase 0 has been letters of intent (LOI). LOIs inform SBIR/STTR who is applying and allows the office to introduce the Phase 0 program to that population. There is a contractor reaching out in different manners for new Phase 0 applicants. The office has also met with the Association of Women in Science to discuss opportunities.

Reed asked about the strength of the connection to university technology transfer offices. **Oliver** said there is some connectivity and technology transfer opportunities have been offered as topics in solicitations. The connection to the technology transfer offices has not been leveraged for recruitment.

Gregurick asked if Phase II applications need to have a Phase I completion. **Oliver** said in the last reauthorization, only three agencies were given authority to go directly to Phase II:

Department of Defense, Department of Education, and Department of Health and Human Services.

Lethin asked how the DOE programs compare to other agencies in terms of mission impact. **Oliver** said the SBIR office has companies who routinely work on technologies that may be useful publicly or even for a DOE lab. The challenge is getting a handle on the public benefits. Department of Defense (DOD) is a procurement agency; DOE is not. Of the five largest agencies, National Aeronautics and Space Administration (NASA) and DOD are contracting agencies for the SBIR programs because they are looking for customers for their technology, whereas National Institutes of Health (NIH), NSF, and DOE are typically looking for the public benefit.

Reed asked about connections to the Innovative Corporations (I-Corps) program. **Oliver** said the I-Corps/NSF program originally targeted university PIs who were working on basic or applied science. NSF sees I-Corps as a feeder program to SBIR. The Office of Energy Efficiency and Renewable Energy (EERE) set up training centers at some of the national labs for PIs. Other agencies are also leveraging the I-Corps program.

MACHINE LEARNING (ML)

Prabhat, NERSC, LBNL discussed Deep Learning (DL) in industry and DL for science. NERSC is a DOE mission HPC user facility. NERSC supports a range of workloads for all of the DOE applied offices. NERSC has 6,000 users, 700 unique projects, 700 codes, and a truly international set of users from all of the states and many different countries. NERSC is now seeing a rapid growth in data from user facilities and more complex scientific workflows. There is an increased requirement for streaming data at HPC facilities; users need real-time and interactive feedback because there are high demands in productivity in the data world. Finally, the typical HPC stack at HPC facilities is simply not up to the task; tools like C, Fortran, C++ are aging. There is a large gap in productivity between the HPC world and the data science community. Users want to see a rich data stack that is available, functional, and performant, on the HPC machines. Important science problems require simulation but increasingly rely on data analytics. The scientific discipline that is best suited to address data analytics is statistics, and ML is simply a modern flavor of statistics, more popularized and certainly made more successful by computer scientists. A lot of data analytics users require statistics and ML tools to be deployed at places like NERSC.

Apart from listening to users and organically understanding the landscape, DOE does conduct Exascale Requirements Workshops to solicit feedback from all of the applied offices on their needs. The focus is on the exascale ecosystem, beyond just the HW, and ML has been called out as an important theme. The types of problems applied programs have include classification, regression, clustering, dimensionality reduction, design of experiments, feature learning, and anomaly detection.

To address these needs NERSC can set-up hardware and add key technologies in the data stack. Cori is currently NERSC's flagship machine, data users predominantly mention I/O therefore NERSC has a 1.5 petabyte burst buffer available on the machines, ESnet provides a high-speed path to move data into the facility, and there are a slew of file systems. In terms of the SW stack, over the last few years NERSC has deliberately identified key technologies in the data stack that users are asking for such as data transfer and access, workflows, data management, data analytics, and data visualization.

The Big Data Center, a joint project between NERSC and Intel, is addressing performance, optimization, and scalability. The Big Data Center is explicitly trying to answer the questions: what are capability applications in the data space, what is NERSC's SW strategy, and how do we best utilize HPC HW?

Data analytics brings linear algebra, graph analytics, and image/ signal processing methods. DL will not supplant or remove all of the statistical approaches; there is a lot of need for statistical modeling, statistical analysis, and structured inference, and are requirements for linear algebra and image and signal processing. DL solves a different type of problem and should be seen as a complementary addition to the toolkit.

Multiple industries spend billions of dollars accentuating their AI strategy. There have been a series of breakthrough results in a number of computer science problems in DL. For example, Image Net is a computer vision challenge to find all objects in a scene and DL-based systems have beaten humans at this task. In speech recognition there is very likely a DL system in the back-end of a smart phone making sense of voice commands. DeepMind, acquired by Google, has successfully beaten two Grand Masters at the game of Go. And self-driving cars are now on the roads. AI is no longer science fiction.

Similar tasks in science exist such as pattern classification, regression, clustering, and feature learning. But there are also differences and unique attributes in scientific data such as multi-variate properties, high precision data, noise and structural artefacts in scientific instruments, and the underlying statistics of how the data is distributed and represented.

Prahbat described six use cases in climate science, cosmology, astronomy, neurology, genomics, and light sources to illustrated DL in science. The use cases are success stories indicating that for pattern classification problems and regression problems as long as there is enough training data DL is likely going to work. Preliminary indications in surrogate simulations, or fast simulations, using GANs are promising. Feature learning is a pre-requisite for solving these problems and DL is able to easily pull out features that are most relevant for the task at hand.

Short-term challenges for DL in science are pragmatic problems. These include complex data sets and ensuring that the DL libraries have 1st class primitives, hyper-parameter optimization, training, layers and filters in the network, learning rates and curriculum, canned functionality, performance and scalability, and scarcity of labeled data. Some core challenges include a lack of theory, protocols, and interpretability.

Prabhat speculated that between 2018-2020, there will be broad deployment of DL tools, libraries, and frameworks at HPC centers, domain scientists will self-organize and conduct label campaigns, there will be a flood of very simple papers published, and a range of pattern classification and regression problems and potentially clustering problems will be nearly completely solved. Beyond 2020 entire data archives will be completely segmented and classified which will enable a more sophisticated analysis such as anomaly detection of patterns, correlation analysis, coherent spatial-temporal regions with semantic labels, and causal analysis. Finally, the community will address long-term challenges such as the generalization limits of these methods, uncertainty quantification, interpretability, and incorporating physical constraints.

Prabhat speculated that the future scientist will have to provide semantic labels which are artifacts of human language and culture, and almost certainly will be doing interactive exploration. Fundamentally the scientist is going to be freed up to think deeply about mechanisms and hypothesis. Prabhat concluded stating that ML is an emerging requirement in the DOE community. NERSC has invested a lot in ML. DL has fundamentally enabled breakthroughs in long-standing problems in computer vision and robotics and speech recognition and there are direct analogs in DOE applications. A lot of low-hanging fruit can be exploited in 2-3 years but there are long-term challenges and the science community has to step up to. Finally, these are exciting times.

Discussion

Lethin referred to incorporating laws of physics and asked what routes Prahbat saw for this or if there is any promising work that was on the right path. **Prabhat** said industry has very little motivation to incorporate physical constraints and he did not think many people are looking at that. The kinds of ideas being discussed internally focus on constraining the family of basis vectors in the DL system that are consistent with the solution space of the physical domain; there are pieces of work in the conventional NIPS Computer Vision community as well. In the objective function, terms that have to do with negativity or sparsity can be incorporated. If 3-4 governing equations behind the phenomena of interest were known, working those in will take some effort, but he was not aware of many people doing that. **Lethin** speculated that there is an aesthetic role for the scientist. **Prabhat** suggested Lethin was alluding to extracting a unique set of equations corresponding to the phenomena; it might be multiple representations that result in the same fundamental dynamics, so how you choose between those multiple equations is an interesting question.

Matsouka asked about auto-encoders replacing simulations. **Prabhat** said he was nervous about the claim of DL replacing simulations; the notion should not be entertained until and unless the limits of these methods can be better characterized. The auto-encoder must be trained on the network and data is necessary to do that. **Matsouka** asked, in light of exascale, if the right infrastructure is being built for the next-generation machines or is something missing. **Prabhat** stated that at the very least benchmarks suites must be extended to add DL benchmarks and to ensure the entire SW stack works well no matter the machine. **Matsouka** stated that he was skeptical whether the I/O capabilities on the current and future machines can sustain the I/O required for training the new machines. **Prabhat** said he thought DL benchmarks must be incorporated into the procurement suites.

Chen asked if there is a way to include DL methods for regression and clustering and feature detection as part of an in-situ workflow together with the simulation. **Prabhat** said similar to the in-situ work occurring in ASCR, DL is simply an additional capability. Prabhat is more comfortable with classification and regression problems for the time being. Anomaly detection methods rely on lots of data. He suggested a complementary approach of using an uncertainty to determine an anomaly.

Crivelli mentioned that in an experiment she ran, Cori was very slow and they had to turn to GPUs, she asked if schemes are going to be available for users in near future. **Prabhat** explained that there is a 3-year gap between CPUs and GPUs. The Big Data Center is working to bridge that gap. The Big Data Center provides a direct connection to product teams at Intel and those teams are ensuring the tools are deployed on NERSC for production use.

Hey asked how much of the commercial big data stack is going to be relevant and used at NERSC. **Prabhat** said that NERSC first listens to users and if users are asking for a technology NERSC will conduct an investigation with vendors to explore deployment. **Hey** asked if advances in ML technologies can be made by using scientific big data. **Prabhat** said yes, easily. Some of the long-term challenges will certainly help move methods forward. The HPC community is well-poised to improve the technology in the single-node performance and multi-node scaling.

Sarkar asked about opportunities where the HPC SW stack is influencing the scalability of DL. **Prabhat** said things are not converging. The data science community has completely different dynamics. For example, Google is happy with remote procedure call (gRPC) and have left MPI to the open-source community. However, if vendors can indicate there is a reasonably stable stack and libraries they are going to optimize, then the HPC community can come in and start extending things.

40th ANNIVERSARY OF DOE – PARTNERSHIP BETWEEN NNSA AND SC IN HPC

Paul Messina, ANL, shared historical information concerning the collaborations between NNSA and SC in HPC. Beginning in the 1950s and 1960s, DOE labs had a major role in producing what was called sub-routine libraries, usually for numerical methods, and that tradition has continued in other fields. The partnerships were predominantly informal, collegial, open, and exploratory efforts that led to impactful discoveries. Messina discussed TOOLPACK, POOMA, Visit, Seven Dwarfs, LINPACK, LWG, CrOS, and PVM.

TOOLPACK (1972) was an effort to facilitate portability and productivity and was used with early LINear equations software PACKage (LINPACK) versions to automatically transfer from single precision to double precision. Parallel Object-Oriented Methods and Applications (POOMA) is a framework developed mostly at Los Alamos National Laboratory (LANL) aimed a different programming models. Visit is a visualization tool which began at LLNL as part of the Accelerated Strategic Computing Initiative (ASCI) and now involves ORNL and LBNL.

The labs have had a lot of activities in programming languages and models. In November 1976 when Cray-1 was delivered at LANL, the Language Working Group (LWG) was formed to recommend a common programming language for all the national labs. The biggest influence of the LWG was on the Fortran 88 standard. In the early to mid-1980s there was a lot of work on design and implementation of language for the new systems. Crystalline Operating System (CrOS) message passing routines were packaged as Express, a commercial SW product; p4 was at Argonne and Sisal was at LLNL. Parallel virtual machine (PVM) was originally developed at ORNL and was widely used until 2000.

In the early 1980s, Applied Mathematical Sciences (AMS), now ASCR, funded three university efforts for advanced architectures: New York University (NYU) Ultracomputer; University of Illinois, Urbana-Champaign (UIUC) CEDAR; and California Institute of Technology (CalTech) Cosmic Cube. Around the time of the Cosmic Cube, LANL had a project called PμPS (parallel microprocessor system). Many collaborations and interactions among the universities and the labs took place because people saw the mutual benefit and they made their systems available to each other. In 1984 via the Field Task Proposal (FTP) collaborative studies among AMS funded groups was discussed. By the very early 1990s, DOE decided to invest in the new advanced architectures and algorithms and SW and formed the Advanced Computing Laboratory (ACL) at LANL, the Center for Computational Sciences (CCS) at ORNL, and NERSC at LLNL.

Multiple labs (ANL, LANL, LLNL, PNNL, and SNL) and multiple agencies (NASA, NSF, and Defense Advanced Research Projects Agency (DARPA)) participated in the Touchstone Delta/Concurrent SuperComputing Consortium (CSCC). By the early 1990s, many of the labs had collections of systems with advanced architectures. For example, Ames Lab had the Scalable Computing Laboratory, ANL had the High-Performance Research Facility, LANL had the

Advanced Computing Laboratory, ORNL had the Center for Computational Sciences, and SNL had the Massively Parallel Computing Research Laboratory. Also, labs were partners with vendors; co-design with vendors began in the early 1950s and has continued in the area of advanced technologies, for example, SNL-Cray partnership for Red Storm, and LLNL-ANL-IBM partnership for Blue Gene.

In terms of data and storage, the labs played a major role. The High Performance Parallel Interface (HiPPI) HW design at LANL had test facilities so companies could go to LANL and test their HW and SW and verify the HiPPI interface worked. Another multi-agency project funded by DOE, NSF, DARPA, and NASA called the Scalable Input/Output (Scalable I/O) project which included (CalTech, ANL, LANL, LLNL, SNL, Princeton, Syracuse) led to Read Only Memory Input/Output (ROMIO) and MPI-IO. LANL had High Performance Data Storage project, and LLNL had the National Storage Laboratory.

DOE has been involved in networking and distributed computing for quite some time. The Casa Gigabit Network Testbed was a mid-1990s partnership project between CalTech, NASA Jet Propulsion Lab (JPL), LANL, and San Diego Supercomputer Center (SDSC) in conjunction with the University of California, Los Angeles (UCLA). The National Magnetic Fusion Energy Network (MFEnet) had a contract with Sprint in 1994 to implement the Asynchronous Transfer Mode (ATM) over Synchronous Optical Network (SONET). Distributed Collaboratory Experiment Environments (DCEE) were led by LBNL and included testbeds collaboratories at ANL, LLNL, ORNL, and PNNL. One of the important aspects of I-WAY (wide-area visual supercomputing) (1995) was that it explored issues of distributed wide-area resources meeting and scheduling.

Other collaborations Messina discussed included an NNSA activity (ASCI) supported by DOE labs, and support at the DOE labs for MPI activities and for visualization R&D work. More recently are the collaborations for big procurements such as the Collaboration of Oak Ridge, Argonne, and Livermore (CORAL), a procurement project between DOE labs and SNL, and the ECP involves both DOE and NNSA labs. In conclusion, most of these collaborations were self-organizing because people had a vision, they wanted to do something, and they saw mutual benefit.

Discussion

Dunning stated one of the categories Messina missed was the impact he had on the science and engineering community. At NorthWestCem (NWCem), although a PNNL project, was inspired by ANL and discussions with Messina and others that made it clear that parallel computing was going to be the future. NWCem is still the only electronics structure code that can treat extremely large molecules because it can make full use of the massively parallel systems. **Messina** agreed the impact of codes is a very important category, they can transform a community and many of them are developed and maintained at the national labs. **Dunning** added that in many cases their genesis arose from discussions between computational scientists and computer scientists at the laboratories. **Messina** said the CalTech Comprehensive Conceptual Curriculum for Physics (C3P) project also had a large number of applications, their impact was in showing that complex, not necessarily uniform applications, could be done effectively on distributed memory systems. **Chapman** added the impact of SW technology and compilers and run-times were instrumental, such as the work at the University of Illinois and CEDAR.

Sarkar asked if Messina thought younger colleagues and early career people have the same culture of independent thinking and empowerment today. **Messina’s** personal impression was that it would be very difficult for an early career person to be as successful as many were in the 1970s and 1980s and 1990s. Unsolicited proposals were looked at favorably but the same is no longer true.

Reed added that one reason things happened as they did because people like Paul Messina made them happen and Messina dramatically understated his contributions to a large number of projects on that list. Reed said thank you to Messina.

40th ANNIVERSARY OF DOE PANEL – DOE SUPPORTED COMPUTING TECHNOLOGIES THAT MADE A DIFFERENCE

Van Jacobson (TCP/IP), Rusty Lusk (MPI/MPICH), Phil Colella (AMR), Matthew Wolf (ADIOS), Barry Smith (PETSc), John Wu (FastBit), Buddy Bland (HPSS) (*invited*)

This year is the 40th anniversary of DOE. The communications group has been collecting vignettes of work the DOE has afforded that has made a difference. Chalk introduced the panelists who discussed notable computing technologies that ASCR either supported or has been involved in that has made a big impact.

Monga stood in for Jacobson and reviewed Transmission Control Protocol/ Internet Protocol (TCP/IP). TCP is still the workhorse of the Internet. TCP is still actively worked on and in the last 27 continues to be a problem that people research and write papers on. There are also newer algorithms that have been added into the Linux stack. TCP congestion control behavior is important because it affects throughput and when it affects throughput it affects users or the user applications. Jacobson also worked on a much improved algorithm with the Google Collaborative, called the Bottleneck Bandwidth and Round-trip time (BBR) TCP. BBR TCP has been rolled up by Google and is being used to download web browsers like Chrome and data from the cloud. What Jacobson did then and what he is doing now is going to have a tremendous impact on what we do on the Internet.

Rusty Lusk discussed MPI/MPICH. MPI is a message passing interface library specification and was designed to become an industry standard for message passing in parallel computing. The goals of MPI were portability of parallel programs, the enabling of performance, and being expressive. Competition among vendors in the late 1980s vendors on their programming systems and HW meant programs were not portable. The MPI Forum was assembled in 1992 and composed of CS, vendor representatives, and application computational scientists who wanted a portable standard in order to expand the market.

MPICH (MPI Chameleon) is one implementation of the MPI standard. MPICH tracked the evolving specifications from the 6-week cycles of the MPI Forum and gave immediate feedback on implementation issues. MPICH was completely rewritten in 2000 to implement the MPI-2 standard. MPICH is still current and is the basis of both research and commercial implementations.

In 1980, ASCR (then MICS) was supporting work on programming standards. In 1983, under Messina’s leadership, MICS established the Advanced Computing Research Facility (ACRF). There was research funding for advanced computer systems concepts as well as multiple researchers to program ACRF machines and free programming classes were offered to spread the word about parallelism. Both p4 and Chameleon arose from this environment. Lusk felt that MPI, not an ASCR project, was generously funded indirectly and traces its roots back to ASCR funding. MPICH, an ASCR project, has been crucial to the success of MPI and will continue into exascale. And the most significant contribution from ASCR has been the creation of a flexible research environment and continued support of MPI and MPICH.

Matthew Wolf discussed adaptive I/O systems (ADIOS) which addresses two key issues in data management: 1) self-description and 2) high performance I/O. ADIOS has had over a decade of investments in the system and its precursor versions. ADIOS is a framework for data intensive science. The system is built out of a set of layers. The core is focused on the base issues around marshalling careful management of metadata. There are API library interfaces. The framework is where all of the pieces fit together.

The ADIOS team is thinking about better engagement with the developer community, and those who are developing better indexing, transports, and compression techniques. Through the ECP effort ADIOS2 is rewriting the interface. ADIOS2 will redevelop the SW stack to be consistent so that the framework port is easy. Then internal data structures will be adapted to make it as high-performance as possible. Community input is needed to drive sustainability in the long-term.

Paul Colella discussed Adaptive Mesh Refinement (AMR), which is a method of adapting the accuracy of a solution within sensitive regions of simulation. Refinement regions are organized into rectangular patches and refinement is performed in time and space. AMR started with Marsha Berger's 1982 thesis which was published in a paper with Oliger in 1984. The foundations of AMR came from DOE Applied Mathematics, and it is impossible to understand what happens when you refine or de-refine meshes without mathematics. This field has had a long-term collaboration between mathematics and CS, yielding data structures beyond rectangular arrays, HPC on vector computers, and distributed memory processors. In 1986, John Bell and Colella both moved to LLNL and decided that AMR was the best way to deal with low-mach number combustion. Other areas using AMR include astrophysics, aerodynamics, subsurface flows, and plasma physics.

Mohrman and Galbraith (2005) published a study, "Dynamics of the Adaptive Mesh Refinement (AMR) Network: The Organizational and Managerial Factors that Contribute to the Stream of Value from the Basic Research Funding of the Office of Science", on how the AMR development process led to innovations. Some of the connections between projects and people circle round and have produced a rich community in these areas.

AMR is successful; it is a widely used and irreplaceable technology. The keys to success for AMR were a broadly-applicable technology being developed in tandem with specific difficult science problems being solved. The broad applicability meant that in the long-term AMR would have a big impact. A roadmap was sketched out in 1976, but the first publication that fully adopted AMR was not published until 1998. The 12 year gap was because technology had to be developed to solve the combustion problem. The roadmap was being developed at the same time as the High Performance Computing & Communications (HPCC) program and AMR was identified and funded as a Grand Challenge project. Visionary science partners helped secure buy-in from some practitioners which allowed scientifically technical engagement as well as political cover. AMR was asked to respond to a SciDAC-2 call and put outreach into the budget. The outreach allowed AMR to find collaborations where people could use the technology, and it helped the technology, SW, and algorithms because it stress-tested areas that might have been overlooked.

Barry Smith discussed a mathematical library, Portable, Extensible Toolkit for Scientific Computing (PETSc). ASCR has funded mathematical libraries that are used world-wide in many complicated situations since the 1970s. Mathematical libraries provide reusable encapsulations of algorithms that can support a large number of different users. Code can almost always be encapsulated in a library so long as the library has the flexibility to customize for specific things.

In addition to encapsulating mathematical ideas the library shields users from a lot of the architectural details of the machine.

Eigensystem Package (EISPACK) (1973) was a portable package that brought robust calculations for eigenvalues for dense matrices. In 1977, LINPACK was introduced and did the same thing for linear solvers. There were a variety of other mathematical libraries in the 1980s, a very nice set of packages for ordinary differential equations (ODE) integrators. Then in 1992 Linear Algebra PACKage (LAPACK), was released and was the last hurrah of the previous generation of numerical software and mathematical libraries. In the early 1990s, massively parallel computing began to disrupt everything. The issue was there were competing standards for programming the new computers. Also, none of the message passing protocols had any support for using them with libraries. The message passing protocols were all based on an assumption that someone would write code for each message passage system. MPI introduced the concept of an MPI communicator and there was suddenly a huge new opportunity for developing numerical libraries.

There are traditional libraries which focus on one particular activity. Bundling, or wrapper, libraries contain lots of their own functionality and provide clean interfaces to using the other libraries. PETSc provides a lot of functionality from the low level linear algebra up to preconditioners, cryloft solver, non-linear algebraic solvers, time integrators, and a TAO portion – a toolkit for advanced optimization providing a variety of scalable optimization algorithms.

John Wu discussed FastBit which has helped address the problem of quickly finding records within a large dataset, similar to a needle-in-a-haystack. The core of FastBit work is a compressed bitmap index. FastBit began in the Grand Challenge program with a storage access system involving a nuclear physics project called Solenoidal Tracker at Relativistic Heavy Ion Collider (STAR) looking for quark-gluon plasma signatures. Researchers in this project built a compressed bitmapping index, the Word Aligned Hybrid (WAH) Compressed Index, which performs an order of magnitude faster than the commercial technology DBMS Compression Index. In terms of applications, STAR was looking for small number of things from collision events. There are two cases: first FastBit enabled automatic extraction of data from archives to run analysis and second, a combustion case led to work with visualization researchers and grew into query-driven visualization. FastBit SW has been picked up by a commercial company who is using it to do molecular docking, finding where a small molecule can fit in crevices of a large protein, this is used for drug design.

Buddy Bland described High Performance Storage System (HPSS) a scalable archive system developed by a collaboration between IBM and five DOE labs (LANL, LBNL, LLNL, ORNL, and SNL). Founded in 1992, the important features are parallel data transfers, access control, redundant arrays of independent tape, scalable metadata performance, bandwidth, and capacity. The HPSS Collaboration has consisted of between 25-70 representatives from the labs and IBM. The technical and executive committees prioritize features and the development teams develop, test, and document changes and then package the versions. A critical piece of the HPSS collaboration is the commercialization element to meet IBM's business needs. IBM's commercialization has funded at least ½ of the development of HPSS.

HPSS is important because it provides a scalable storage system that has kept up with the demands of tera-, peta-, and exascale computing. This ability has required a sustained investment from ASCR and NNSA, which will ensure the system readiness for exascale as well as attracting commercial prospects to invest in such long-term projects. HPSS is well positioned for exascale because it has continuously evolved and remains a world-class SW that has demonstrated scalability and I/O rates and data storage by factors of thousands. HPSS pioneered the separation of data from metadata and I/O from storage, and is cloud-enabled.

The HPSS Collaboration is still going strong and most DOE HPC facilities rely on HPSS as their primary archival storage system. HPSS supports universities, weather and climate data centers, nuclear energy programs, defense systems and research labs in ~40 sites world-wide. And DOE SC and NNSA continue to support the development of HPSS.

Discussion

Chalk opened the floor for discussion and posed a challenge to help determine how to continue to develop technologies in the future given DOE is being asked to do more with less.

Lethin suggested a way to stretch resources is to consider engagement similar to that found in HPSS, especially since large data centers of Google and Facebook need the HPC technologies the labs and ASCR have spearheaded.

Sarkar stated that the feeling of autonomy and independence seems to be tied to the funding limitations, therefore if the future leaders can operate with more autonomy, they may be able to find ways to do more with less. **Colella** stated that patience is the one parameter you do control with less money; things take time to mature and have large impact. **Hey** suggested that the labs should collaborate, because it seems they all have competing products.

Smith indicated that the mathematical libraries community has been meeting and discussing ways to avoid duplication, the Program Managers realize this and are providing funding. **Dunning** recommended more base programs and fewer initiatives that require spending time writing proposals. **Reed** added that the system was fluid enough to allow researchers to pursue ideas without red tape. **Smith** said that the expenses are so high now that even trying to do a little thing costs a lot of money. **Colella** was unsure research was more expensive, there is just less money. **Helland** stated that one motivation for the reverse lab visits was to look for ways to do scientific focus areas.

Williams asked Lusk to describe what he did to keep the program afloat in the 1-2 years of decreased funding. **Lusk** explained a combination of things helped, a retirement, carryover money, and no student that year. **Hey** asked if the program's survival had anything to do with exascale and MPI. **Lusk** said it could have been, noting that a colleague told the program office, it is unclear how computers will be programmed in the next 25 years but it would be with MPI.

PUBLIC COMMENT

None.

Reed adjourned the ASCAC meeting for the day at 5:16 p.m.

Wednesday, September 27, 2017
Morning Session

Reed called the ASCAC meeting to order at 8:36 a.m.

VIEW FROM WASHINGTON

Steve Binkley, Acting Director of the Office of Science, thanked all of the ASCAC members for their participation noting their service is important. He mentioned that the Lab-Directed Research and Development (LDRD) subpanel's recommendations have been very far reaching throughout the DOE. The LDRD report has been circulated to other parts of DOE and has received quite a bit of attention.

Binkley provided an update on the DOE's budget and political appointees. The Secretary of Energy, Rick Perry, called out exascale computing as a priority and the shift to early-stage R&D caused a draw-back to some of the SC programs. DOE is awaiting the final FY18 appropriations votes and anticipates a continuing resolution until early/mid-December 2017. Dan Brouillette was sworn in on August 7, 2017 as the Deputy Secretary of Energy. Brouillette has been at DOE before as the Assistant Secretary for Congressional and Intergovernmental Affairs during the George W. Bush administration (2001-2003) and has worked in other areas of the government. Paul Dabbar was nominated in July 2017 to become the Under Secretary for Science. His Senate hearing was held in July; he has been voted out of committee and is now awaiting full Senate confirmation. Dabbar is currently on-board in DOE as a special government employee. Dabbar is very knowledgeable about the operations of DOE primarily through his service on the Environmental Management Advisory Board. He has also worked with the labs, served on directorate level advisory boards for several national labs, and in general has a deep appreciation for science and technology. Dabbar has said that he will be an advocate for science programs. Binkley expects that Dabbar will strongly back exascale and the activities that are currently under development in ASCR and other programs in quantum computing and ML.

In the FY18 PBR SC was given specific assignments to emphasize and incorporate in the budget perform early stage R&D, continue operations of the national labs, increase funds to exascale, and finish construction projects. The guidance given with the FY18 PBR gave a budget target that was ~$900M (17%) below the FY17 appropriations. The FY18 Congressional mark-ups have been issued. The House mark is close to FY17 appropriations budget and the Senate mark is slightly higher than FY17. A memorandum from OMB in August laid out guidelines for the FY19 process and DOE's FY19 target is essentially equal to FY18 target.

In conclusion, Binkley conveyed a strong message that it is important that the individual programs continue to produce the best possible science within the resources available and continue the well-established tradition of excellence in operating the scientific user facilities.

Discussion

Reed asked for the most useful thing ASCAC can do to help. **Binkley** said the most useful thing is to keep talking up the importance of science in the national economic and national strategic context and the importance of innovation. There seems to be a mindset that science is optional, and that it is fine to be second in certain areas. There are a number of areas where the U.S. cannot afford to be second. HPC and the types of research and technical activities that go on in the ASCR program are very important to national security and economic competitiveness. Secondly, the cornerstone of U.S. preeminence in science has always been the strength of the technical people working in the various disciplines. It is important to keep in mind what is required to continue to attract and retain people who are able to do science at the extreme edges.

Dunning asked if it is useful to share the contributions that DOE has made over the years in a more popular format. **Binkley** said it would be worthwhile to have a concrete story about how DOE and its predecessors have played a role in HPC, modeling, and simulation and what the impacts of those research activities and capabilities have been over time. Mark Castner, in conjunction with BESAC, is looking at the major accomplishments that the BES Chemistry and Materials Programs have made over the last 40 years.

Crivelli asked about the workforce development programs. **Binkley** said that the real mission space of SC is along the graduate fellowship lines and will continue to emphasize CSGF and the early career research program. The other education programs are in the Workforce Development for Teachers and Scientists (WDTS) area. WDTS is funded at a very small level and that will likely remain the same. Education and workforce development are areas in the mission space of the Department of Education. The national labs all have localized programs for earlier education outreach.

Chen asked about the connection of early stage science and the applied offices. **Binkley** said there is definitely a push in the applied offices to move to early stage research, this translates to the decrease in their respective budgets. There were already strong linkages between the applied programs and parts of SC, most notably in ASCR and BES and to some extent BER. The current Acting Undersecretary of Science is looking for those linkages.

Juan Meza noted that one of the Secretary of Energy's priorities was to reach 40% research and asked for the overall number for SC and for ASCR. **Binkley** stated that 40% has been a guideline from OMB for a while; the level of research funding in SC is not to fall below 40%. In the FY18 budget, SC's research ended at 39% and the OMB examiner is pushing to see what SC can do to reach 40%.

Hey stated that the focus toward basic science is in complete contrast to the United Kingdom (U.K.) where the focus is towards applied research and business development. **Binkley** said notably the U.K.'s quantum area is much more applied than the U.S. approach.

Jack Dongarra confirmed that the FY19 budget would similar to the FY18 budget. **Binkley** said the OMB direction, released publicly during the summer, made it clear that agencies like DOE would have similar target budgets to FY18. **Dongarra** asked what the impact is on programs SC is expected to grow in the out years, like ECP. **Binkley** added that with programs like ECP, there is also guidance in the budget process with certain targets to hit. The Secretary of Energy's office held the line on funding for high priority initiatives.

Helland responded to Meza's question and stated that ASCR is 48% research because ECP counts as research.

ECP APPLICATIONS

Paul Kent, Oak Ridge National Laboratory, shared information about Quantum Monte Carlo code (QMCPACK), an application development project within ECP. QMCPACK is a framework for systematically improvable quantum-mechanics based simulations of materials. The researchers are determining how to evolve the existing QMCPACK code to develop a performance portable implementation. The project seeks to do QMC for materials because it has the potential for very high accuracy predictive calculations. QMC for materials also has a range of potential applications and it can be used directly or for benchmarking and upscaling. Exascale will enable treatment of realistic, complex materials which are impossible at the peta-scale.

BES's Computational Materials Sciences Center (CMSC) goal is to develop and validate methods for the properties of functional materials for energy applications. Within ECP there are two classes of QMC methods being developed, real space QMC and auxiliary-field QMC, to enable cross validation. To keep the QMCPACK project synergistic and maintain a clear definition between ECP and the CMSC, ECP focuses on performance portability (single code) for exascale architectures, while CMSC is focusing on science and methods development, concentrating on the critical path for the science and existing code base.

QMCPACK code is a fully open-source code for computing the electronic structure of atoms, molecules, and solids. QMCPACK is currently used for several INCITE projects, it is a benchmark for CORAL, a readiness code for Summit at OLCF, and there is an early science project at ALCF. One thing QMCPACK is trying to solve is the two incompatible execution pathways, OpenMP and CUDA. QMCPACK's challenge problem consists of calculating a supercell of NiO (nickel oxide) and ensuring that computation within one day.

Kent shared some of the highlights of the first year of QMCPACK project. There has been significant new analysis of QMC algorithms, various implementations, memory usage, plus readiness for updating QMCPACK in later years. Technical reports have been written for each milestone (met 5 of the 7 milestones as of the ASCAC meeting with the remaining two to be delivered the same week). Most importantly QMCPACK has determined the parallelization strategy based on target problems. QMCPACK will need 20-50x speed-up on the node through increased parallelization. A large emphasis has been placed on improving development practices. Continuous integration based on GitHub Pull requests was adopted.

From the initial performance portability assessments KOKKOS look most deployable but some features are needed. OpenMP is promising but mature compilers are needed. There is more analysis work to do, including testing on additional architectures and programming models. Human factors include the skill level required to implement the methods and improvements to lower the on-ramp, either changes to the interfaces or better training materials and workshops. The QMCPACK project is growing quickly with 30-50 pull-requests per week and 177 contributors worldwide. A QMCPACK package was made for Spack revealing that additional features were needed in Spack to support QMCPACK. In FY18 the project will be doing more programming model assessments and will choose one by December 2018.

There are three risks for this project. The number one risk is personnel because there are a range of skills represented on the project and a limited number of people who can put one foot in both camps (who know enough about the methodology, the development trajectory but at the same time are able to touch the code or use KOKKOS or OpenMP). The second risk is the timeline for maturation of the SW technologies for performance portability, and third is the flexibility of SW technologies to address novel HW architectures. The researchers are optimistic about achieving their goals and they have identified limitations and desired improvements in SW technologies.

Discussion

Matsouka asked about the criteria for selecting program models, in terms of metrics or procedures. **Kent** said fundamentally they have to be capable, sufficiently available, and appear to be implemented for a long time. **Matsouka** asked if Kent will write down the criteria at some point. **Kent** confirmed that they will write out the criteria and justifications, but they do not have a point system; some subjective elements exist.

Chapman asked how the ECP platforms are begin taken into account. **Kent** said the project team is staying informed with the rest of the ECP. If something wholly new comes in, the team will adjust their plans and work with the SW technology partners to cope with the changes.

Dunning asked for Kent's thoughts on adopting new SW technology in the applications development process. **Kent** said the team is mostly concerned about libraries with a much smaller user base and less concerned about KOKKOS or OpenMP. An indication of the lifetime of a SW product affects the likelihood of the team using it.

Matsouka stated the continuity for SW assets will become critically important and is an area where international collaboration would be very important to scientists elsewhere. **Kent** agreed that international collaboration is essential because it will assure a wider user base for the SW, further helping with longevity and the identification of issues and improvements.

Lethin asked about pleasant surprises or unanticipated challenges. **Kent** said the biggest surprise was the amount of work needed on the OpenMP side in terms of compilers and the GPU limitation offload. There were a lot of start-up costs and efficiency improvements in communication are still being evaluated, that is a challenge because it is critical.

Reed stated that the issue of long-term maintenance is a challenge because it competes with resources for new activities. **Kent** said that a very nice, advertisable outcome from ECP activities would be if we influence a popular language to have much better HPC support. **Dunning** added that application developers will not risk building their code on a technology without long-term support. This is an issue to be addressed because many good technologies may be discontinued because of the risk.

Chapman commented that in the OpenMP space the feedback from the application teams has been taken very seriously and is influencing the standard.

Chen suggested a risk mitigation strategy against application developers sticking with safe technologies is to work closely in the co-design sense with some of the SW technology computer scientists and try out some of the bleeding edge programming models and runtime systems. A lot may be learned in the process and some of those key features might get absorbed into standards like OpenMP. Chen thinks application teams may have to code up key kernels in multiple programming models because things are not all that well known yet.

Break 10.10 a.m. to 10.28 a.m.

EXASCALE PROJECT UPDATE

Paul Messina, Argonne National Laboratory, updated ASCAC on ECP. Doug Kothe will become Director of ECP in October 2017 and Stephen Lee will continue to be Deputy Director of ECP. Some ECP Leadership Team members will be stepping down and a list of individuals to fill the positions is being vetted.

The fourth focus area, exascale systems which included NRE testbeds, is being phased out and moved to facilities, it is extremely important to have closer engagement with facilities. ECP still has three focus areas: applications development, SW technology, and HW technology. The scope of the HW technology may be expanded to include more activities that focus on integration and engagement with the facilities. ALCF has 10 applications and 5 of these teams receive funding from ECP. At OLCF, 6 of the 13 applications receive funding from ECP. At NERSC, there are 11 applications which are also funded by ECP. The seven seed applications in ECP have been renewed for FY18 and their funding has been increased to $1M per year. Five of the ECP applications projects were highlighted.

An interagency review of ECP SW stack was held in June 2017. A SW gap analysis draft has been completed and sent for review to the Laboratory Operations Task Force (LOTF), Computational Research Leadership Council (CRLC), and ECP program managers. A plan for coordinated regular ECP SW stack releases has been developed. Some SW milestones for the 3rd quarter include improved OpenMP 4.5 implementation in the LLVM Compiler, scalable memory usage in MPI, release of HPCToolkit for ECP testbeds, C++ API specification for BLAS (Basic

Linear Algebra Subprograms) and LAPACK math libraries, API specification for scalable checkpoint-restart, in-situ algorithms, and node power abstractions.

PathForward supports DOE-vendor collaborative R&D activities. There are six companies, Advanced Micro Devices (AMD), Cray Inc. (CRAY), Hewlett Packard Enterprise (HPE), International Business Machines (IBM), Intel Corp. (Intel), and NVIDIA Corp. (NVIDIA), with which ECP has contracts. DOE is funding these at $252M over 3 years which is supplemented by companies at $400M over 3 years. The ASCR Leadership Computing Challenge (ALCC) has allocated ~1B processor hours parsed out over three computing centers (ALCF, OLCF, NERSC) for ECP.

ALCF has selected its first exascale system. The exascale systems will provide characteristics information that will be used to focus ECP activities in the future. There will be an independent project review in January 2018 with the 2nd annual meeting held in February 2018. In summary, Messina stated that ECP is making excellent progress and there will be continued substantial management and technical challenges that will be tackled effectively.

Discussion

Lethin asked how proprietary information is managed since it needs to be encapsulated in the tools. **Messina** said historically that has been done by the “tool team” who works with the company. There still can be an open version but mostly this is handled on a case-by-case basis.

Bergman asked to what extent are there plans, collaborations, or benchmarking with international efforts. **Messina** said there are currently no plans. There has been no pushback on international collaborations but it is not high priority. There is one memorandum of understanding (MOU) in place with Japan which predates ECP.

Dunning asked for Messina’s thoughts on performance portability as a key performance parameter (KPP). **Messina** said that one KPP in the early drafts was on performance portability. However, the changes in the ECP require a reexamination of the definition of KPP success criteria for the project. Removing performance portability as a KPP has been discussed; encouraging the application to aim for it but not to make it a pass/fail option. **Dunning** expressed concern that performance portability would be a mandatory trait. **Messina** said application teams understand there is value in performance portability and good performance. The application will be running for decades and it is clear it will not be running on the same systems.

Matsouka discussed exascale metrics and application bottlenecks. He argued that FLOPS is a poor metric because the application could be being bottlenecked by something else. The application’s performance portability should be measured in terms of the bottleneck the application corresponds to. **Messina** felt that the emphasis on mathematical libraries can help. Careful tuning of the library to different applications may help achieve performance portability.

Crivelli asked how ECP encourages collaborations and if there is a repository for the codes. **Messina** said ECP is using the Confluence SW to facilitate collaboration and Confluence also acts as a repository for the codes. The projects also organize audio and video conferences to facilitate collaboration.

Reed asked if there were any particularly striking surprises in ECP. **Messina** said he was naïve about some things, specifically that with such a well-articulated national plan (2015 NCIS) that ECP would sail through, have higher budget levels, and experience closer collaborations with the other federal agencies. Messina did expect more cohesive, multi-agency activities.

Reed thanked Messina for all he has done (applause).

PRELIMINARY COMMENTS FROM THE FUTURE COMPUTING TECHNOLOGIES SUBCOMMITTEE

Vivek Sarkar, ASCAC, discussed the charge of the Future Computing Technologies Subcommittee, current thinking about opportunities and challenges, and preliminary recommendations. The Subcommittee charge is to review opportunities and challenges for HPC capabilities and determine R&D priority research areas and emerging technologies. The subcommittee felt the charge clearly referred to the post-exascale timeframe and that it was appropriate to focus on different timeframes for different technologies. The subcommittee will identify potential technology areas for future HPC systems, identify synergistic activities in these areas, estimate timeframes for different levels of readiness, and create a framework for assessing application's abilities to exploit future technologies. The subcommittee felt the biggest impacts have been that Dennard scaling ended over 10 years ago, that technology experts want to continue shrinking and diminishing feature sizes, and that performance improvement is flattening.

The future technologies being considered include von Neumann approaches with specialized computing such as GPU accelerators, reconfigurable logic, and CPU-integrated accelerators; Memory-centric computing; Photonics; and non-von Neumann approaches such as neuromorphic computing, analog computing, and quantum computing. The subcommittee sees extreme heterogeneity, specialization, and hybrid systems as dominant themes. The subcommittee is using the notion of technology readiness levels (TRL) to estimate the stages and timeframes.

Sarkar reviewed reconfigurable logic (FPGAs), memory-centric processing, photonics, neuromorphic computing, analog computing, and quantum computing. Sarkar shared some of the promises, uses, and challenges of each of these technologies. The subcommittee will offer a framework as to how application readiness for new architectures could be considered.

Four preliminary recommendations were: first, the space of post-Moore and the range of technologies being discussed plays to DOE's strengths and it is important for SC to play a leadership role in these areas. Second, there must be investment in application readiness preparations. Third, the workforce talent and staff members who are able to map applications onto emerging HW should be identified and fostered. And fourth, any exploration of post-Moore should include some stakeholder ownership by the facilities.

In summary, there is a wide range of technologies for future HPC. The subcommittee is studying different areas of research and emerging technologies. Heterogeneity and hybridization are common themes for future HPC. Applications need to be agile. And SC should play a leadership role in developing the post-Moore strategy for science in HPC.

Discussion

Reed asked Sarkar to provide a personal sense of the technologies. **Sarkar** said that big data have taught us not to only focus on compute. The neuro-memory processing trend is definitely going to be a contributor, and after that, it is going to be a hybrid system. However, different applications may have different forms of accelerators for different kinds of applications. **Reed** added, is the limiting case of that argument that every solution is a custom scientific instrument designed for that specific domain? **Sarkar** said that economies of scale will dictate that people who have related needs will cluster around the same technology rather than design their own.

Lethin asked if the subcommittee's report goes in depth on the needs of new algorithms and mathematics. **Sarkar** said new algorithms and mathematics are key enablers for quantum, but for neuro-memory processing people see differences in the algorithms, however there is also an idea of recasting algorithms to fit some of these capabilities.

Hey asked if the subcommittee had considered non-silicon technologies such as carbon nanotubes. **Sarkar** said they did not consider these but are open to suggestions to do so. **Hey** added that HPC needs to be made easier for the ordinary user if the desire is to build businesses on HPC. If someone who has no knowledge of parallel computing can simply port their serial code to a modest scale parallel system and get speed-up of 100, then more businesses may be engaged. That is a view of post-Moore's law. **Sarkar** thought that opportunity may not need to wait for post-Moore. The most promising commercial interest seems to be in the data analytics frameworks rather than simulations.

Matsouka suggested that one view that needs to be in the report is the fact that Moore's law is a law of continuing performance increase based on lithography decrease. That is over time there is a continuum of performance development. A distinction must be made, when going to more disruptive technologies, whether these technologies will lead to continued increased performance development or that they are just one-time speed-bumps. **Dunning** said that is a valuable addition to this whole concept of what is meant by the post-Moore effort. A pathway going forward needs to show increased capability. Although many of the labs are heavily involved in the ECP, there are other labs that could also contribute in this area. **Sarkar** said he felt optimistic about an exponential decrease in cost per unit. As the costs decrease the capacity can be increased to yield a contributing effect. **Reed** said that continual scaling for a technology has to be exothermic; otherwise the money runs out before the physics.

Crivelli asked about the Anton machine. **Sarkar** said Anton has been mentioned in the subcommittee's discussions and is an example of a specialized instrument for a certain task; Anton will definitely be cited in the report as an example.

Reed asked ASCAC if they would prefer a virtual meeting in December, and if they would rather have a 1-day virtual meeting or two ½ day virtual meetings, back-to-back. **John Negele** stated he strongly recommended two ½ day meetings.

PUBLIC COMMENT

None.

Reed recognized and thanked Juan Meza for his contributions to ASCAC as this is Meza's last ASCAC meeting.

Reed adjourned the April 2017 ASCAC meeting at 12:13 p.m. (ET).

Respectfully submitted,
Tiffani R. Conner, PhD, PMP, AHIP
Science Writer
ORISE