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

MEETING MINUTES

March 29-30, 2022

Videoconference
ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE

The U.S. Department of Energy (DOE) Advanced Scientific Computing Advisory Committee (ASCAC) convened a Videoconference on Tuesday and Wednesday, March 29-30, 2022 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 Arthur
Keren Bergman
Martin Berzins
Vinton Cerf
Barbara Chapman
Jacqueline Chen
Silvia Crivelli
Jack Dongarra
Timothy Germann
Roscoe Giles
Susan Gregurick
Bruce Hendrickson
Anthony Hey
Alexandra Landsberg
Richard Lethin
Satoshi Matsouka
Jill Mesirov
John Negele
Vivek Sarkar
Krysta Svore

ASCAC Members Absent

John Dolbow
Thom Dunning

Also Participating

Amber Boehnlein, Thomas Jefferson National Accelerator Facility (JLab)
David Brown, Lawrence Berkeley National Laboratory (LBNL)
Richard Carlson, ASCR
Christine Chalk, ASCAC Designated Federal Officer, Oak Ridge Leadership Computing Facility (OLCF), Advanced Scientific Computing Research (ASCR)
Lois Curfman McInnes, Argonne National Laboratory (ANL)
Lori Diachin, Lawrence Livermore National Laboratory (LLNL)
Hal Finkel, ASCR
Ian Foster, ANL
Al Geist, Oak Ridge National Laboratory (ORNL)
Barbara Helland, Office of Science for ASCR
Douglas Kothe, ORNL
Ti Leggett, ANL
Geraldine Richmond, DOE
Irfan Siddiqi, LBNL
Ceren Susut, ASCR

Attending

There were approximately 290 individuals present in total for all or part of the virtual meeting.

OPENING REMARKS, **Reed**, ASCAC Chair, convened the meeting at 11:01 a.m. Eastern Time and welcomed attendees.

The new President's proposed budget was released on March 28, 2022. The 2022 Omnibus Appropriations Bill passed on March 15.

Reed and **Dongarra** are part of a new National Academies of Science (NAS) study on *Post-Exascale Computing for the National Nuclear Security Administration* (NNSA). A recent article about the future of high performance computing (HPC) is available at https://arxiv.org/abs/2203.02544.

Bergman has been named a person to watch by HPC Wire Magazine.

VIEW FROM WASHINGTON, Geraldine Richmond, U.S. Department of Energy Undersecretary for Science and Innovation

ASCR's development and deployment of computational networking capabilities to analyze complex scientific phenomena with big data are pushing DOE research frontiers and contributing to the strategic direction of SC priorities. ASCR's state-of-the-art computing facilities and research program help to secure the nation's competitiveness advantage in high technology and information technology industries. ASCR's technological advances are also tackling climate, a high priority for the DOE Secretary of Energy, while the program's focus on innovation promotes American leadership in microelectronics manufacturing. Richmond frequently highlights ASCR's research accomplishments in daily readouts with SC leadership.

ASCR's unprecedented technological progress is an inspiration to younger generations to pursue Science, Technology, Engineering and Mathematics (STEM) careers. The Computational Sciences Graduate Fellowship (CSGF) provides outstanding benefits to students pursuing doctoral degrees in HPC in addition to building a scientific workforce pipeline. However, DOE must worker harder to diversify its workforce. The promotion and stewardship of Diversity, Equity and Inclusion (DEI) are foundational to advancing scientific innovation and delivering the SC's mission. New SC initiatives like Reaching a New Energy Sciences Workforce (RENEW) will significantly expand training opportunities for undergraduate and graduate students from underrepresented and underserved groups. Such efforts will place greater emphasis on competitive solicitations for financial assistance from Minority Serving Institutions (MSIs) and partnerships with MSIs, including Historically Black Colleges and Universities (HBCUs) as well as increased outreach to institutions underrepresented in the SC's research portfolio. DOE is committed in the strongest possible terms to accelerating efforts to ensure the next generation of scientists and engineers better represent the nation's diversity. All community members are urged to take meaningful steps to address DEI challenges.

DISCUSSION

Speaking hypothetically, **Cerf** sought Richmond's opinion on the best immediate improvement to ASCR. **Richmond** replied such an improvement would be finding that not just ASCR's, but the entire DOE's workforce resembles the nation's composition. Additionally, rather than fearing new technology, the nation would welcome the changes brought by artificial intelligence (AI) and quantum computing. People would realize they do not need understand these capabilities to use them.

Lethin requested more information about DOE's current efforts to address the U.S. microelectronics challenges. **Richmond** stated DOE is working to increase the visibility of this issue in many different sectors. The national laboratories and user facilities deserve recognition for work already completed in this area. DOE welcomes partnerships with other agencies, especially the Office of Commerce and the National Institute of Standards and Technology.

Reed encouraged the community to review the Enacted FY22 Budget.

COMPUTATIONAL SCIENCES GRADUATE FELLOWSHIP UPDATE, David Brown, Lawrence Berkeley National Laboratory

The CSGF is a four-year PhD fellowship supporting the study of science, mathematics, engineering, and computer science using HPC. This workforce development program ensures an adequate supply of scientists and engineers, raises the visibility of careers in computational sciences, provides work experience, and strengthens collaborative ties between academia and the national laboratories. Eligible students can apply for the traditional HPC track, or since 2018, the Mathematics/ Computer Science (Math/CS) track. Fellows pursue an interdisciplinary program of study including at least one 12-week practicum at a national laboratory in areas distinct from their thesis topic and participate in the CSGF Annual Program Review, which provides networking opportunities. Starting in 2019, the Annual Program Review has been held concurrently with the Supercomputing Conference. Students have attended virtually for the last two years due to the coronavirus (COVID-19) pandemic.

CSGF's total annual budget increased to $17M for Fiscal Year 2022 (FY22), with ASCR's support increasing from $10M to $15M per year. The NNSA and the Advanced Simulation and Computing (ASC) program continue to contribute $2M per year. These funding levels represent a historic high. Every fellow's stipend will increase from $38K to $45K beginning in the fall of 2022.

The DOE CSGF Management Partnership and the DOE CSGF Steering Committee, respectively, administer and advise the fellowship program, guided by monitoring activities conducted by the Krell Institute. In 2021, Krell surveyed 56% of present fellows and past alums to evaluate CSGF's quality and investigate diversity and inclusion (D&I). The report indicates CSGF is a highly sought-after fellowship. CSGF actively takes steps to include a diverse set of individuals and works to create an environment in which those from various backgrounds feel supported and their perspectives valued. CGSF positively affected alum careers and publication records. Alums were employed in domestic and international sectors, with most working in industry (38%), academia (28%), and DOE Labs (14%). Fellows and alumni show very high satisfaction with the practicum experience. Current fellows attend 76 different universities. COVID-19 led to temporary practicum suspension in 2020 with most completed virtually in 2021. Onsite practicums are planned for 2022. Krell reported mixed impacts of virtual learning on CSGF experience.

Applicant pool diversity has been strong in recent years, supported by Krell's recruitment efforts targeting diverse communities through conferences, career fairs, and virtual information sessions. The applicant pool's recent gender diversity reflects that of the general scientific community, with near gender parity obtained among fellows. Both CSGF tracks employ a multi-stage downselection to screen CSGF applicants; every application is read by at least two reviewers. An additional screening step was introduced in 2022 to better retain self-identified diversity applicants. CSGF offers to the 43 finalists in 2022 traversed gender, race, and ethnicities. Finalists represented 31 undergraduate institutions, and 22 ranked graduate schools as their first choice for attendance with programs spanning 20 fields of study.

DISCUSSION

Landsberg asked about the large increase in female awardees in the recent cohort, application selection, and practicum concentration at select laboratories. **Brown** said 2022 female applicants were better. Reviewers select applicants, not projects. Research statements are evaluated for articulation of the scientific process and creative ideas. Recommendation letters identify students with aptitude, research experience, and critical thinking skills. Application area diversity is also important. Candidates with HPC knowledge are preferred. There are more NNSA than SC laboratories, and NNSA labs are at least twice the size of many SC labs. Many fellows end up in weapons laboratories where their training and computational science background support the labs' mission. **Jeffrey Hittinger** (LLNL, chat) added the percent of women in the pool increased from ~25% to ~50% while the number of men in the pool dropped from ~66% to ~50% during 2022 downselection.

Hendrickson inquired about program-limiting factors. **Brown** remarked the final 2022 downselection from 42 to 23 traditional track applicants was extremely difficult. There are enough high-quality candidates to double the fellowship size.

Giles sought perspectives on CSGF's success in diversifying its applicant pool while creating a workforce cohort with new capabilities. Can CSGF's model be generalized to other fields? **Brown** stated CSGF was created in 1990 because the national laboratories lacked a capable computational sciences workforce. Krell runs other DOE fellowships that similarly include requirements influencing participant career direction.

James Ang (Pacific Northwest National Laboratory, PNNL; chat) inquired about tracking subfields in Computer Science (CS). **Reed** (chat) stated CS fellows are not tracked at that level of detail.

Berzins asked how CSGF assesses applicants' future potential when their present performance is limited by circumstances. Some never have opportunities to fulfill application criteria, yet will perform well if given the chance. These individuals are missed in paper-based systems. **Brown** agreed this is a challenge. Applicants submit practical experiences, publications, and leadership skills, including in non-research areas. Reviewers synthesize information and evaluate whether applicants have taken advantage of the opportunities available to them. There is no rubric for acceptance. **Hittinger** clarified applications are not scored but evaluated holistically. Reviewers try to understand personal circumstances, incorporate diversity, and take risks on high-quality candidates who may have had fewer opportunities. **Reed** said the National Science Foundation (NSF) is considering socioeconomic factors as well as other parameters to aid in program diversification.

Helland pointed to CSGF support for fellows. The Steering Committee talks to new fellows and introduces expectations. Many alums sit on the Committee and engage in mentoring activities. **Brown** explained Krell maintains a close relationship will fellows for the duration of their award. Krell and the Steering Committee guide fellows through topic or course changes and offer flexibilities for students encountering life circumstances. The Annual Review provides fellows with networking opportunities with each other, the Steering Committee, and the scientific community.

Landsberg requested information about the Review Committee. **Brown** said Committee consistency is important; a portion of members is changed every two years. The first screening step engages >40 reviewers, most of whom are alums. This year, 11 of the 12 Committee members were alums. Diversity is sought in Committee topical expertise and across other axes.

THE EXASCALE COMPUTING PROJECT'S BROADER ENGAGEMENT INITIATIVE, Lois Curfman McInnes, Argonne National Laboratory

With the expansion of scientific computing, ASCR faces urgent workforce challenges. Without intervention, forecasts predict a growing workforce gap and continued underrepresentation of women and minorities. Based on recommendations from the *Transitioning ASCR after ECP* report, an Exascale Computing Project (ECP) Task Force on Broader Engagement was formed in August 2021. The Task Force engages several national laboratories; the Argonne Leadership Computing Facility (ALCF), the Oak Ridge Leadership Computing Facility (OLCF), and the National Energy Research Scientific Computing (NERSC) Center; and is partnering with the Sustainable Horizons Institute (SHI). The Task Force's mission is to establish a sustainable plan to recruit and retain a diverse workforce in the DOE HPC community by fostering an inclusive and supportive culture within computing science and the DOE national laboratories. Efforts will engage talented people from underrepresented groups, raise awareness of DOE activities and needs, and provide pathways for community interactions. Actions will leverage ongoing activities in the SHI, DOE-wide and laboratory-specific programs, and local and regional HPC communities.

Strategies to promote HPC workforce development and retention center on influencing laboratory and community culture. Assembly of an HPC Workforce Development and Retention (HPC-WDR) Action Group is underway. The Action Group will host HPC community meetings to share best practices and ideas, develop key initiatives, and generate recommendations and strategies. Creation of a complementary website curating workforce development and retention resources is in progress. Additionally, a new best practices webinar series, scheduled to begin May 24, 2022, will engage speakers from the DOE and the broader community.

Basic HPC is not typically taught at early stages of students' careers, leading to gaps in the worker pipeline since HPC skills are prerequisites for many existing advanced computing programs and HPC careers. A two-pronged strategy will meet students where they are, opening pathways to HPC careers. The Direct-to-Student Approach will leverage national laboratory resources to provide an intensive, two-week course in HPC/ AI targeting junior and senior undergraduates, early graduate students, and underrepresented groups. Materials will be made freely available to the community. Guided by listening sessions, the University Approach will build capabilities and programs at MSIs and EPSCoR institutions. Phase-1 project funding was approved in December 2021, and activities supporting both approaches are ongoing.

SHI's Sustainable Research Pathways program builds research collaborations between students from underrepresented backgrounds, faculty working with underrepresented students, and DOE HPC laboratories. This mentoring and research program has grown each year since its 2015 inception at LBNL and is expanding this year to include a multi-lab ECP partnership. Out of the 64 offers issued by the 2022 Sustainable Research Pathways for HPC (SRP-HPC) program, 57 individuals have accepted representing 34 matches between students or student and faculty teams and research laboratories.

With ECP concluding in 2023, ECP will be seeking sustainable ways to support efforts and partner with the HPC community. More information is available at https://www.exascaleproject.org/hpc-workforce/.

DISCUSSION

Bergman appreciated ECP efforts to broaden opportunities and address educational gaps for undergraduate and graduate students. University computer science departments are expanding rapidly, but HPC educational components are not yet clear. There is a gap during critical educational periods for undergraduates and perhaps earlier. **Curfman McInnes** stated the Task Force is passionate about leveraging community resources to provide students with pathways to advanced HPC programs.

Chapman strongly endorsed outreach to MSIs where there is a lot of untapped potential and perhaps lack of awareness of HPC opportunities. Retention remains an issue requiring Task Force attention. **Curfman McInnes** commented there is no easy solution to retention; the Task Force created the HPC-WDR Action Group to begin addressing this issue. Many in the community care deeply about this topic and are working together towards solutions.

Giles asked about partnerships with outside entities and tracking programmatic impact. **Curfman McInnes** explained the Task Force has not had time to consider outside partnerships. There is interest in working with the community to identify a future, sustainable path. SHI is already tracking program impacts on student and faculty participants; those accepting SRP-HPC offers agree to remain in contact with SHI after the program.

Crivelli remarked reaching out and presenting to high schools, MSIs, and HBCUs is a simple way to have an impact. Such experiences are rewarding and do not require large effort. **Curfman McInnes** agreed outreach is important. There are opportunities to capture the interest of early career individuals in new and compelling ways.

Cerf suggested connecting with the American Institute for Physics (AIP) National Task Force to Elevate African Americans in Undergraduate Physics (TEAM-UP), which aims to double the number of black PhDs in physics by 2030.

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

Helland reviewed ASCR personnel changes and retirements as well as current COVID-19 workplace policies.

SC FY22 Enacted Appropriations provide ≥\$120M, \$2M and \$245, respectively, for artificial intelligence and machine learning (AI/ ML), collaboration with the National Institutes of Health (NIH), and quantum information science (QIS). The QIS figure includes ≥\$120M and \$125M, respectively, for research and the National Quantum Information Science Research Centers (NQISRCs).

ASCR's FY22 Enacted Appropriations designate \$286M for Research, with ~\$51M for Applied Mathematics, including ~\$24M for AI and big data; ~\$50M for Computer Science, including ~\$16M for AI and big data and ~\$7M for QIS; and ~\$79M for Computational Partnerships, including \$19M for AI and big data and \$14M for QIS. The ~\$106M Research and Evaluation Prototypes budget includes \$15M for CSGF and ~\$78M for QIS. The \$620M HPC and Network Facilities budget allocates \$118M for NERSC and \$2M for the High Performance Production Facility; \$160M for the ALCF; \$250M for the OLCF; and \$90M for the Energy Sciences Network (ESNet) Facilities and Testbeds. In addition to funds designated within the ALCF and OLCF budgets, an additional \$129K is allotted to the ECP.

ASCR continues rollout of the fifth Scientific Discovery through Advanced Computing (SciDAC) cycle and is planning to re-compete the Mathematical Multifaceted Integrated Capability Centers (MMICCs). In partnership with other SC programs, ASCR has supported the NQISRCs. ASCR has also maintained basic research investments in quantum computing and communications while supporting testbeds. ASCR continues promoting the use of AI/ ML in scientific simulations and data-intensive applications to accelerate discovery.

ASCR 2022 workshops addressed parallel discrete event simulation (PDES); cybersecurity and privacy; scientific software development and use; and data management, storage, and visualization towards transitioning ECP efforts back to core research areas. The SC's Biopreparedness Research Virtual Environment (BraVE) Roundtable gathered information on the unique role SC can play in basic science research and development (R&D) for future pandemics and related crises.

ASCR facilities continued optimal operations in 2022, delivering HPC resources to >10K users and network operations to all DOE laboratories and sites. ALCF-3 (Aurora), NERSC-9 (Perlmutter), and OLCF-5 (Frontier) facility upgrades are at Critical Decision-2 (CD-2). CD-1 planning for NERSC-10 and the High Performance Data Facility (HPDF) continues.

The SC-wide FY22 RENEW initiative aims to increase STEM participation of historically underrepresented individuals and institutions. Based on listening sessions with MSIs and the National Strategic Plan for Quantum Information Science and Technology (QIST) Workforce Development, ASCR will organize networking events and support partnerships among DOE labs, underserved institutions and other entities to 1) strengthen or establish new quantum computing and network research collaborations; 2) expand undergraduate and graduate student opportunities in quantum computing or networking; and 3) facilitate access to infrastructure. Efforts will be coordinated with other initiatives such as NSF's Expanding Capacity in Quantum Information Science (ExpandQISE) program.

The FY23 President's Budget Request (PBR) of ~$7.8B for the SC and ~$1.1B for ASCR represent ~4% and ~3% increases, respectively, over FY22 Enacted Appropriations. Ongoing SC initiatives continue with support for RENEW ($30M); BRaVE (~$52M); QIS (~$293M); AI/ ML ($169M); Exascale Computing ($268M); and Advanced Computing (~$38M). New FY23 SC initiatives include SC Energy Earthshots ($204M); Funding for Accelerated Inclusive Research (FAIR) ($36M); and Accelerate Innovations in Emerging Technologies (ACCELERTE) ($40M).

SC Energy Earthshot funding will launch the Energy Earthshot Research Centers (EERCs). These multi-investigator and multi-disciplinary teams will be complemented by small group awards addressing knowledge gaps. ASCR will use its ~$50M FY23 budget to partner with Basic Energy Sciences (BES) and Biological and Environmental Research (BER) to establish EERCs in coordination with the applied technology offices. ASCR will also invest in small group awards in applied math and computer science incorporating AI/ ML.

FAIR will complement RENEW in building research capacity, infrastructure, and expertise at MSIs by 1) developing relationships between MSIs, national laboratories, and user facilities; 2) providing support for single Primary Investigators (PIs), research teams, and equipment and infrastructure; and 3) delivering the majority of funds to MSIs with a smaller portion to partnering institutions. ASCR will invest $4M in FY23 funds in midrange equipment to broaden participation in QIS research programs and establish SciDAC-like partnerships in underserved communities.

ACCELERATE will drive discovery of sustainable production methods for new technologies; train a STEM workforce to support future industries; and meet the nation's clean energy and security needs. ASCR will use $5M in FY23 funds to support development of new AI/ ML, data analysis, and visualization tools to accelerate the design and deployment of clean energy technologies and promote energy-efficient computing.

ASCR's FY23 PBR seeks ~$379M for Mathematical, Computational, and Computer Sciences Research; ~$613M for HPC and Network Facilities; and $77M for the SC ECP. The

PBR prioritizes foundational research across ASCR to transition and sustain critical ECP technologies. Core research investments will continue to address heterogeneous architectures; changing ways HPC systems and data storage are used; development of new scalable energy efficient algorithms and software; and new data management and visualization tools. A requested $13M increase within FY23 funds will continue incorporation of AI/ ML into simulations and data-intensive applications while fostering greater connectivity with distributed resources. A requested $6M increase will support the DOE-National Cancer Institute (DOE-NCI) partnership as well as continue efforts to deliver rapid and robust responses to public health emergencies. ASCR will maintain full support for basic research investments in quantum computing, communications, internet testbeds, and the NQISRCs.

The FY23 PBR allows for continued optimal ASCR facility operations. ALCF-3 will continue acceptance testing while NERS-9 will complete Phase 2 acceptance and continue NERSC-10 planning. OLCF will complete OLCF-5. HPDF will continue planning for CD-1. In partnership with the Research program, ASCR Facilities will initiate select actions outlined in the FY22 *SC Integrated Research Infrastructure Architecture Blueprint Activity.*

DOE laboratories will issue a joint Request for Information (RFI) in the third quarter of FY22 to solicit input from >35 computing technology vendors on their capabilities to meet DOE requirements for future systems. Responses will drive strategic planning.

DISCUSSION

Reed asked about inflation. **Helland** commented labs are experiencing inflation effects, which may be addressed in the FY24 budget. Fixed-price contracts at facilities are beneficial.

Hendrickson inquired about increasing average grant sizes to compensate for rising labor costs within the labs. **Helland** indicated the FY22 budget does not have funds to address rising labor costs; these may be addressed in FY23. To stem ongoing employee loss to computing companies, DOE is considering mid-term raises. Solicitations may also need to consider salary increases. **Reed** agreed.

Hey commented FAIR is already widely used in the context of data fairness for Findable, Accessible, Interoperable and Reusable. Do FY22 and FY23 budgets enable a coherent, ASCR-wide AI/ ML strategy? Are there possibilities for NSF or international collaborations? **Helland** said such a cohesive effort, when initiated, would need to be SC-wide. The FY22 SC budget allocates $120M for AI/ ML. ASCR will receive ~$59M in FY22 and ~$70M in FY23 for AI/ ML. Facilities are planning to revisit the 2019 *AI for Science* report to envision possible next steps. There are ongoing discussions about NSF collaborations. A QIS joint oversight group was established in 2021, creating opportunities for international collaborations. This Administration is focused on building international partnerships.

Cerf (chat) raised microgrid demonstrations. **Helland** (chat) replied ASCR recently joined the DOE Grid Modernization Task Force and is learning about demonstration plans.

NATIONAL QIS RESEARCH CENTERS: MISSION AND ACTIVITIES, Irfan Siddiqi, Lawrence Berkeley National Laboratory

Five NQISRCs, each led by a national laboratory, were launched in 2020: Co-design Center for Quantum Advantage (C^2QA); Next Generation Quantum Science and Engineering (Q-NEXT); Quantum Systems Accelerator (QSA); Quantum Science Center (QSC); and Superconducting Quantum Materials and Systems Center (SQMS). Each Center represents a partnership of labs, universities and industry. NQISRCs take distinct but complementary approaches to tackling major challenges in advanced materials for quantum technologies; entanglement distribution networks; high-performance instruments and sensors; and full-stack quantum computation. Solutions address the entire QIS innovation chain for emerging technologies. An Executive Council coordinates activities across Centers.

The NQISRCs contribute to QIS ecosystem stewardship by 1) supporting workforce development programs aimed at broadening and diversifying the workforce; 2) broadly engaging with industry partners to accelerate deployment of quantum-enabled technologies; and 3) creating new community synergies with DOE programs and user facilities. In their first year of operation, the Centers jointly conducted activities in technical coordination, facility instrumentation, workforce development, cross-center management, and outreach. Future plans address continued coordination in technical areas, instrumentation and facilities, and ecosystem stewardship. Implementation of a QIS summer school is planned for 2023.

NQISRC science highlights include synthesis of tunable molecular color centers from Q-NEXT; investigation of a spin system's quantum phases using a programmable quantum simulator by QSA; discovery of room-temperature, single-photon emitters in SiN by QSC; generation of an efficient, fully-coherent Hamiltonian simulation by C^2QA; and discovery of niobium nanohydride precipitates in superconducting transmon qubits by SQMS.

DISCUSSION

Cerf (chat) asked about an acronym. **Svore** (chat) defined VGE as Variational Quantum Eigensolver.

Gregurick commented on the level of quantum computing technological readiness and impact in biophysics and molecular biology. **Siddiqi** advised it is difficult to put a timeline on innovation. Quantum technologies could evolve rapidly and will undoubtedly impact the world, but the way in which they will do so is not yet known. New use case will require consideration of physical, structural, and computational issues when selecting quantum algorithms.

Svore (chat) drew attention to fault tolerance. **Siddiqi** agreed fault tolerance is important and can be accomplished by creating a noiseless environment or a system tolerant of noise. Ideal systems will be robust to both noise processes.

Hey asked about coherent connection of ion traps. **Siddiqi** said remote entanglement has been demonstrated on most platforms; distributed quantum computing is under development. There are several different architectures for ion traps.

COMMUNITY OF INTEREST UNCONFERENCE, Ian Foster, Argonne National Laboratory; Richard Carlson, Office of Advanced Scientific Computing Research; and Amber Boehnlein, Jefferson National Accelerator Laboratory

The *Envisioning Science in 2050* report from the Community of Interest Workshop on Future Scientific Methodologies is under review. Held in November 2020, this virtual workshop engaged ~150 participants over three non-consecutive days in speculating about possible DOE computing futures while remaining grounded in today's reality. Following a pre-meeting process to gather ideas, curated topics were presented as illustrative vignettes to stimulate creative group discussion. Nine scenarios for possible futures were explored: 1) Smart self-driving facilities automate experimentation; 2) A representative workforce turbocharges the labs; 3) A universal theory machine transforms science; 4) Linked facilities form a discovery cloud; 5) New missions drive new modes of working; 6) A universal data service make all data easily accessible; 7) Science in a world of purpose-built computers; 8) Every system has its doppelganger; and 9)

Delphi, a universal knowledge map. Participants identified the implications and possible consequences of new technologies or methodologies, signposts representing progress, and keys to addressing possible pitfalls. Material captured from each scenario may form the basis for more focused Basic Research Needs (BRN) workshops.

DISCUSSION

Gregurick complimented the organizers on the unconference's success.

Arthur inquired about documenting predictions' assumptions and likelihoods to track unknowns. **Foster** commented daily guidelines moved participants from ideas to signposts to success. While not every detail was documented, groups tracked assumptions using a range of technologies like Google Documents. Attendees were encouraged to imagine what would happen if some things were possible. **Boehnlein** added, unlike typical unconferences, this meeting set expectations through framing. Processes were used to refine starting points, and group notes were checked for some degree of a ground strap. **Carlson** said daily activities provided structure for participants to build on new technology ideas while remaining grounded in reality.

Reed asked how participants were encouraged to think about the future while remaining unencumbered by current constraints. Sometimes individuals outside of the field are better at identifying new possibilities. **Boehnlein** remarked this was difficult. Participants are used to providing a different kind of BRN input. Breaks between meeting days gave organizers the chance to reframe content. Facilitators had extra training to encourage interactions. **Foster** observed participants were selected from diverse disciplines. Questions were used to prompt creative thinking. **Carlson** credited the meeting's structure. There was only one plenary session on the first day, and there were no end-of-the-day report-outs. Small groups fostered discussion.

David Coster (Max Plank Institute for Plasma Physics) speculated the meeting benefited from its remote format, which gave participants breaks between meeting days. **Foster** agreed, adding the remote format also enabled fluid movement among the many small breakout sessions. Due to COVID-19 shut downs, participants were also eager to interact.

Landsberg posed a question about traditional versus unconference processes. **Foster** said the meeting's goal was to consider different future scenarios, not to determine the future. This lead to selection of the unconference process. **Boehnlein** noted unconference methods can be complimented by traditional processes such as surveys and interviews. There may be opportunities for applying traditional processes after the report is released.

Hey contrasted unconference scenarios with other groups' research goals, such as designing robots that could defeat a World Cup soccer team by 2050 or designing an AI system that could receive a Nobel Prize by 2015. **Foster** personally hoped humans and robots would be working together by 2040.

Giles (chat) asked if the 30-year timeframe was limiting. **Foster** (chat) clarified the exact timeframe was not the focus; the goal was to keep attendees thinking five years into the future.

UPDATE FROM THE EXASCALE COMPUTING PROJECT, Douglas Kothe, Oak Ridge National Laboratory and Lori Diachin, Lawrence Livermore National Laboratory

Baselined in FY20, the ECP is nearing CD-4 in FY24. ECP received positive responses on all 2022 Independent Project Review (IPR) questions. Following 2021 IPR recommendations, ECP developed a *Contingency Management Plan* with DOE and is preparing for a successful project conclusion through ECP's new *End of Project Management Plan, IT Tool Handbook* and existing *Transition to Operations Research* document. These materials and ongoing activities address sustained support for 1) maintaining research teams on the exascale platform; 2) transitioning ECP applications and software to the broader community; and 3) documenting tools and project management practices. Evolving plans to support the software ecosystem post ECP envision a core software center administering sprint-like campaigns building on ECP's Extreme-Scale Scientific Software Stack (E4S).

The 2019 ECP Final Design Review set four Key Performance Parameters (KPPs). KPP-1 and KPP-2, respectively, address the success of ECP applications teams in demonstrating challenge problems designated as mission critical or broadening the reach of exascale science and mission capabilities. KPP-3 measures Software Technology (ST) and Co-Design (CD) projects integration and creation of a productive and sustainable environment. Evaluation spans parameters related to clients, tool usage, facilities development, community ecosystem, and vendor development. KPP-4 appraises PathForward milestones.

The multi-stage, rolling KPP review process requires approval from subject matter experts and ultimately the Federal Project Director. For KPP-1 and KPP-2, applications teams are providing reports defining challenge problems, key performance steps, and figures of merit to be used in evaluation. All ST and CD teams have defined KPP-3 integration strategies. ECP's living dependence database is a key information source for managing integration points and KPP planning. KPP verification is a critical activity for the next 18 months, and emerging risks traversing compiler bugs to staff retention are being actively monitored or mitigated.

Crusher, Frontier's Test and Deployment System (TDS), was opened to ECP users in November 2021 and will be available until ECP conclusion. Diverse applications representing KPP-1 or KPP-2 challenge problems have demonstrated improved performance on Crusher. ECP's application portfolio annual review is in draft report format, and quarterly release of E4S and Software Development Kits are providing community value. Access to the full Frontier system and El Capitan early hardware are on track for FY22, followed by access to Aurora TDS and the full Aurora system in FY23. Full El Capitan system access is scheduled for FY24.

DISCUSSION

Berzins (chat) asked about OpenMP offload and emphasized the importance of portability across GPU architectures. **Diachin** (chat) responded OpenMP is very important to Fortran codes. Many C and C++ codes use programming abstractions, such as Kokkos and RAJA, along with HIP, CUDA, and SYCL. **Kothe** stressed OpenMP's importance to sustainability. Abstraction layers are popular and now also have an OpenMP option.

Lethin (chat) inquired about code abstraction and performance. **Diachin** explained HIP, CUDA, and SYCL can be abstracted using Kokkos and RAJA. Several applications have found Kokkos and RAJA abstractions to be competitive when compared to those of CUDA and HIP, though results are application dependent. **Kothe** noted QMCPACK is one of several applications addressing GPU performance.

Giles valued how KPP efforts benefit stakeholders and a deeper understanding of systems. Are additional workforce challenges anticipated over the next year? **Kothe** responded KPP verification is not only about 413.3B requirements, but also community performance in reproducibility and verification with SME vetting before transitioning to Frontier. Opportunity investments regarding workforce sustainability can be tapped to increase the probability of uptake and sustainment beyond KPP metrics. **Diachin** added there is a direct line of sight from stakeholder mission-critical science needs to ECP challenge problems. These new capabilities will have broader implications for DOE and the nation than just meeting the KPPs.

PUBLIC COMMENT

Sarkar called attention to possible synergies with Schmidt Futures, which has launched the Virtual Institutes for Scientific Software initiative to help manage and sustain scientific software used for research and scientific applications. **Kothe** suggested engaging via monthly town hall meetings held through the Leadership Scientific Software portal.

Reed dismissed the meeting for the day at 3:30 p.m.

THURSDAY, MARCH 30, 2022

OPENING REMARKS, **Reed** convened the meeting at 11:01 a.m. and led other attendees in congratulating **Dongarra** on receipt of the 2021 A.M. Turing Award from the Association of Computing Machinery.

POLARIS: A SCALABLE TESTBED TOWARDS AURORA, Ti Leggett, Argonne National Laboratory

Polaris follows Crux as the second testbed bridging Theta and Aurora at the ALCF. Installation of the 44-PF Polaris system began in 2021 to provide select technologies and architectures similar to that of Aurora for testing. Polaris uses the HPE Apollo Gen10+ platform. Each of the 560 nodes comprises one AMD EPYC Rome Central Processing Unit (CPU) and four NVIDIA A100 GPUs with Unified Memory Access (UMA), two fabric endpoints, and two Non-Volatile Memory Express Solid State Drives (NVMe SSDs). GPU architecture additionally employs High Bandwidth Memory (HBM). CPU-GPU and GPU-GPU interconnects utilize peripheral component interconnect express (PCIe) and NVIDA Link (NVLink), respectively. The HPE Slingshot 10 system interconnect features dragonfly topology with adaptive routing, and the network switch delivers 25.6 terabytes per second (TB/s) per switch. Polaris has an aggregate 368-TB memory connected to existing ALCF storage resources, including the Grand, Eagle and Home file systems. Later this year, Polaris CPUs will be upgrades to AMD Milan models and the high-speed network (HSN) to HPE Slingshot 11. Polaris successfully completed acceptance testing last week and full user access is scheduled for the summer of 2022.

The ≥2 exaflop (EF) Aurora system is currently being delivered and installed. Aurora uses the HPE Cray-Ex platform and will contain >9K nodes, each comprising two Sapphire Rapids with HBM Intel Xeon CPUs and six Ponte Vecchio Intel GPUs with UMA and eight fabric endpoints. The tile-based chiplet GPU architecture employs HBM and Foveros 3D Integration. Aurora will use the same HSN and network switch as Polaris. The system will have ≥10 petabytes (PB) of aggregate memory and 220 PB of high-performance storage with ≥ 25 TB/s of decentralized autonomous organizations (DAOs). Since Intel architectures cannot be tested at scale on Polaris, other testbeds have been provided to early Aurora users.

ECP and ALCF Early Science Program (ESP) research groups will have early access to Polaris. Polaris will lead the expansion of HPC activities from those focused primarily on model-based simulation to those including data and learning paradigms. Thus, ESP projects incorporate simulation, data, and learning components as well as select workflows. Though there is not one-to-one component overlap, Polaris and Aurora will share High Performance Cluster Management system software; MPI, OpenMP, DPC++, Kokkos, RAJA, and HIP programming models; HPE Cray MPI and MPICH for message passing interfaces; the HPE Slingshot 11 system interconnect; and DL frameworks, the Cray AI stack, Python, Numba, Spakr and Containers for data and learning. Polaris uses cu* from CUDA math library while Aurora will use oneAPI.

ALCF is investigating ways to provide new services that enhance workflows at facilities by providing users with faster feedback during experiments.

DISCUSSION

Cerf asked whether the ALCF influences manufacturer chip design and about the impact of the increasingly heterogeneous computing base on writing compilers. **Leggett** explained Polaris used as many existing technologies as possible due to the short timeline. Feedback will be provided to vendors for Aurora. Polaris's design allows for future introduction of heterogeneity; future compilers and applications must be able to manage diverse architectures. Intel's oneAPI aims to support different hardware components.

Germann raised concerns about bridging from Polaris's CUDA math library to Aurora's oneAPI. **Leggett** explained no Intel products were available for scaling on Polaris. ECP and ESP users have access to smaller testbeds and other platforms for testing oneAPI libraries on Intel hardware. The concern is not the libraries, but how the underlying kernels interact with the other components, such as Slingshot.

Berzins stressed the importance of RAJA and Kokkos; DOE exascale machines rely heavily on these codes. DPC++ may not be as useful a tool, especially when considering portability. **Leggett** explained Kokkos and RAJA teams have access to Intel platforms, with support activities continuing for Slingshot and NVIDIA. HIP will be a first step for porting over to the new platform. **Reed** opined heterogeneity and maintaining software portability while still being performance-driven will be important issues over the next few years.

REPORT FROM THE COV, ASCAC DISCUSSION AND VOTE ON COV REPORT, Alexandra Landsberg, COV Chair

The ASCR Research Committee of Visitors (COV) met virtually in August 2021 to review the Applied Mathematics, Computer Science, Computational Partnerships, and Research and Evaluation Prototypes programs for FY16-FY19.

Key findings pointed to 1) The impact of the ECP as well as new AI/ ML and QIS efforts to ASCR's research portfolio, related funding cuts, and the need for a holistic budget plan; 2) The need for clarity in how programmatic shifts are made and communicated; 3) The positive impact of the Early Career Research Program (ECRP) on awardees' careers but the negative impact of the congressional mandate to fully fund awards <$1M on award numbers; 4) The need to increase use of preproposals; and 5) The COV's need for additional time to digest dense presentations and help in navigating Portfolio Analysis and Management Software.

Key recommendations suggest that ASCR Research 1) Identify and document their North Star with a clear vision, mission statement, and five-year plan; 2) Develop procedures to better communicate the impact of programmatic shifts; 3) Investigate strategies to identify early- to early-mid-career researchers with significant promise and ways to enable them to develop into PIs of large DOE projects; 4) Implement a preproposal process to reduce the community burden of writing and reviewing proposals with little chance of being funded; and 5) Provide solicitation summary statistics in COV presentations to facilitate evaluation.

Applied Math findings addressed limited funding, solicitations, awards, portfolio elements, and quantum testbeds. Comments touched on ECRP awardees, navigation of funding challenges, program management, and emerging research areas. Recommendations discussed use of preproposals, mechanisms to diversify PIs including re-establishment of small university groups, effectiveness measures for math centers and long-term lab projects, and exploration of new emerging research areas.

Computer Science findings referred to solicitations, awards, program management, limited funding, university PI opportunities, workshops, and portfolio elements. Comments dealt with unsolicited awards, R&D risk, and workshop fatigue. Recommendations considered communication, PI diversification, emerging technologies, and success targets.

Computational Partnerships findings called attention to solicitation workload, the reviewer pool, program management, cross-disciplinary partnerships and SciDAC's impact, and software migration to exascale platforms. Comments referred to the SciDAC-4 Coordination Committee and documentation of feedback loops from partnerships to ASCR base research programs. Recommendations considered preproposals, communication of SciDAC strategic goals, and external SciDAC review.

Research and Evaluation Prototypes findings discussed solicitations, program development and monitoring, awards, and workforce development. Comments remarked on award recipients, program flexibility, and CSGF. Recommendations dealt with quantum testbed utilization and CSGF diversity.

DISCUSSION AND VOTE

Cerf inquired if recommendations materialized in 2019-2022. **Landsberg** explained the review period was from FY16-FY19. ASCR has implemented use of preproposals, and the budget has grown from $30M to $50M. **Helland** said targeted funding opportunity announcements (FOAs) and preproposals are helping to prevent reviewer burnout. **Susut** added the SC FOAs now outline the preproposal competitiveness process. **Arthur** (chat) remarked the preproposal system has worked well.

Chen asked if authors of rejected preproposals receive reviewer feedback. **Susut** confirmed, and **Chen, Landsberg, and Reed** approved.

Germann raised restoration of funding to single PIs and small groups of investigators. **Helland** commented with increased funding, additional FOAs may be released, and there will be increased focus on outreach to underserved institutions.

Mesirov posed a question about the lifetime of large DOE centers. **Landsberg** remarked math centers were originally stood up for five years followed by an open re-competition for an additional five years of funding. **Susut** explained continuing programs like MMICCs and SciDAC institutes typically offer five years of funding per cycle. Re-competitions are open to all, including existing PIs. **Helland** added the quantum centers received an initial five years of funding and will have a re-competition opportunity. Centers may be sunset after ten years. Five-year re-competitions allow funding of new ideas.

Lethin asked about diversity of review panel and workshop participants, including academic and industry involvement. **Landsberg** explained workshop participation was not examined. Spot-checking review panels indicated new academic reviewers were brought in, including international reviewers for SciDAC. Lower engagement of industry reviewers suggests industry is potentially an untapped resource for ASCR reviewers.

Reed and Helland appreciated the COV's work.

All present ASCAC members voted to approve the draft COV report. ASCR will issue a program response in 30 days.

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

The SC has charged ASCAC with forming an International Benchmarking Subcommittee to answer four topics related to 1) maintenance of critical international cooperation in an increasingly competitive environment for both talent and resources; 2) identification of key areas where the U.S. has, or could aspire to, leadership roles in advanced computing and high-end computational science and engineering; 3) identification of technical areas or capabilities where additional emphasis within reasonable constraints or through partnerships could foster U.S. leadership roles; and 4) how programs and facilities can be structured to retain talented individuals and reduce barriers to successfully advance the careers of scientific and technical personnel in advanced computing, computational sciences, and engineering and related fields. A final report is due in the spring of 2023.

DISCUSSION

Cerf raised the possibility of collaborating on the charge with other organizations seeking to advance STEM capacity, such as AIP's TEAM-UP program. **Helland** said this is a possibility and will examine the subcommittee composition of other programs that have already begun addressing or have addressed this charge.

Reed will begin assembling a subcommittee.

UPDATE ON FRONTIER EXACALE SYSTEM AND EARLY SCIENCE, Al Geist, Oak Ridge National Laboratory.

ORNL's journey from petascale to exascale computing began with the 2.3-PF Jaguar system in 2009, progressed to the 27-PF Titan system in 2012 and the 200-PF Summit system in 2017, and recently culminated in the 2-EF Frontier system in 2021.

Frontier's 74 Olympus computer racks occupy 4K square feet and contain a total of 9,408 nodes. Each rack holds 128 nodes containing one AMD Trento CPU with 512 Gibibytes (GiB) of double-data rate fourth-generation synchronous dynamic random-access memory **(**DDR4) memory, four AMD MI250X GPUs with 128 GiB HBM per GPU, four TB of non-volatile memory (NVM), and four Cassini NICs. There is coherent memory across each node, and the GPUs and CPUs are fully connected via AMD Infinity Fabric. Overall, Frontier features 9.2 PB of memory with 37 PB of local node storage and 716 PB of center-wide storage. The system's Cray Slingshot network has dragonfly topology. Compared to that of Jaguar, Frontier's 15-megawatt/EF operations has realized a >200x reduction in energy requirements due to ASCR investments in vendors targeting reduced power consumption by computing chips and use of a water cooling system in Frontier. Frontier uses and builds on Summit's successful architecture, which excels at advanced simulations, high-performance data analytics, and AI. Furthermore, Frontier's multi-tier storage system offers users 37 PB of local storage capable of 11B input/output operations per second (IOPS) and mechanisms to prevent storage system aging.

During Frontier's build, the chip shortage made acquiring parts challenging. Order lead times increased by six to 12 months, and delivery commitments were broken weekly due to supply chain issues. ORNL worked with ACSCR to obtain a Defense Property Accountability System (DPAS) rating for Frontier that helped prioritize U.S. part orders, and HPE's size was leveraged to negotiate with suppliers. The last of the 60M parts required for Frontier arrived in October 2021. System testing and stabilization are ongoing.

To prepare for Frontier access, early science teams in the Center for Accelerated Application Readiness (CAAR) and ECP received access to Crusher in November 2021. The TDS contains two cabinets of Frontier hardware comprising 192 nodes; the system has Slingshot 11, Cassini NICs, and the same software environment as Frontier. All ECP users gained Crusher access in January 2022. Crusher training was provided the same month by AMD and HPE, while ORNL provided login instructions and a *Crusher Quick Start Guide.* Two Crusher hackathons followed in February 2022 for CAAR and ECP early science teams. Initial CAAR early science team results show Crusher speedups over Summit ranging from small gains to 15x. The Combinatorial Metrics (CoMet) application has successfully run on ~3.2K Frontier nodes. To date, the majority of ECP KPP-1 and KPP-2 applications are already running on Crusher. Once KPP goals are achieved, ECP teams will transition to early science allocations on Frontier. ECP is scheduled for full Frontier access in July 2022, and the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program in January 2023.

DISCUSSION

Cerf asked about coherent memory and the NIC design interface with the GPU. **Geist** explained coherent memory means the GPU can see and read the DDR memory directly from the CPUs or read the HBM memory on another GPU. There is a PCIe link from the GPU to the NIC.

Reed wondered about dependence on international supply chains. **Geist** explained ~ 80% of the parts were U.S.-based. Two European vendors were not required to abide by the DPAS rating but were willing to prioritize parts. Other suppliers were based in China and Canada. The supply chain was affected by COVID-19 and factory shutdowns.

Mesirov asked about CoMet. **Rachel McDowell** (ORNL, chat) said CoMet enables large scale Genome-Wide Epistasis Studies, co-evolution, pleiotropy, and other studies.

Berzins inquired about the proportion of applications bound by the data bus limit as opposed to making use of the multiple floating point operations per transfer word that are possible with the GPU architecture. Modern CPU architectures can face similar limitations. **Geist** explained; though the 3.2-TB/s transfer speed between the GPU and HBM memory is the fastest speed possible, this exchange still limits GPU speed.

Cerf wondered about machine allocation and usage to optimize cycle consumption. **Geist** relayed INCITE proposals for Summit are oversubscribed by a factor of three every year. All possible system hours are allocated, and system utilization is ~90%. INCITE proposals are screened to ensure a large machine is needed, and 40%-70% Summit jobs have used >25% of the machine. **Benjamin Brown** (DOE, chat) reinforced that ASCR requires HPC facilities to monitor and report on utilization. **Arthur** (chat) commented utilization has limited value as a driving metric. Social dynamics can also affect utilization.

Tiffany Trader (HPC Wire) asked about node usage, Frontier deployment and TDSs. **Geist** observed some researchers have used the two Graphics Compute Dies (GCDs) as a pair hooked together for a faster interconnect while others use the node like eight connected GPUs. Each usage presents advantages. The last Frontier cabinet was delivered in October, followed by cable wiring. Stabilization is ongoing. Crusher was turned over to ECP in November 2021 and will run until the end of ECP. The one-cabinet Borg TDS is available to system administrators for pre-software testing. **Helland** clarified that Frontier will be considered ready for operations after stabilization and acceptance testing.

Matsuoka drew attention to team maintenance of a code base that work across Summit and Aurora. **Berzins** (chat) replied Kokkos/ RAJA seem to be the main code base. Personally, Berzins found Kokkos/ CUDA was not immediately portable due to C++/ CUDA inconsistencies in loop code. This lack of portability may also apply to Kokkos/ SYCL or Kokkos/ HIP.

UPDATES ON THE SOFTWARE-STEWARDSHIP RFI, THE WORKSHOP ON REIMAGINING CODESIGN, AND THE ROUNDTABLE ON PARALLEL DISCRETE EVENT SIMULATION, Hal Finkel, Office of Science for Advanced Scientific Computing Research

The ASCR Software-Stewardship RFI was released in October 2021 with responses due in December 2021. The 37 responses yielded >360 pages of text from ECP groups, DOE national laboratories, non-profit organizations, the U.S. Research Software Engineer Association, and small and large businesses. Responses pertained to training on software development and use; workforce support; infrastructure for common development needs; curation and governance processes; maintenance of situational awareness; shared engineering resources; project support; application engagement and support; best practices training; community outreach and networking; and legal and administration support. Responses also proposed varying stewardship management and oversight structures as well as criteria to assess stewardship success. Future steps include finalizing the target scope of stewardship activities for FY23; defining the relationship between stewardship and synergistic activities in the Facilities, Research and Advanced Computing Technologies Divisions; pursuing the definition and release of FOAs in the targeted scope; and working with ECP, ASCR facilities, and others to enable a common understanding of how stakeholders will contribute to the overall process.

The ASCR Workshop on Reimagining Codesign (ReCode), held in March 2021, engaged 146 participants and 86 observers in addressing four codesign technology targets: 1) Leadership-scale systems and software; 2) Computing at the edge; 3) Accelerator modules and chiplets; and 4) Smart networking, memory, and storage. Resulting Priority Research Directions were: 1) Driving breakthrough computing capabilities with targeted heterogeneity and rapid design; 2) Software and applications that embrace radical architecture diversity; 3) Engineered security and integrity from transistors to applications; and 4) Design with data-rich processes.

The ASCR Roundtable on PDES, held in September 2020, was attended by 21 participants from eight national laboratories. Motivating use cases discussed applications related to transportation and mobility; the energy grid; internet and cybersecurity simulations; materials science using kinetic Monte Carlo processes; epidemiological planning, response, policy, and decision making; and hardware codesign and large-scale scientific infrastructure simulations. Research opportunities lie in advancements for core or inner technologies; usability and outer technologies; scientific enterprise and mission applications; and cross-cut technologies.

DISCUSSION

Pointing to opportunity costs, **Arthur** asked about funding to automate time-consuming aspects of software stewardship and incentives. **Finkel** said RFI responses highlighted challenges around incentives. There is a lot of interest in process automation; this is addressed by ASCR's X-stack: Programming Environments for Scientific Computing FOA.

Cerf stressed the importance of incentivizing software curation and stewardship. Does the ASCR supercomputing community use open source structures such as GitHub and employ automated debugging? Fuzzing is a common practice at Google. Do new computing facilities present a challenge for retrofitting HPC applications? **Finkel** explained the ECP has made a significant investment in tooling. Use of GitHub is considered best practice for teams working on projects that can be shared openly. There has also been work on a GitLab-based system for managing internal projects and their testing. Automated debugging and fuzzing are commonly raised issues. Some X-Stack projects are studying the application of fuzzing-based technologies to numerical and scientific software. Much of this work has been inspired by Google. Changes to underlying systems and software present challenges, and some issues are difficult to track down. Best user and system administrator practices include collecting system provenance information.

Sarkar drew attention to workforce gaps. The workforce is comparable to an inverted pyramid, with fewer individuals with assembly language skills at the base, software engineers following, and more individuals at the top level of domain science with Python and other language skills. There is a potential vision for no-code AI. Hiring practices should reconsider required qualifications for positions. **Finkel** remarked several RFI responses commented on requirements related to workflow software and Python packages. ASCR is trying to make HPC more accessible to individuals that are not C++ or Fortran experts.

Cerf (chat) asked about classified code security. **Hendrickson** (chat) indicated platforms that run classified applications are not on ESNet. **John Shalf** (LBNL, chat) added many security concerns discussed in ReCode are driven by more mundane issues than security classification. For example, participation in NIH collaborations requires handling of sensitive, not classified, data while enabling industry use of DOE systems involves handling export controlled data. Connecting experiments in real-time to supercomputers entails a patchwork of ad-hoc security systems. Current vendor hardware support for Trusted Execution Environments (TEEs) only work within a single node and do not offer an appropriate solution for HPC. There is an opportunity for codesign to enable the development of TEE's that are suitable for HPC and distributed capabilities needed by the open science community.

Robert Bartolo (Transformational Liaisons, chat) commented on European software stewardship activities.

UPDATE FROM WORKING GROUP ON COLLABORATION WITH NNSA AND NCI

Tony Hey, Working Group Chair

The ASCAC Subcommittee on the DOE-NCI collaboration reviewed the status of the Modeling Outcomes using Surveillance data and Scalable AI for Cancer (MOSSAIC) project on March 28, 2022. MOSSAIC is the third pilot study, now a project, stemming from the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program. MOSSAIC uses deep learning (DL), Natural Language Processing (NLP) approaches to extract information from electronic health records to accelerate reporting of cancer diagnoses, understand treatment efficacy, and improve outcomes while advancing exascale computational development. MOSSAIC is built on partnerships and community outreach engaging the NCI, DOE national Laboratories (ORNL and Los Alamos National Laboratory [LANL]), 10 cancer registries (nine Surveillance, Epidemiology, and End Results [SEER] Registries and one non-SEER Registry), industry and stakeholders, academia, and other federal agencies. MOSSAIC iterates cycles of R&D, rapid testing at scale, and API release to partners.

In total, the ASCR Leadership Computing Challenge (ALCC) and ECP, respectively, have contributed 270K and 300K Summit node hours over MOSSAIC's lifetime. The OLCF has allocated 130K Summit node hours for 2021-2022 and ongoing efforts are using CITADEL to train MOSSAIC models with protected health information data. Models have been tested on three Frontier TDSs (Tulip, Spock and Crusher), and day-one model readiness is expected for Frontier. MOSSAIC is successfully advancing DOE's mission in supercomputing and AI through algorithms, privacy-preserving federated learning, and distributed AI and prototyping Integrated Research Infrastructures.

MOSSAIC's 2022 priorities comprise: 1) Scalable transformer language models for clinical information extraction; 2) Data collection and development of new recurrence and biomarker APIs; 3) Uncertainty quantification (UQ) and interpretability; 4) Clinical integration and translation; and 5) Enabling transformer training on Leadership Computing Facilities for CANcer Distributed Learning Environment (CANDLE).

The Subcommittee was impressed by MOSSAIC's achievements, including deployment of a production NLP system at several SEER centers, and interested in other application areas. Concerns were raised about sustainability. Future Subcommittee meetings will address the additional JDACS4C projects: AI-Driven Multiscale Investigation of Ras-Raf Activation Life cycle (ADMIRRAL); Innovative Methodologies and New Data for Predictive Oncology Model Evaluation (IMPROVE); and CANDLE.

DISCUSSION

Vint asked about transformers. **Hey** said transformers are widely used in NLP.

Mesirov posed questions about project goals, input data, performance, and metrics of success. **Gina Tourassi** (ORNL) explained MOSSAIC's overarching goal is the development and deployment of AI tools to enable cancer registries across the US to achieve near real-time collection of information for the national cancer surveillance system. Currently, U.S. cancer registries employ a manual process that takes ~28 months from diagnosis until surveillance database incorporation. MOSSAIC AI tools extract information about five to seven cancer attributes, biomarkers, and metastasis and recurrence from unstructured pathology and radiology reports. The program's measure of success is deployment across registries. Accuracy is measured through API comparison to retrospective data. Manual annotation takes 55 seconds per report versus AI's five to seven milliseconds, reducing the data entry cycle to 30 days. Guided by conversations with registries, the current approach leverages UQ and autocodes reports for which the system has demonstrated >97% accuracy across all data elements. System implementation results in a 20%-30% workload reduction, depending if reports are analyzed individually or in the context of longitudinal AI tools.

QUANTUM REQUEST FOR INFORMATION, Ceren Susut, Office of Science for Advanced Scientific Computing Research

The SC's RFI on Access to Quantum System sought responses to 12 questions. Twenty-three distinct comments from national laboratories, universities, companies and other stakeholders were submitted during the comment period from August to September 2021. Responses fell into six broad categories: 1) types of quantum systems; 2) the role of federal agencies; 3) serving a broad community of users; 4) access criteria and success metrics; 5) intellectual property; and 6) other factors, issues, and opportunities. Overall, a mix of perspectives were provided on the need for access to both commercial systems and earlier-stage systems not yet ready for commercialization. Respondents advocated for a broad range of technologies and platforms and adoption of useful practices and lessons learned from DOE user facilities. In addition to providing access, DOE's role was envisioned as providing training, internships, and support for users. A draft road map will be developed in Spring 2022 and feedback from the National Quantum Coordination Office (NQCO) and interagency partners will be gathered in Summer 2022. A final roadmap will be presented to ASCAC in September 2022.

DISCUSSION

Cerf asked about industry responses, DOE investments in developing algorithms, and if conventional machines can emulate quantum machines to produce useful computations. **Susut** commented there are many industry-DOE collaborations in the QIS space. Industry is interested in both QIS access and addressing basic research challenges. DOE is working on near- and long-term use cases, as illustrated by the NQISCs. The literature contains examples of supercomputers emulating quantum machines. Classical computers have also been used to simulate small quantum computers.

Svore commented on challenges surrounding dual-access technology related to physical machine and solutions access. There is potential to automate access through detection mechanisms built into machines, but doing so will require focused effort, especially as machines are developed and scaled. **Susut** remarked that respondents emphasized providing access to quantum systems as well as classical simulators and supercomputers in parallel to quantum systems. **Helland** reinforced that anyone given user facility access must provide a proposal, but back-end tracking to ensure users do what they have proposed to do could be improved.

PUBLIC COMMENT

John Negele (Massachusetts Institute of Technology, Emeritus) asked if ASCR has considered developing algorithms or technology to reduce energy consumption related to Bitcoin mining. **Reed** commented energy efficiency has been a focus of ECP. **Helland** said no specific projects are looking at Bitcoin at this time. Projections suggest AI will also consume a lot of future energy; the FY23 budget will begin addressing energy-efficient algorithms. Dongarra has led work in mixed precision efforts. **Dongarra** agreed mixed precision will be an important future tool. **Thomas Wong** (ASCR, chat) stated an executive order released earlier this month charged the Office of Science and Technology Policy, in consultation with DOE and other agencies, with producing a report including material addressing cryptocurrency energy consumption.

Reed and Chalk remarked the next meeting will be held in July via a hybrid format.

Reed adjourned the meeting at 3:06 p.m.

Respectfully submitted on April 19, 2022,
Holly Holt, PhD and Sonia Isotov, MLIS,
Science Writers, Oak Ridge Institute for Science and Education (ORISE)