ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE
to the
U.S. DEPARTMENT OF ENERGY

MEETING MINUTES

July 29, 2021

Videoconference
ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE

The U.S. Department of Energy (DOE) Advanced Scientific Computing Advisory Committee (ASCAC) convened a Videoconference on Thursday, July 29, 2021 via Zoom. The meeting was open to the public and conducted in accordance with the requirements of the Federal Advisory Committee Act (FACA). Information about ASCAC and this meeting can be found at http://science.osti.gov/ascr/ascac.

ASCAC Members Present

Daniel Reed (Chairperson)
Richard Author
Martin Berzins
Vinton Cerf
Barbara Chapman
Jacqueline Chen
Silvia Crivelli
John Dolbow
Jack Dongarra
Timothy Germann
Roscoe Giles
Susan Gregurick
Anthony Hey
Bruce Hendrickson
Sandy Landsberg
Richard Lethin
Satoshi Matsouka
Jill Mesirov
John Negele
Krysta Svore

ASCAC Members Absent

Keren Bergman
Thom Dunning
Gwendolyn Huntoon
Vivek Sarkar

Also Participating

Katie Antypas, Lawrence Berkeley National Laboratory (LBNL)
Steve Binkley, Office of Science
Christine Chalk, ASCAC Designated Federal Officer, Oak Ridge Leadership Computing Facility (OLCF), Advanced Scientific Computing Research (ASCR)
Michelle Buchanan, Office of Deputy Director for Science Programs
Hal Finkel, Office of Science for ASCR
Emily Greenspan, National Cancer Institute (NCI)
Barbara Helland, Office of Science for ASCR
Mary Ann Leung, Sustainable Horizons Institute

Attending

There were approximately 151 attendees of the virtual meeting.

OPENING REMARKS FROM THE COMMITTEE CHAIR, Daniel Reed
Reed welcomed everyone and introduced new ASCAC members, including Richard Arthur (General Electric), Roscoe Giles (Boston University), Bruce Hendrickson (Lawrence Berkeley National Laboratory) and Jill Mesirov (University of California, San Diego).

VIEW FROM WASHINGTON, Steve Binkley, Acting Director of the Office of Science.

Binkley, currently serving as the as the SC Acting Director, will return to the SC Principal Deputy Director role when nominees are confirmed under the new administration,

Jennifer Granholm and David Turk were sworn in during the spring of 2021 as the DOE Secretary and Deputy Secretary, respectively. Geraldine Richmond and Asmeret Behre are the nominees for Under Secretary for Science and SC Director, respectively. Both nominees are scheduled for their confirmation hearing on August 3, 2021.

Under the Biden Administration, the DOE Applied Energy Programs were returned to the purview of the Under Secretary for Science and Energy, as organized during the second term of the Obama Administration. Positioning SC programs in closer proximity to Applied Energy Programs facilitates collaboration.

Tanya Das is new SC Chief of Staff and Natalie Thomas is a new Executive Assistant.

The FY22 President's Budget Request (PBR) seeks $7.44B for the SC, which is a 5.89% ($414M) increase over FY21's PBR. The House Energy and Water Development Subcommittee issued a lower markup of $7.32B. The Senate markup will be announced this week.

Discussion

Lethin asked about new political appointees' SC priorities. **Binkley** replied that there has been a resurgence in interest for the Clean Energy; Climate; and Diversity, Equity and Inclusion (DEI) priorities. Two Clean Energy Earthshot initiatives include increasing hydrogen use for energy production and direct air-capture of carbon. More information will be announced in the coming months. SC maintains a Clean Energy and Climate crosscut and has funded $2.7B in related research. The budget request for Climate research has grown, primarily in biological and environmental sciences, reflecting the need for better modeling and simulation to analyze climate effects. Created under former DOE Director Chris Fall, the new Diversity, Equity and Inclusion Office is currently being staffed, and the DOE is focusing on DEI in the federal workforce. The DOE has also been working with the national laboratories on diversity plans for approximately five years.

VIEW FROM GERMANTOWN, Barbara Helland, Associate Director of the Office of Science for Advanced Scientific Computing Research.

Benjamin Brown is the new Facilities Division Director for ASCR as of October 2020. Margaret Lentz is the new Computer Science Program Manager as of April 2021. Ceren Suset became the permanent ASCR Research Division Director in May 2021. Sonia Sachs, an ASCR Program Manager, is retiring to the private sector at the end of July 2021.

The FY22 PBR of $1.04B reflects increases over the FY21 enacted appropriations of $1.01B to the budget lines for the Applied Math ($51.05M) and Computer Science Programs ($49.77M), the Computational Partnerships Program ($86.03M) and the Computational Science Graduate Fellowship (CSGF) Program ($15M). CSGF funded a class of 32 fellows. The PBR also increases funding for the establishment of Energy Science Network (ESnet) High Performance Network Facilities and Testbeds ($93.96M).

The FY22 House Mark of $1.03B is $15M less than the PBR and recommends $160M+ for the Argonne Leadership Computing Facility (ALCF); $250M+ for the Oak Ridge Leadership Computing Facility (OLCF); $115M+ for the National Energy Research Scientific Computing (NERSC) Center; and $90M+ for ESnet. Additionally, the recommendation provides $250M+ for Mathematical, Computational, and Computer Sciences Research, including $15M+ for Computational Sciences Workforce Programs. Finally, the recommendation includes at least $10M and up to $40M for the development of artificial intelligence (AI)-optimized emerging memory technology. As planned, Exascale funding decreased for FY22.

The FY22 budget narrative highlighted several initiatives. ASCR is partnering with Basic Energy Sciences (BES) and Biological and Environmental Research (BER) on Biopreparedness Research Virtual Environment (BRaVE) to extend work created through the National Virtual Biotechnology Laboratory (NVBL). Considering lessons learned from the NVBL COVID-19 High Performance Computing (HPC) Consortium, ASCR is particularly interested in understanding the simulation and data management needs of SC programs and other agencies. BraVE funding is included in the Computational Partnerships budget line.

All SC programs, including ASCR, are supporting the Reaching a New Energy Science Workforce (RENEW) initiative. ASCR is using FY21 to conduct outreach to and listening tours of Minority Serving Institutions (MSIs), professional societies and representatives from underserved communities. ASCR has targeted $5M for RENEW in the Research and Evaluation Prototypes budget line.

The Integrated Computational and Data Infrastructure (ICDI) effort, a preexisting initiative, aims to connect SC user facilities to HPC capabilities, AI tools and data stores. This initiative will open doors for new scientific communities to access the DOE resources. Funding derives from the Applied Mathematics, Computational Partnerships and Facilities budget lines.

The Scientific HPC Data Facility (HPDF) addresses the need for a virtual data facility and is being designed from the ground up to support data handling, workflow management and real-time data processing. The HPDF will integrate with edge computing to provide access to real-time, architecturally diverse computing resources while interfacing with ASCR facilities.

Community Data Sets, along with Research and User Facilities, comprise the three pillars of the DOE SC Community Research Enterprise. The Public Reusable Research (PuRe) Data Resources effort consists of data repositories, knowledge bases, analysis platforms and other activities that aim to make data publicly available. Data sets from the Atmospheric Radiation Measurement Data Center; Joint Genome Institute; Materials Project; National Nuclear Data Center; Particle Data Group; and Systems Biology Knowledgebase are the initial recipients of the PuRe Data Resource designation.

ASCR selected ten Early Career Fellows in 2021. Their research focuses on Applied Math, AI, Computer Science and Quantum Information Science (QIS). ASCR released ten funding opportunity announcements (FOAs) in 2021 (four are restricted to applicants from national laboratories): 5G Enabled Energy Innovation: Advanced Wireless Networks for Science (lab only); and Microelectronics Co-Design Research (lab only); EXPRESS – Randomized Algorithms for Extreme-Scale Science; Data Reduction; Advancing the Quantum Internet Backbone for DOE Mission Science (lab only); National Institutes of Health (NIH) Collaboration: Bridge2AI and Privacy Preserving AI Research (lab only); X-Stack: Programming Models, Runtime Systems and Tools; Entanglement Management and Control in Transparent Optical Quantum Networks; Integrated Computational and Data Infrastructure Initiative; and Data-Intensive Scientific Machine Learning and Analysis;

A chapter in the *Office of Science User Facilities Roundtable: Lessons from the COVID Era and Visions for the Future* report, released following the December 2020 virtual roundtable, highlights challenges and opportunities for improving computing and data resources for SC user facilities.

A memorandum of agreement between the SC, NIH and National Science Foundation (NSF) was reached to form a Joint Oversight Group for QIS efforts shared between the NSF and SC and to cooperate in funding the NSF Collaborative Research in Computational Neuroscience across all three entities. The NCI, SC and National Nuclear Security Administration (NNSA) renewed their memorandum of understanding (MOU) to fund the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program.

Based on the report, *Transitioning ASCR after ECP*, in October 2020, ASCR established an internal software stewardship task force in March 2021 to study challenges posed by the increasing complexity of the software ecosystem supported by the Exascale Computing Project (ECP). The task force has met with key groups including the former ASCAC subcommittee on the ECP Transition report as well as ASCR facilities and ECP leadership. Upcoming meetings with the Computational Research Leadership Council are scheduled and the task force is preparing a request for information (RFI) to seek broader community feedback. Pending DOE travel policy, further discussions may be held in person at the 2021 SuperComputing (SC21) Conference.

Discussion

Cerf asked about budget numbers. **Helland** explained that values shown in parentheses were not necessarily negative; some referenced funds dedicated to specific AI and QIS initiatives. Funds targeted for ALCF and OLCF count towards the ECP. Numbers in red are not negative but instead highlight shifts in budget line organization. Starting with the FY21 Enacted Appropriations, Small Business Innovations Research/Small Business Technology Transfer (STTR/STTR) program values are included in subprogram lines. **Lethin** inquired if SBIR/STTR allocations are uniform across subprograms. **Helland** replied that SBIR/STTR allocations are uniform excepting education programs and facilities equipment; the $15M designated for CSGF and $5M distributed across three facilities are not taxed. Additionally, since the ECP is a construction project, it does not pay SBIR/STTR.

Cerf queried if ASCR is making use of commercial cloud hardware in addition to designing its own hardware for AI efforts. Is it possible for workloads to remain unclassified and run on the public cloud? **Helland** responded that NVIDIA chips incorporate TensorFlow. Though there is not an existing commercial contract, there may be future opportunities to work with cloud vendors on developing next generation computer hardware. Several DOE programs already use the public cloud.

Gregurick expressed excitement about the HPDF and asked about its management structure and relationship to the PuRe Data Resources. Recognizing that the HPDF is at an early planning stage, **Helland** stated the HPDF will be open to all, much like other user facilities. ESnet will be key in streamlining access to PuRe Data Resources enabling data integration with simulations to guide experiments.

Dongarra observed that there is an approximately 2.5% increase in ASCR's FY22 PBR over the FY21 Enacted Budget. Yet, the SC is targeted for an approximate 5.9% budget increase. Why is ASCR's percentage budget increase lower than that of the overall SC? **Helland** explained that SC budget increases are not meted out proportionally to programs. Some programs' budgets will likely increase by more than 5.9%. To be fair, ASCR received increases under previous administrations when other programs did not.

Reed requested more information about federal recognition for investment in basic research needs in the context of the SC's and ASCR's budgets. **Helland** commented that the current administration is focused on applied programs, and ASCR has benefited by enabling technologies in Applied Math, Computational and Computer Sciences Research. In particular, the SC budget for QIS has increased. AI, ML and a number of large projects have also seen increases. These basic research areas and associated technologies are key to the entire research infrastructure. **Giles** suggested that a figure illustrating how the Applied Math and Computing research budgets have trended relative to ECP decreases and AI and Quantum increases would be helpful. **Helland** agreed and suggested updating a table from the ECP Transition report. Applied Math is so critical to AI that ASCR has been focusing on the priority research directions identified by the Scientific Machine Learning Basic Research Needs workshop.

Noting an approximately 50/50 distribution between university and laboratory selections, **Mesirov** inquired if the Early Career Selection Committee predetermines award allocations across organizations. **Helland** said that award allocations are not predetermined. However, universities can request up to $750K per year for four years, and laboratories can request up to $500K per year for five years. Monetary figures represent total costs, including direct and indirect costs. The selection distribution reflects what the budget can afford.

Cerf remarked that software library maintenance is essential to sustaining the software ecosystem but is often poorly supported. Will the task force address maintenance issues? **Helland** agreed that policies and test suites developed to make ECP libraries interoperable are important for future exascale machines and will be key concerns for the stewardship task force.

Citing speculative figures of $50B to construct a US-based chip fabrication plant to be shared by US chipmakers, **Cerf** mused whether such a facility is worthy of DOE and ASCAC discussion given the state of supply chains and the potential for China to interfere with Taiwanese facilities. **Helland** confirmed that the DOE has already been discussing this issue for a number of years and is participating in a microelectronics interagency working group. Ensuring chip availability to keep the US competitive is a concern. BES is generating new materials, and it is not possible to predict what computers will look like in five years. **Cerf** mentioned the 5th Generation Effort launched by Japan in the 1980s and suggested that it might be time for the US to engage in state-of-the-art efforts domestically. **Helland** commented that Congress passed the Creating Helpful Incentives to Produce Semiconductors (CHIPS) for America Act in the FY20 National Defense Authorization Act. **Arthur** added that the Defense Advanced Research Projects Agency (DARPA) Electronics Resurgence Initiative also addresses this issue. **Ang** acknowledged DOE/SC and NNSA/Advanced Scientific Computing support for the Semiconductor Research Corporation Decadal Plan for Semiconductors Workshop Series. **Lethin** remarked that the cost of one new leading edge fabrication facility construction at $20B is more than twice the entire SC budget. **Reed** concurred that this speaks to the scale of the challenge.

Giles suggested that the ASCR stewardship task force consider support for career pathways and engineering activities when evaluating software sustainability.

Chen commented that consideration of data science, AI/ML methods and libraries at scale is important since ECP codes operate at scale, and many workflows include data science elements necessary for simulations. Incorporating data interfaces that allow many applications to access information is also vital. **Helland** agreed, noting the possibility of maintaining ExaLearn, the Co-design Center for Exascale Machine Learning. Data interfaces are key to the ICDI under development, and COVID underscored the importance of making data interoperable.

ASCR ROUNDTABLE DISCUSSIONS ON OPERATING-SYSTEMS RESEARCH, Hal Finkel, Advanced Scientific Computing Research, Program Manager

ASCR hosted a virtual roundtable discussion on operating systems (OS) research on January 25, 2021. Sixteen national laboratory scientists and more than 100 registered observers attended. Five invited speakers presented on Extreme Heterogeneity; Edge Computing; Data and AI; Containers, Virtualization and Programming Environment Management; and Resiliency and Correctness.

Two white papers captured the roundtable discussion: *Research Opportunities in Operating Systems for High-Performance Scientific Computing*; and *Research Opportunities in Operating Systems for Scientific Edge Computing.* Each paper delineates current topical trends; highlights research challenges, requirements and open questions; discusses the benefits of research success and identifies contributing research communities. The primary aspects of the modern computing ecosystem key to understanding future research OS opportunities are Full-stack co-design for extreme heterogeneity and scalability; Adaptive management and partitioning of resources; and Smart supercomputer systems and facilities. Notably, performing experiments via Field Programmable Gate Array (FPGA)-accelerated simulation tools may aid research, since investigators have a limited ability to affect hardware designs, and they may have limited access to large-scale HPC platforms for booting custom OS software. Scientific edge computing presents a number of unique challenges. As scientific experiments generate ever-increasing amounts of data and grow in operational complexity, modern experimental science demands unprecedented computational capabilities.

Discussion

Cerf asked if FPGA simulations can generate misleading distortion. **Finkel** commented that FPGAs are used as accelerators or to simulate what hardware will do at a cycle level. With many caveats, the way programs are optimized for FPGAs is different from the way codes are optimized. Application-Specific Integrated Circuit (ASIC)-targeted codes will run slowly on a FPGA and may lack a full cache hierarchy. These caveats aside, it should be possible to understand the main performance characteristics of an architecture. In fact, multiple FPGAs are sometimes grouped for large simulations. **Cerf** further inquired about clocking race conditions and high density chips. **Finkel** acknowledged that timing differences can be an issue. However, researchers using FPGAs take specific steps to understand where such differences will cause relevant effects. **Eric Van Hensbergen** (Arm) added that companies use large-scale FPGAs to validate design functionality. However, companies lean heavily on formal methods for race conditions, especially in the context of security.

Hendrickson recognized that OS for HPC platforms were historically considered a core research area, but run times are emerging as an important topic as systems engage more heterogeneous hardware and complicated workflows. Reframing discussions to address run time systems, not just operating systems, is worth considering. **Reed** agreed, noting that the boundary between microkernels and run time systems is subtle.

Gregurick drew attention to adaptive management push-and-pull, from vendors not keen on adaptive OS and from researchers who have concerns about the number of software iterations requiring configuration. **Finkel** observed that adaptive software should not impede but aid users in effectively managing available resources. If the portability challenge is thought of solely as a static problem, then success is unlikely in context of increasingly heterogeneous systems. Future research effort must focus on systems that address these challenges, including those with adaptive run times and where users can selectively direct system operations. **Cerf** observed that abstraction may aid resource management in a heterogeneous environment.

Lethin asked if there is a role for formal methods across general and HPC OS research communities. **Finkel** affirmed. Some of the recently announced X-Stack awardees propose combining formal and other methods to test and verify scientific software. Formal methods have challenges; they are difficult for non-experts to employ and often present scalability issues. The selected projects offer key insights in addressing challenges relevant to SC portfolios. The white papers also identify areas where formal methods are especially important, such as security, privacy, trust, integrity and provenance. Such verification methods should be employed to the extent that they can be used effectively.

Berzins stated that there is a desperate need for a portable way to run applications across machines with unique architectures and distinct vendor-supplied heterogeneity fixes. The solution does not need to be complicated. The ECP Kokkos and RAJA frameworks operate at a code scale that is likely too fine. A solution that functions at a higher level without the complexity of existing systems is needed. **Finkel** agreed. The ECP made major investments in this area, including C++ frameworks and modern Fortran. With the advent of AI and ML, researchers have also built on TensorFlow and PyTorch toolkits for Monte Carlo and particle-based simulations. It will be possible to create future applications able to use higher-level interfaces and to provide optimized solutions that run across heterogeneous platforms. At the same time, the development of such solutions and specific applications may require going down a level to optimize performance. We are seeing an ecosystem with increased capabilities at the higher levels and increasing capabilities for portable programming at intermediate levels. All these components are important. **Berzins** acknowledged the complexity of the situation, but reiterated the need to look for simple solutions. **Finkel** concurred. **Shih** (LBNL) added that there is ongoing work at LBNL to address Quantum-Accelerated HPC for quantum-automated workflow partitioning onto topology-networked heterogeneous parallel processors. Researchers are cross-testing the LBNL Quantum TIGER and an AI Adaptive Algorithm-Topology named ATMapper to design a hybrid of quantum latency-aware parallel processor assignment and scheduling. **A chat participant** noted that there are examples of TensorFlow used to build computational fluid dynamics (CFD) simulations that perform well. **Arthur** relayed that there is a CFD ASIC startup operated by vorticity.xyz that in theory, with memristors, could have amazing matrix inversion performance. Most intriguing is the exploitation of analog phenomena, such as Optalysys, to break some of the Big-O complexity limits that are artifacts of digital representation. However, lacking an effective way to generate software designed systems, this enters into very complicated and brittle system design and operations territory.

Cerf inquired if there is ongoing work on hardware reinforced security, such as the Capability Hardware Enhanced Reduced Instruction Set Computer (RISC) Instructions (CHERI) work at DARPA. **Van Hensbergen** replied that there is now an Arm-based prototype of the CHERI technology code named Morello to explore its potential in different scenarios, including in edge computing and HPC. Additional work addresses the M-class version of CHERI as part of the DARPA System Security Integration through Hardware and Firmware program.

Cerf suggested that a list of hard problems which, if solved, would make a huge difference in ASR would be helpful. **Arthur** relayed that the Council on Competitiveness

Advanced Computing Roundtable has been working with Neil Thompson from the Massachusetts Institute of Technology (MIT) on advocating for a microelectronics thrust to enable an ecosystem with more agile ASIC/Systems-on-Chip (SoC) development toward important problems underserved by hardware.

CHALLENGES AND LESSONS LEARNED IN EXPANDING PARTICIPATION IN COMPUTATIONAL SCIENCE AND ENGINEERING, Mary Ann Leung, Sustainable Horizons Institute

The Sustainability Horizons Institute (SHI) is supporting diversity in the ASC community through several programs. Since 2014, reports and research have indicated that DOE labs face workforce challenges and need new techniques to attract and retain diverse talent. At the 3-part Diversity and Inclusion panel, held during the2021 Society of Industrial and Applied Mathematics (SIAM) Computational Science and Engineering Conference (CSE21) panelists shared experiences of being a minority in CSE and discussed effective strategies for promoting diversity and inclusion. These challenges include a lack of visible role models, hidden needs taken for granted by others and explicit and implicit bias. Stories and discussion noted that institutional change is often slow and challenging, but dramatic shifts can be realized. Talent recognition, having an advocate and access to resources, programs and opportunities make a different for CSE minorities. Effective DEI approaches identified by this panel include changing the paradigm by adopting teaching models that appeal to a broader demographic; utilizing tactical versus strategic hiring approaches; delivering implicit bias training; partnering with diversity organizations; recognizing that equity fundamentally stems from diversity and inclusion practices; re-tooling to cultivate authenticity; and working together to support creative and inclusive environments.

The Sustainable Research Pathways (SRP) Program, run collaboratively by LBNL and SHI, matches professors working with underrepresented students to students from underrepresented backgrounds. From 2015-2020, >450 applications representing >1,100 students were received. Of the successful applicants, 90% were matched with staff, engaging a total of 80 faculty and approximately 200 students. Of these students, 62% were undergraduates, 35% female, 17% Black/African American, 25% Hispanic or Latino, 40% first generation scholars and 3% had a disability. In surveys evaluating the program's impact on students, 50% of participants indicated that the experience changed their vision of themselves as scientists. Prior to participation, 75% anticipated seeking industry employment. After participation, 25% planned to seek industry employment, with 25% planning to attend graduate school, 13% planning to get a postdoc, 25% unsure, and the remainder having other plans. Aside from changing student career trajectories, the SRP has positive impacts on participating faculty and staff and builds new, productive programmatic collaborations.

The Broader Engagement (BE) program at the SIAM CSE meeting focuses on students and early career professionals from underrepresented populations. This program supports participants through special conference activities including guided affinity groups; a mentor-protégé program; the Diversity and Inclusion Panel; networking events; roommates; opportunities to present research; and tutorials.

Discussion

Giles asked about post-program relationships and mechanisms to help participants maintain connections. How can these programs be scaled, and when will participants join

ASCAC? **Leung** remarked that program participants often develop life-long networks and relationships. While there will be challenges, programs are scalable. SHI has been approached by other labs to implement SRP, and SIAM has requested expansion of the BE program beyond the CSE conference. Some BE participants are CSGF awardees, and this is a pathway to ASCAC for some. **Dolbow** added that rather than seeking one-to-one matches between program participants and supervisors, forming participant research teams that are supervised by post-doctoral scholars could aid in program scalability. With this research structure, participants also experience a research framework that is much closer to lab operations. **Leung** agreed, noting that the BE program's affinity groups assign a single mentor to several participants.

Dolbow appreciated the high percentage of first generation scholars in the SRP program and asked about outreach to or other strategies targeting this community. **Leung** confirmed that SHI deploys intentional outreach strategies. However, defining diversity to include first generation scholars and awarding points to applications, which many other programs do not do, has made the biggest difference. Consequently, white men who are first generation scholars may enter the program; such individuals also need assistance.

Hendrickson asked if the DOE collects and reports diversity statistics from graduate program recipients. **Chalk** said that the internal and external review processes required to collect demographics on a voluntary basis in the Portfolio Analysis and Management System (PAMS) are underway. **Leung** commented that data is critical, and the SHI has provided recommendations to some of the national laboratories after reviewing their data.

Referring to strategic versus tactical hiring, **Berzins** asked how the DOE can recognize potential when conventional metrics are missing or when applicant activities were undertaken in challenging circumstances. **Leung** remarked that metrics may be need to be changed. Rather than looking for someone with a specific technical skill, the DOE can acquire talent and teach particular skills.

UPDATE FROM THE EXASCALE COMPUTING PROJECT, Katie Antypas, Hardware and Integration Director for the ECP, LBNL

To enable ECP national goals, Hardware and Integration (HI) links applications, software and hardware innovations with training, outreach and allocation management through partnerships with DOE facilities. The HI leadership team reflects research and communication goals, including leads on PathForward, Hardware Evaluation, Application Integration at Facilities, Software Deployment at Facilities, Facility Resource Utilization and Training and Productivity.

The PathForward sub-project is completed and has achieved stretch goals. The early access hardware (HW) systems named Arcticus and Spock are available for ECP early users at Aurora (ALCF) and Frontier (OLCF), respectively. Arcticus has 17 nodes with Intel Xe-HP Graphics Processing Units (GPUs). Spock has 12 nodes featuring 1x64 core AMD EPYC Central Processing Units (CPUs), 4x MI100 GPUs with 32 gibibytes (GiB) of High Bandwith Memory (HBM) each and access to OLCF home and project areas. Single node, functionality, performance and software tests in addition to compiling and debugging are ongoing. The Perlmutter system at NERSC is identifying bugs in system software and networking for Frontier and Aurora risk mitigation. The Frontier ECP Test and Development System and the Full User Access System are anticipated in the fourth quarter of FY21 and the third quarter of FY22, respectively. Aurora dates are not yet publicly available.

The Application Integration team is using early HW to aid Application Development (AD) teams in optimization and porting. A staff of 45 Application Integration engineers supports vendors, AD teams and facility staff in implementing algorithms and new architectures, optimizing applications for new HW and software and deploying new programming models. The milestone report *Deploy and Integrate Software Technologies and Applications to Facilities* highlights recent accomplishments in applications and software development as well as contributions from Application Integration engineers. Significant performance boosts have been achieved for a number of codes across different HW iterations, including General Atomic and Molecular Electronic Structure System (GAMESS), ExaBiome software, and NorthWest computational Chemistry for Exascale (NWChemEx) which respectively address computational quantum chemistry, genome assembly and quantum chemistry simulations. Based on feedback, support has been increased for Software Technologies (ST) teams through Application Integration for Math library engagement; Hierarchical Data Format 5 (HDF5) and input-output (I/O); and software libraries at ALCF and OLCF.

The Software Deployment team has converged around the Extreme-scale Scientific Software Stack (E4S) and packaging as a vehicle for incorporating ST products into exascale platforms. Segregating deployment from Continuous Integration (CI) has simplified integration plans. E4S-built pipelines with CI have been deployed at each facility on production systems and software testing on early access hardware is underway.

The Hardware Evaluation team has identified ways to reduce excess data movement in CPU/GPU interactions and is creating new tools. Hardware Evaluation is also set to complete final studies on memory technologies, analytical modeling and network simulation.

The Training and Productivity sub-project has shifted focus to training on new hardware and is partnering with facilities to hold joint training events. Nearly 30 tutorials were offered to the ECP community in 2021 with 2,050 attending. Virtual platforms offer a future means for ACSR-wide training programs.

Discussion

None.

DOE COVID-19 SCIENCE & TECNOLOGY RESPONSE: NATIONAL VIRTUAL BIOTECHNOLOGY LABORATORY, Michelle Buchanan, Senior Technical Advisor, Office of Deputy Director for Science Programs

The NVBL was initiated in March 2020 and funded by the Coronavirus Aid, Relief, and Economic Security (CARES) Act. The NVBL incorporated broad expertise and capabilities across all 17 of the DOE’s national laboratories and collaborated with industry, universities and other federal agencies to support decision makers from local to national levels. The DOE’s High Performance and Leadership Computing Facilities, Light and Neutron Sources and Nanoscale Science Research Centers have been instrumental to the NVBL’s accomplishments.

The NVBL has focused on five topics: addressing supply chain bottlenecks by harnessing advanced manufacturing; generating innovations in testing capabilities; identifying medical therapeutics via computational drug discovery and structural biology; understanding viral fate and transport in the environment; and providing epidemiological modeling and logistical support. To address face mask shortages, the NVBL, Cummins facility and other collaborators created a melt blown carbon media and a novel in-line charging device to enable production of N95 filter media. NIOSH certification was received in October 2020, and DemeTECH is now using the media in mask production. This effort created 600 jobs. A collaboration between national laboratories addressed a shortage in test kits and swabs; work culminated in the design and manufacture processes for COVID test kits, with sterilized preforms for sampling tubes supplied by the Coca-Cola Company. This collaborative work also enabled Thermo Fisher Scientific to produce 8M automated tests kits per week; a $40M factory in Kansas is employing 300 new workers. The Testing and Protocol group is evaluating a device that samples breath for viral infection and is generating a digital microfluidic device that will expedite COVID testing in addition to detection of other diseases via droplet Polymerase Chain Reaction (PCR) methods. User facilities provided pivotal insight into viral structure and vaccine development, and DOE computationally screened billions of compounds against hundreds of viral targets to expedite treatment discovery. Models were generated to predict drug efficacy and safety. DOE also supported the Center for Disease Control (CDC), Federal Drug Administration (FDA) and other agencies in identifying contaminated test kits. Finally, DOE computational tools aided in epidemiological, infrastructure, resource demand and economic modeling to assist decision makers. Simulations also established methods to understand and control risk of viral spread in buildings and other environments.

NVBL achievements through February 2021 are summarized in the *R&D for Rapid Response to COVID-19 Crisis* report. This report will be updated as NVBL funding comes to a close.

Discussion

Gregurick thanked DOE for COVID-19 work done at the HPC Consortium and at the national labs. Many investigators that participated in this effort were NIH-funded, and finding computational resources is nontrivial. The DOE’s ability to provide these resources allowed scientists to quickly reconfigure their research programs. All efforts to continue support through this program are greatly appreciated.

Cerf canvassed reactions to the recent announcement by DeepMind that AlphaFold was able to compute most of the proteins produced from human DNA and expressed hope that the project will be able to share capabilities for other proteome calculations. **Hey** enthused that the database of predicted protein structures could act as a laboratory for understanding biological mechanisms and drug design and discovery. **Reed** agreed that the DeepMind platform is big news. **Mesirov** cautioned that predicted shapes still need to be confirmed. Methods are published in Nature and codes will be made available. It will be important to review the accuracy and validation of results. **Cerf** and **Reed** concurred. **Hey** added that David Baker's group at the University of Washington has produced an open source version of AlphaFold, enabling engagement of the global research community.

OVERVIEW OF COLLABORATION BETWEEN THE DOE, NNSA AND NCI, Emily Greenspan, National Cancer Institute

In 2014, the National Strategic Computing Initiative (NSCI) and Precision Medicine Initiative (PMI) enabled potential for a joint HPC-focused collaboration between the NCI and DOE. Three pilot projects were identified and a five-year MOU was finalized in 2016. Original program governance consisted of the FNL Advisory Committee (FNLAC) NCI-DOE Collaboration Working Group; NCI-DOE Collaboration Governance Review Committee; and FNLAC NCI-DOE Collaboration Task Force. The Collaboration Task Force’s final report from October 2020 recommended continued collaboration with guidance for the selection and review of future projects, as well as the formation of project advisory groups and engagement with the NCI extramural community. Conclusion of Pilot 1 was advised due to insufficient data and research integration. Pilots 2 and 3 will continue. FACA oversight will shift to ASCAC. A Scientific and Technical Advisory Committee will be created for each project. The new NCI-DOE Collaborations Executive Committee will oversee the status of the partnership, funding, program priorities and implementation of ASCAC recommendations. A new five-year MOU, incorporating an automatic five-year renewal every five years, was finalized in June 2021.

JDACS4C, the major program under the partnership, funded three pilot projects: 1) Predictive Modeling for Pre-Clinical Screening; 2) RAS Biology on Membranes; and 3) Population Information Integration, Analysis and Modeling for Precision Cancer Surveillance. The new Innovative Methodologies and New Data for Predictive Oncology Model Evaluation (IMPROVE) project stems from Pilot 1 and aims to develop a robust framework for comparing deep learning cancer drug response models and to execute AI-oriented high-throughput experiments. NCI will fund 3-5 competitively determined investigators to actively collaborate on deep learning models. Pilot 2 current capabilities include reduction of molecular dynamics simulation output with the P2B1 autoencoder; simulation of membrane surfaces with MemSurfer; and identification of key samples in a changing dataset via Dynim. The team will further elucidate activation of an important RAS effector protein named RAF by building on the new Multiscale ML Modeling Infrastructure (MuMMI). To modernize the NCI's Surveillance, Epidemiology and End Results (SEER) program, Pilot 3 utilizes convolutional neural networks, self-attention networks and large-scale Transformer language models for clinical Natural Language Processing (NLP) and further deploys uncertainty quantification (UQ). DOE infrastructure, including Summit, Frontier, ORNL's Knowledge Discovery Instructure and Citadel, have been critical in determining how to protect patient health information (PHI). The seven SEER registries incorporate over 1M unique cancer patients and 3.5M pathology reports. The team will advance deployment to six or more registries this year.

Accelerating Therapeutics for Opportunities in Medicine (ATOM) began in 2017 as a public-private partnership between FNL, the University of California San Francisco, LLNL and GlaxoSmithKline aimed at using AI to accelerate drug discovery. ATOM has recently transitioned to a 501(c)(3) nonprofit organization engaging Argonne National Laboratory (ANL), Brookhaven National Laboratory (BNL) and ORNL via Cooperative Research and Development Agreements (CRADAs) with no direct SC funding.

The CANcer Distributed Learning Environment (CANDLE) is an open source, deep learning platform that accelerates research and scales efficiently on supercomputers to enable exascale capabilities. Current functionalities on NIH's Biowulf supercomputer include hyperparameter optimization of deep learning models. CANDLE is also being used for UQ for some of the pilot computational models. Program benchmarks are available on Github.

The new Predictive Oncology Model and Data Clearinghouse (MoDaC) hosts annotated mathematical models, software and data sets from NCI-DOE collaborations. The platform features a user-friendly web interface; multiple data download options; public REpresentational State Transfer (REST) Application Programming Interface (API) and Digital Object Identifier (DOI) support. A cross-referenced list of all program capabilities and publications is available at https://datascience.cancer.gov/collaborations/nci-doe-capabilities.

In 2019, NCI and DOE established the Envisioning Computational Innovations for Cancer Challenges (ECICC) to advanced new collaborations in the computing and predictive oncology communities addressing four challenge areas: digital twins; adaptive treatment;

synthetic data; and ML-driven hypothesis generation. The online ECICC community platform has >250 members. Workshops advancing digital twin and adaptive treatment via predictive radiation oncology efforts were held in 2020 and 2021. More information is available via the online community platform.

Discussion

Mesirov, a former FNLAC member expressed her pleasure that all pilot projects were evaluated seriously by the Collaboration Task Force. She noted that many were dubious about data availability when Pilot 1 was presented, and it is good that this project has been reformulated through IMPROVE. However, there will still be challenges. Cancer data is complex and not easily predictable because of the stochastic nature of biology. Medical record data is also very messy. The NCI computational biology program has been working on the problem of drug response prediction, termed patient stratification, and extracting meaningful data from medical records for many years. There are an incredible breadth and depth of expertise and useful tools targeted at every one of the challenges described, and the collaboration's effort to engage more of the NCI extramural community will be critical to advancing projects and cross training DOE staff. **Greenspan** remarked that Pilot 3 has many collaborators from academia. The team is also developing privacy-preserving APIs to protect the PHI data used to build the algorithms in collaboration with the CDC and Veterans Affairs (VA). **Mesirov** voiced approval. Every cancer center wants to make use of its data; Epic Electronic Medical Records software helps, but does not solve the problem. It is important to set up trusted data brokers to ensure that only authorized individuals have access. The NIH has been funding many databases for a long time where only trusted brokers have access. Opening collaborations to the NCI and NIH communities at large will be a powerful way to facilitate research.

Gregurick encouraged responses to the NIH RFI on Streamlining Access to Controlled Data from NIH Data Repositories. **Van Hensbergen** requested more information about privacy preserving deep learning NLP models. **Georgia Tourassi** (ORNL) replied that there are multiple publications on the topic. A starting point is the article *Privacy-Preserving Deep Learning NLP Models for Cancer Registries.*

Cerf asked how MoDaC uses digital object identifiers (DOI). **Greenspan** explained that MoDaC assigns DOIs to individual computational models. **Cerf** observed that this is interesting since DOIs are potentially executable.

Cerf inquired if anyone is using ML to detect deleterious drug interactions. **Gregurick** replied that the NIH has FOAs in AI/ML for drug discovery, and some support investigation of deleterious drug interactions. The National Institute on Drug Abuse (NIDA) is also interested in this topic.

NEW CHARGE, Barbara Helland, Associate Director of the Office of Science for Advanced Scientific Computing Research

On July 27, 2021, the SC charged ASCAC to review NCI-DOE collaboration activities under JDACSC4 and to provide advice regarding new opportunities and challenges. The working group will report findings through ASCAC public meetings as needed, but at least once a year. The ASCAC Chair will transmit report findings in letter form to the SC Director following acceptance of the report by the full committee in a public meeting.

Discussion

Helland thanked Berzins who served on previous NCI review process and Hey for agreeing to lead the new working group. Projects 2, 3 and IMPROVE fall under the scope of this charge. ATOM will not be considered because it is now a nonprofit.

Chalk reiterated thanks to Hey. The working group will conduct a high-level evaluation of the program and track overall progress since each project will form their own steering committee to provide detailed insights.

Reed noted that interested parties should reach out to ASCAC. Working group participation requests will be forthcoming.

UPDATE ON FUGAKU, Satoshi Matsuoka, Director of RIKEN Center for Computational Science (R-CCS)

Fugaku is currently the world's fastest supercomputer. Fugaku achieved a Top500-Linpack benchmark Rmax of 0.442 exaflops (EF) and an Rpeak of 0.537 EF. Fugaku's single precision (FP32) peak, half precision (FP16) peak and High Performance Learning-AI (HPL-AI) benchmark performances exceeded 2 EF. Fugaku reached an 11-petaflop (PF) Rmax on nine target applications, a performance approximately 70x higher than that of the K supercomputer. In the exascale era, application performance metrics are most important.

Fugaku's new high performance Arm A64FX CPU was developed via extensive co-design. The server CPU has nearly 160K nodes and is the first with a 7-nanometer (nm) process; Fugaku's CPU operates three times faster than the latest CPUs from US competitors. The Tofu Interconnect D (TofuD) features a 400-gigabits per second (Gbps) class network per node with a 0.5-microsecond latency making Fugaku's server the first with on-die Network Interface Card (NIC) switches. The 6-petabyte/second injection bandwidth corresponds to ten times the aggregate Google, Apple, Facebook, Amazon and Microsoft (GAFAM) traffic. Fugaku boasts the first ultra-scale disaggregated architecture with 48 CPU cores supplemented by 2 or 4 assistant operating system cores per node, HBM2 and NIC are connected via on-chip network with multiple Direct Memory Access Controls (DMACs). Any nodal memory region is accessible by any CPU via the Remote Direct Memory Access (RMDA) and can be injected onto the on-die Level 2 (L2) cache with sub-microsecond latency. The accelerated GPU-like processor has memory enhancing features including 512 bit x 2 Scalable Vector Extensions (SVEs) from Arm and Fujitsu, sector cache, HBM2 and intra-chip barrier synchronization. These attributes confer GPU-like performance for real-world computing, especially for CFD processes. Fugaku's high-efficiency design is power saving; Fugaku operates at 30 megawatts (MW) and is nearly 3.7 times more efficient than the latest x86 processor.

Fugaku implements a converged software stack of approximately 3K applications supported by an open source management tool and an LLNL Spack. Most software on x86 HPC clusters and clouds function after recompiling. A multi-organizational partnership is currently providing cloud services to trial users to evaluate services and identify issues. Fujitsu, RIKEN, Arm and others collaborated on the development of a deep learning software stack that achieved unrivaled speed on the MLPerf HPC Machine Learning Processing Benchmark.

Fugaku's Applications First initiative targeted nine priority sustainable development goals spanning health and medicine; environment and disasters; energy; materials and manufacturing and basic sciences. The FY20-FY22 Fugaku Research Promotion Program builds on priority areas to address grand challenges pertaining to solving the universal problems of mankind and pioneering the future; reinforcing efforts to protect lives and property; enhancing industrial competitiveness; and research infrastructure.

DOE national laboratories currently access Fugaku via the DOE-Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) partnership with continued collaboration at least through March 2022. Many teams have successfully ported code and obtained excellent performance and scalability results. Fugaku joined the US-led COVID-19 HPC Consortium and its simulations aided research in medical therapeutics and societal epidemiology.

Fugaku is central to the RIKEN Center for Computational Science (R-CCS) five pillars of research addressing grand challenges; the "Simulations First" initiative to move Japan towards Society 5.0; the science of convergence of first principle simulations, empirical AI methods and big data instrumentation on large-scale HPC systems; the expansion of workload analysis and increased generality of HPC to the broad IT ecosystem; and the new computing paradigm for the post-Moore era.

Discussion

Hey articulated concerns about Arm being purchased by NVIDIA because Arm's business model does not belong to chip manufacturer. He asked if Fugaku would have been possible if NVIDIA owned Arm. **Matsouka** observed that there are always pros and cons. NVIDIA is a high performance chip company and may compel Arm to have a stronger interest in high performance arenas overlapping with Fugaku, which will benefit HPC overall. Fugaku is willing to work with all interested parties including NVIDIA. The US, UK and Japan are among the friendliest allies in world. **Shalf** (LBNL) noted that Arm is not the only company that offers licensable processor Intellectual Property (IP). Japan previously pivoted from Sparc ISA to Arm. Perhaps a solution is in building using licensable IP blocks as opposed to depending on the specific IP provider.

Van Hensbergen inquired about the potential to expand the collaboration with the US around the continuum of connecting edge computing to smart cities. **Matsouka** commented that the R-CSS Digital Twin and Smart Cities team is very interested in connecting to the edge. Discussions with companies, including NTT and SoftBank, are underway to emulate a smart city within Fugaku. Results could inform traffic parameters, building physics during earthquakes and other topics. However, it will be necessary to emulate information technology (IT) components because smart cities will have hundreds of thousands or even millions of sensors. One advantage of using Arm is that Fugaku can directly emulate an embedded Arduino chip. Since there are 160K nodes, it is possible to run thousands of virtual machine instances with Arm code.

PUBLIC COMMENT

None.

Reed adjourned the meeting at 6:05 pm.

Respectfully submitted August 25, 2021
Holly Holt, PhD
Science Writer, ORISE/ORAU