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

PUBLIC MEETING MINUTES

September 26-27, 2024

HYBRID MEETING

ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE

The U.S. Department of Energy (DOE) Advanced Scientific Computing Advisory Committee (ASCAC) convened a hybrid meeting on Thursday, September 26th, and Friday, September 27th, 2024 at Hilton Arlington National Landing, 2399 Richmond Hwy, Arlington, VA 22201 and 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.

Designated Federal Officer (DFO) for ASCAC
Ceren Susut, Associate Director, Advanced Scientific Computing Research (ASCR)

ASCAC members present in-person
Martin Berzins (Chair), University of Utah
Roscoe Giles (Vice Chair), Boston University
Richard Arthur, General Electric (GE)
Tina Brower-Thomas, Howard University
Timothy Germann, Los Alamos National Laboratory (LANL)
Mary Ann Leung, Sustainable Horizons Institute
Vanessa Lopez-Marrero, Brookhaven National Laboratory (BNL)
Irene Qualters, LANL
Sameer Shende, ParaTools Inc.
Valerie Taylor, Argonne National Laboratory (ANL)
Cristina Thomas, 3M Company (retired)

ASCAC members present virtually
Keren Bergman, Columbia University
Jacqueline (Jackie) Chen, Sandia National Laboratories (SNL)
Gilbert Herrera, Department of Defense (DOD)
Anthony Hey, University of Washington
Alice Koniges, University of Hawaii
Jill Mesirov, University of California, San Diego
John Negele, Massachusetts Institute of Technology (MIT)
Krysta Svore, Microsoft
David Torres, Northern New Mexico College
Juan Torres, National Renewable Energy Laboratory (NREL)
James Whitfield, Dartmouth College
Theresa Windus, Iowa State University

ASCAC members absent
Vinton G. Cerf, Google Inc.
Sunita Chandrasekaran, University of Delaware
Mark Dean, University of Tennessee
Susan Gregurick, National Institutes of Health (NIH)
Alexandra (Sandy) Landsberg, Office of Naval Research
Satoshi Matsuoka, RIKEN Center for Computational Science
Vivek Sarkar, Georgia Institute of Technology
Edward Seidel, University of Wyoming

Office of Science personnel present in person

Ben Brown, ASCR
Christine Chalk, ASCR
Tanner Crowder, ASCR
Hal Finkel, ASCR
Jeremy Crampton, ASCR
Harriet Kung, Acting Director, DOE Office of Science (SC)

Presenters

Katie Antypas, National Science Foundation (NSF)
Ben Brown, ASCR
Hal Finkle, ASCR
Helena Fu, Director, Office of Critical and Emerging Technologies
Si Hammond, National Nuclear Security Administration (NNSA)
Kibaek Kim, ANL
Bronson Messer, Oak Ridge National Laboratory (ORNL)
Ojas Parekh, SNL

Attending in-person

Richard Carlson
Leland Cogliani
Ben Kallen
Michael Parks
Robert Rallo
Griffin Reinecke
Arjun Shankar
Gina Tourassi
Sterling Waugh
Andrew Wiedlea

Attending virtually

Lali Chatterjee
Marco Fornari
Carol Hawk
Saswata Hier-Majumder
Thuc Hoang
Dorothy Koch
Kalyan Perumalla
Robinson Pino
Ashley Predith
David Rabson
Bill Spotz
Julie Stambaugh
Jordan Thomas

There were approximately 260 individuals present for all or part of the meeting.

Thursday, September 26, 2024

OPENING REMARKS, Martin Berzins, University of Utah

Berzins, ASCAC Chair, convened the meeting at 10:06 a.m. Eastern Time (ET), welcomed the in-person and virtual attendees, and noted that upcoming ASCAC meetings will increasingly consider goals for the future of ASCR, SC, and DOE.

REMARKS FROM THE UNDER SECRETARY FOR SCIENCE AND INNOVATION, Geraldine Richmond, Under Secretary for Science and Innovation at the Department of Energy (prerecorded)

Richmond expressed gratitude to ASCAC for its critical role in keeping DOE and ASCR at the forefront of research and technology.

DOE and ASCR are continuing to reap the benefits of the Exascale Computing Project (ECP), with several Gordon Bell Prize finalists having leveraged the exascale capabilities of Frontier at ORNL. Frontier has enabled scientific advances across a broad range of applications, including the recent breakthrough achieving double precision exaflop performance for molecular dynamics simulations for highly precise predictions of electron behavior and interaction. The ability to simulate atomic interactions with high levels of speed and accuracy is transformative for the field of computational chemistry. Industry continues to leverage exascale computing to drive energy innovation in pursuit of net-zero goals, with companies developing more efficient technologies in areas such as aviation. Additionally, digital twins for health applications are crucial to improving patient outcomes, advancing individualized care, and contributing to the Cancer Moonshot initiative.

In the realm of artificial intelligence (AI), ASCR continues to advance research on AI for science to meet the nation's future needs. ASCR has recently announced new awards focusing on foundational models for computational models, automated workflows and laboratories, scientific programming and knowledge-management systems, federated and privacy-preserving technologies, and energy-efficient AI algorithms for hardware for science. Because of the importance of developing AI technologies sustainably and responsibly, ASCR plays a leadership role in driving innovation for energy-efficient advanced computing. On September 12, 2024, DOE issued a request for information (RFI) focused on the Frontiers in AI for Science, Security, and Technology (FASST) Initiative, seeking public input to inform how DOE and the national laboratories can leverage existing assets to provide a national AI capability.

ASCR is pushing the vanguard of quantum information science (QIS), and over the past year has released funding opportunity announcements (FOAs) related to quantum hardware emulation and accelerated research in quantum computing, which advance efforts towards creating modular software stacks and utility demonstrations. The September 17-19, 2024 DOE National QIS Research Centers (NQISRC) Principal Investigator meeting demonstrated the great work being done at the Centers, including advancements in quantum error correction. Additionally, the SC and Defense Advanced Research Projects Agency (DARPA) announced a signed memorandum of understanding (MOU) to coordinate activities related to quantum computing.

In conclusion, **Richmond** paid tribute to two recently passed DOE luminaries, Ed Temple and Charles McMillan, who made significant contributions throughout their careers to SC and LANL, respectively.

VIEW FROM WASHINGTON, Harriet Kung, Acting Director of the Office of Science

The FY25 SC Budget Request totals ~$8.583B, a 4.2% increase over the FY24 appropriation. Increased investments in Administration priorities include: AI ($259M; +$93M from FY 24); Fusion Innovation Research Engine (FIRE) Collaboratives ($60M; +$15M from FY24); Reaching a New Energy Sciences Workforce (RENEW) ($120M; +$69M from FY24); Funding for Accelerated, Inclusive Research (FAIR) ($64M; +$32M from FY24); Climate Initiative ($20M); Microelectronics ($59M; +$22M from FY24), and SC Energy Earthshots ($115M; +$95M from FY24). SC continues to fund facility and laboratory operations and infrastructure projects, with investments including: supporting user facility operations at ~88.3% of the re-baselined optimal funding levels; $50M for upgrading core laboratory infrastructure (+$32M from FY24); the Laboratory Operations Apprentice Program ($5M; +$2M from FY24); and continued support for ongoing infrastructure and scientific user facility upgrade projects.

The House and Senate marks both support SC's FY25 budget request for increased funding over FY24 levels, which is critical to ensuring ASCR's continued leadership. Facility operations and construction projects continue to receive bipartisan support from the House and Senate. SC is waiting for the final conference between the House and Senate to resolve discrepancies.

The FY25 House mark funds SC at $8.39B, which is $150M over the FY24 Enacted Budget but $193M below the FY25 Request. Challenges with the House mark includes no funding for RENEW or FAIR initiatives, Energy Earthshots being funded at $20M ($95M below the FY25 Request), FIRE Collaboratives being funded at $40M ($20M below the FY25 Request), and that the language regarding specified funding levels for Quantum User Expansion for Science and Technology (QUEST) and high performance-computing (HPC)-quantum integration testbeds make it difficult for ASCR to execute. Additionally, the House was supportive of microelectronics research but cautioned that tight budgets may not support the assumed budget over the next three years. The House mark also directed the establishment of a Carbon Sequestration Research and Geological Computational Science Initiative.

The FY25 Senate mark funds SC at $8.6B, which is $360M over the FY24 Enacted Budget and $17M above the FY25 Request. The Senate demonstrated strong support for AI ($160M) and FASST ($100M), and designated FAAST as a separate congressional control line. QIS research and the five National QIS Research Centers were supported at $265M ($15M below the FY25 Request), Energy Earthshots was supported at $60M ($55M below the FY25 Request), microelectronics was supported at $110M ($15M over the FY25 Request), and the FIRE Collaboratives were supported at not less than $45M ($15M below the FY25 Request). Unlike the House, the Senate supports the RENEW and FAIR initiatives to broaden participation. Additionally, the Senate mark designated $25M to establish a Carbon Sequestration Research and Geologic Computational Science Initiative and $10M for atmospheric methane removal research.

SC leadership updates include Chris Landers assuming the role of Director of the Office of Isotope Research and Development (R&D) Production and Sarah Stanton assuming the role of the Director of the Office of International Activities, Research Security, and Interagency Coordination. Additionally, SC has renamed the Scientific Workforce Diversity, Equity, and Inclusion Office to the Office of Scientific Workforce Integrity (SWI) and the Office of Equity and Workforce Development to the Office of Energy Sciences Workforce Stewardship (ESWS). Although the missions of these offices have not changed, the names have been adjusted to better reflect the scope and stewardship roles of the offices.

The FASST program represents an important opportunity for ASCR to lead on AI within SC and the wider community. Additional roundtables are being organized to build on the foundation ASCR has established.

On the topic of research security, DOE has been developing research, technology, and economic security (RTES) policies for nearly a decade following concerns of foreign influence on U.S. funded research and technologies. RTES policies must address a broad range of risks while also balancing U.S. competitiveness by ensuring the DOE and national laboratories can continue to attract and retain the best and the brightest as well as promote principled international collaborations. In 2019, DOE formalized RTES efforts by prohibiting federal employees and contractors at national laboratories from participating in foreign talent programs sponsored by countries of concern, such as China, Russia, Iran, and North Korea. In addition to on-site research security, the DOE and national laboratories developed the Science and Technology (S&T) Risk Matrix to categorize research topics into green, yellow, and red zones based on the level of risk and required vetting and oversight for collaboration. The green zone means collaboration could benefit the U.S., the yellow zone requires some further review before proceeding with collaboration, and the red zone requires review and approval at the undersecretary level. Six critical sensitive emerging technology areas were identified: AI, QIS, HPC, batteries and energy storage, battery technology, and accelerator R&D. This goal of risk-based approach is to protect these critical technologies while allowing for beneficial international collaboration where appropriate.

Early RTES efforts focused on the national laboratories and held off on financial assistance while the interagency process was launching. With the release of the National Security Presidential Memorandum 33 (NSPM-33) in 2021 and the NSPM-33 Implementation Guidance in 2022, the requirements and expectations of what to protect from countries of concern was further codified. Concurrently, the Office of Science and Technology Policy (OSTP) led an interagency effort to establish and harmonize guidelines and policies between agencies. As a result, DOE efforts and policies governing national laboratories research security have been strengthened. There is also a heightened emphasis on transparency, especially foreign nexus, affiliation, and sources of support. SC must demonstrate there are guardrails protecting the intellectual property generated by taxpayer funded research while not compromising the integrity and openness of the U.S. innovation ecosystem.

In 2023, the DOE RTES Policy Working Group (PWG) was established to address RTES policy development and consistency with interagency processes. In 2023, the RTES Office was established to perform due diligence reviews and risk mitigation at three phases: prior to FOA release, after peer review but before final selection of awards, and during the grant lifecycle

when any key personnel changes occur. SC has already begun interfacing with the RTES Office on a few FOAs, and the process adds time and complexity to the grant award process but ensures thorough vetting for foreign nexus risks. There is a process of to provide input to the DOE RTES office on risks levels specific to the each of the critical technology areas. In March of 2023, DOE updated the S&T Risk Matrix to reflect consistency with major scientific and technological developments.

SC financial assistance has not changed significantly. SC recommends the use of universal disclosures, the use of Science Experts Network Curriculum Vitae (SciENcv) to reduce administrative burden and has announced the acceptance of interagency common formats for current pending support and bio-sketches. Additionally, SC strongly supports recent efforts from interagency partners, including DoD and NIH decision matrices, the NSF Trusted Research Using Safeguards and Transparency (TRUST) Framework, and the continued development of the NSF Safeguarding the Entire Community of the US Research Ecosystem (SECURE) Center. DOE and SC are attempting to build an understanding and awareness of risks into the culture of the community and research enterprise. SC invites feedback from the community and interagency partners regarding RTES policies.

DISCUSSION

Giles asked if current feedback mechanisms were effective at providing a view of ground-level research activities, policies, and procedures. **Kung** noted one concern is the lengthening of time to award due to the administrative burdens associated with the addition of two more stages to the review and due-diligence process. The RTES Office will need to rely on the technical expertise of SC's program managers, and as the review process gets underway there will be increased understanding of the appropriate risk posture needed for certain activities. For instance, open literature research activities will hopefully have a speedy review and approval process. In the long-term, the interagency working group is considering how to measure the effectiveness of the security guardrails. One concern is that the increased security processes may have a chilling effect. The challenge lies in balancing the need for security, the maintenance of international research collaborations, and the health of the innovation ecosystem.

Thomas raised concerns about a potential chilling effect where researchers might abandon collaborations due to the perceived complexity of security measures. **Kung** acknowledged this risk and stressed the importance of clear, transparent policies that make it easy to comply while protecting sensitive data. Researchers will be able to have a conversation with RTES and the sponsoring research office to discuss potential concerns. The goal is to maintain a robust international talent pipeline without stifling innovation.

Qualters asked if RTES will engage in related topics, such as International Traffic in Arms Regulations (ITAR). **Kung** clarified that different departments handle ITAR, as well as Freedom of Information Act (FOIA) requests. However, there is value in coordinating some of these related topics. **Arthurs** asked if there was consistency in reviewing FOIA requests. **Kung** added it would be up to the program office, such as ASCR, to ensure a consistent approach to FOIA requests. So far, there have not been many FOIA requests which have intersected with RTES.

Berzins questioned how to balance attracting international talent, particularly from countries of concern, while sufficiently safeguarding research security. **Kung** emphasized the need to maintain a clear, transparent process that allows for collaboration in low-risk areas while ensuring approvals for higher-risk engagements. This balance is critical to preserving the U.S. innovation ecosystem and talent pipeline, which relies on contributions from international researchers. There is no prohibition signaling out a nationality, and RTES has given approvals for people from countries of concern.

VIEW FROM GERMANTOWN, Ceren Susut, Associate Director of the Office of Science for Advanced Scientific Computing Research

Susut reviewed organizational and personnel changes within ASCR. There are two new hires: Dr. Tanner Crowder, Senior Technical Advisor, and Dr. Hal Finkel, Director of the Computational Science Research & Partnerships Division. Additionally, the Facilities Division welcomed two new members: Mahfuzun Nabi, Pathways Intern, and Dr. Jeremy Crampton, American Association for the Advancement of Science (AAAS) Fellow.

The FY25 ASCR Budget Request of $1.152B has received House and Senate marks. In sum, the House and Senate marks designate ASCR receive not less than $1.105B and $1.152B, respectively. Additionally, the House and Senate marks include differing language and allocations regarding non-ASCR programs which relate to ASCR areas of interest.

The House mark provides: not less than $219M for the Argonne Leadership Computing Facility (ALCF); $260M for the Oak Ridge Leadership Computing Facility (OLCF); $146.5M for the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory (LBNL); $93.5M for infrastructure upgrades and operations for the Energy Sciences Network (ESnet); $16M in other project costs for the High Performance Data Facility (HPDF); $330M for Mathematical, Computational, and Computer Sciences research; and up to $35M to support research to develop a new path to energy efficient computing with large, shared memory pools. DOE is directed to provide to the House Committee a report which analyzes the ongoing efforts to acquire high performance and quantum computing systems, advance research in quantum error correction, and develop a strategy for expanding and integrating quantum error correction research activities within the ASCR program. As it relates to ASCR areas of interest, the broader House markup includes language for: not less than $245M for QIS, including not less than $120M for research and $125M for the five National QIS Research Centers (Quantum Centers); establishment of a roadmap integrating the scientific goals of each of the Quantum Centers; support for expanded quantum internet, networking, and communications testbeds; up to $15M to conduct research in support of the QUEST; $20M to strengthen testbeds on HPC facilities to study integration of quantum processes with traditional HPC; and support for expanded relationships with the NIH and the National Institute of Mental Health (NIMH). Additionally, other guidance in the House markup includes: no funding for RENEW or FAIR; $20M for the Energy Earthshots program including $5M from ASCR; concerns about the sustainability of the Microelectronics Centers and the recommendation for SC to ensure flexible funding; not less than $35M for the Established Program to Stimulate Competitive Research (EPSCoR); and commendations for the DOE's strategy to adopt and implement AI in a scalable, secure, and interoperable manner.

The Senate mark indicates strong support for ASCR's leadership in AI, QIS, HPC, and the ECP. The Senate mark allocates: $225M for the ALCF; $260M for the OLCF; $146M for NERSC; $93M for ESnet; $7M in other project costs for the HPDF; and a directive to provide the Committee a briefing on a coordinated DOE plan for what leadership computing facilities should accomplish post-exascale. On the research side, the Senate mark provides: not less than $300M for Mathematical, Computational, and Computer Sciences Research; not less than $20M for the Computer Science Graduate Fellowship (CSGF); up to $35M to support research to develop a new path to energy efficient computing with large, shared memory pools; a recommendation for the development of advanced memory technologies to advance AI and analytics for science applications of very large-scale memory systems and memory semantic storage; and a directive to implement a hybrid HPC / Quantum Computing Pathfinder Program at a Leadership Computing Facility (LCF), with a recommendation of up to $15M to acquire an on-premise quantum computer and up to $10M for a parallel R&D program to address basic research challenges in algorithms and software stack. Further language in the Senate markup related to ASCR includes: not less than $265M for QIS, including not less than $120M for research and $125M for the five Quantum Centers; $100M to implement the FASST initiative in coordination with the Critical Emerging Technologies Office and the NNSA; not less than $160M for AI and machine learning (ML) across SC programs; the expectation that ASCR will take a leading role in DOE's AI / ML activities; and a recommendation of $5M for the Critical and Emerging Technologies Office. Additionally, other guidance in the Senate markup includes: support for the RENEW and FAIR initiatives; not less than $110M for microelectronics and semiconductor manufacturing innovation; encouragement for SC to continue supporting the Accelerate Innovations in Emerging Technologies program; $35M for EPSCoR; and $60 for the Energy Earthshots program, including up to $15M from ASCR.

On July 26, 2024, SC charged ASCAC to evaluate the effectiveness of the CSGF program along eight charge question areas. Irene Qualters is serving as the subcommittee chair addressing the CSGF charge.

In recent ASCR community news, DOE and DARPA are collaborating in quantum benchmarking and have established a MOU to coordinate future research, development, engineering, and evaluation activities related to quantum computing. In addition, ASCR hosted three Workshops in Energy-Efficient, Analog, and Neuromorphic Computing. Recent PI meetings included the Scientific Discovery through Advanced Computing (SciDAC) PI meeting, the first DOE Quantum Centers PI meeting, and a Science Summit for Energy Earthshot Innovation.

The ECP is officially completed and represents a great ASCR and DOE success story. The ECP was supported by the research community, industry users, and interagency partners. Following the completion of the ECP, ASCR will continue to push the frontiers in scientific breakthroughs and push innovations in HPC. ASCR thrives on daring and ambitious ideas, such as the ECP, and relies upon a strong research foundation, facilities ecosystem, partnerships, and the computing community.

Lastly, ASCR holds virtual office hours on the second Tuesday of each month at 2 pm ET to broaden awareness of programs. The ASCR website has more information, including slides and recordings of past events.

DISCUSSION

Giles inquired about the level of funding for the facilities, in particular the data facility. **Susut** said the House mark was consistent with the FY25 request, however the ASCR leadership is exploring the discrepancy in the Senate mark. Regardless, ASCR will support building the team necessary for projects to be successful within the available funds. Additionally, **Giles** expressed disappointment with the House mark for RENEW and FAIR.

Thomas asked about general areas of excitement or concern regarding the budget. **Susut** responded that ASCR has the opportunity to shape the future of computing and is well-positioned with a one-of-a-kind ecosystem created from the ECP. ASCR has the ability to combine the power of high-fidelity simulations and AI to advance science. Additionally, ASCR has the opportunity to attempt merging different methods of computing. One slight concern is the challenge of communicating the value of how sustained investments in basic research lay the groundwork for future advancements.

Svore asked about the objectives of the DARPA and DOE MOU. **Susut** replied that DOE is currently working with DARPA on the details for implementation for the different collaboration opportunities. One area of interest is collaborating in investigating the utility of quantum computing, a topic which DARPA is addressing through their recently announced Quantum Benchmarking Initiative. There are many other synergies between DOE and DARPA. In regard to the DARPA MOU, **Herrera** encouraged ASCR and the DOE labs to offer more than just verification and validation. **Susut** agreed that ASCR brings many unique strengths to the relationship, and the collaboration strategy focuses on leveraging these strengths (including testbeds and research investments in algorithms) and complementing DARPA's efforts.

Svore raised the question of what is needed to enable the next version of an ECP-scale initiative aimed at integrating across different disruptive technologies. **Susut** said a post-ECP initiative would build on community input and engagement, such as feedback gathered through recent ASCR workshops. As a federal agency, ASCR can lead the community in setting ambitious goals, for example through energy efficiency goals. The challenge moving forward is in integrating different technologies and advancing HPC as a whole.

Torres inquired how ASCR's investments compare to China's investments. **Susut** answered that although this information is monitored closely, direct comparisons are difficult because there is a lack of precise information about what China is doing specifically. Generally, China has very ambitious goals in advancing quantum computing, HPC, AI, and post-exascale computing. China is not specifically focusing on one technology but is working towards advancing computing as a whole. Attracting more students and foreign researchers to build the research community will help ASCR reach the same goals.

Berzins asked about the reasoning behind the markups regarding large, shared memory pools. **Susut** said both committees had a large, shared interest in advancing large, shared memory pools, which is a capability relevant to many different technologies.

UPDATE FROM ASCR RESEARCH DIVISION, Hal Finkel, ASCR

Finkel presented FY24 ASCR Research Division updates. ASCR set ambitious goals for FY24, including revitalizing future planning, making substantial investments in AI and quantum

technologies, and establishing partnerships to maximize the capabilities of exascale computing. All goals were successfully met, which is a testament to the hard work of the ASCR team and research community.

Key thematic areas of ASCR research included: advanced modeling, simulation, and visualization; frontier AI and data for science; heterogeneous, distributed, co-designed, energy-efficient computing and algorithms; software complexity for increased versatility; and HPC and networking across experiments, exascale, and the edge.

Regarding quantum computing, ASCR had a major solicitation in FY24 for Accelerated Research in Quantum Computing. This represents a significant investment of $65M for the portfolio of 5-year projects. The solicitation's two topic areas are the creation of modular software stacks for quantum computing as well as the investigation of quantum utility. Many of the portfolio's awardees are operating as collaborative awards between multiple universities, national laboratories, agency, and industry partners.

ASCR competed a solicitation in FY24 for Data Rection for Science, which sought projects related to efficient and accurate compression, streaming, storage, and usage of data. This solicitation represents $5.5M for a portfolio of 3-year projects.

The October 2023 AI executive order directed activities related to ASCR's core mission, including supporting the development of safe, secure, and trustworthy AI technologies, supporting the development of Privacy Enhancing Technologies (PETs), and developing tools that facilitate foundation models useful for basic and applied science. Investing in AI for science is necessary to produce scientifically relevant and scientifically valid results, as traditional AI models are not always applicable to scientific data analysis. The Advancements in Artificial Intelligence in Science FY24 solicitation received over 500 pre-proposals, leading to 11 selected awards representing $67M in planned funding for the portfolio of 3-year projects. These project cover areas such as foundation models, privacy-preserving and federated training, automated workflows, and energy-efficient AI algorithms and hardware.

The FY24 Exploratory Research for Extreme-Scale Science (EXPRESS) solicitation represents $10.8M in planned funding for the portfolio. New areas explored this year included enhanced visualization, discrete event simulation and agent-based modeling, neuromorphic computing, advanced wireless, and quantum hardware emulation. This program is important in growing through exploratory research to foster long-term success in the broader portfolio.

In the era of exascale computing, it is important to engage in strategic partnerships and investments to ensure the potential of exascale computing is fully realized. To this point, the FY24 SciDAC: Partnership in Electricity solicitation, operated in partnership with the DOE Office of Electricity, represents $6M in planned funding for the portfolio of 3-year projects which focus on the electric grid. Additionally, the FY24 Codes for High Performance Computing for Manufacturing solicitation, operated in partnership with the DOE Advanced Materials & Manufacturing Technologies Office, represents $3M in investment to improve national laboratory-developed software with the goal of increasing the ability to leverage HPC resources to help U.S. manufacturers solve problems through advanced modeling and simulation.

ASCR's core research investment portfolio was reinvigorated through the Competitive Portfolios for Advanced Scientific Computing Research solicitation representing $87M in planned funding for the portfolio of 4-year projects. The program's goals include supporting

long-term high-impact research, responding to emerging science and technology trends, and collaborating with a diverse community of academic and industry partners. This full re-competition of ASCR's "base program" expanded the program to include applied mathematics, computer science, and advanced computing technologies and testbeds. Fifteen of the national laboratories proposed over sixty distinct research thrusts, many of which build on the success of ECP investments.

There was a call this year for Microelectronics Science Research Centers, which builds upon the success of the Abisko Microelectronics Codesign portfolio and will contribute to addressing foundational challenges in the design, development, characterization, prototyping, demonstration, and fabrication of microelectronics. This call represents $60M of funding per year to establish the Microelectronics Science Research Centers; the two focus areas for FY24 were energy efficiency and extreme environments.

ASCR has continued to invest in growing and diversifying the research community, such as through continued investment in the Early Career Research Program which bring more early career scientists into the ASCR community. Two awards were made through the EPSCoR program focused on quantum networking and neuromorphic hardware. For FY24, ACSR has planned investments for RENEW and FAIR, with selection announcements forthcoming. Additionally, the Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs continue to invest in the business community. For FY25, the two SBIR/STTR Phase I topics are accelerating the deployment of advanced software technologies and HPC cybersecurity. ASCR also continues to make investments in the next-generation software stack through the Consortium for the Advancement of Scientific Software (CASS) program.

In FY24, BRN workshop reports were published for Quantum Computing and Networking and for Visualization for Scientific Discovery, Decision-Making, & Communication. Three workshops were held on the Future of Computing for Science, covering the key topic areas of energy-efficient computing, analog computing, and neuromorphic computing. These workshops were widely attended, and workshop slides and plenary sessions are published online.

DISCUSSION

Taylor asked about the goals for FY25. **Finkel** noted that while FY25 budget and planning process is currently underway, the ASCR Research Division plans to continue core portfolio investments in SciDAC and in computational partnerships. Additionally, it is expected there will be major investments in quantum technologies, and new research areas will likely be explored through programs like EXPRESS. **Finkel** encouraged the research community to look at the forthcoming open call which will include language describing the overall research portfolio areas and scope of investment.

Taylor inquired if there were any plans to address the significant demand in AI for science in the upcoming year. **Finkel** acknowledged the uncertainty around the budget and external political factors but emphasized that AI remains a critical and exciting focus area for ASCR. For example, the recent workshops on energy-efficient computing, neuromorphic computing, and analog computing highlight research areas which will contribute to enabling the next-generation of AI technologies, hardware, and algorithms. The goal is to have a full pipeline

for AI research and development, where basic research investments can mature through large-scale testbeds and be picked up by facilities and partners. Although the competitive portfolio selections represent a range of topic areas, the ASCR Research Division is thinking about how to best address any gaps in the current portfolio in the future.

Arthur suggested that there is an opportunity to use AI trained on technical libraries or frameworks to provide cognitive assistance and productivity support. **Finkel** responded positively, pointing to two specific projects within the AI portfolio that focus on advanced programming systems for computational science as well as the development of AI agents for automated laboratory and scientific workflows. These projects, as well as concurrent advancements, aim to address how the latest AI technologies can be paired with the scientific tools developed for exascale to further scientific research.

Berzins raised the idea that software shelf-life may be increasingly limited due to the rapidly evolving nature of AI, therefore it is recommended to keep an eye on future AI developments. **Finkel** agreed and stressed the importance of ensuring that the next-generation software stack is not static but capable of evolving alongside hardware and computation innovations and science needs.

CRITICAL AND EMERGING TECHNOLOGIES, Helena Fu, Director, Office of Critical and Emerging Technologies

The Office of Critical and Emerging Technologies was established in December 2023, in response to the White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI. The office focuses on coordinating efforts within the DOE and with external agencies and stakeholders in four critical and emerging technology areas: biotechnology, QIS, AI, and microelectronics. These areas are considered foundational, dual-use, and have been historically supported by DOE.

The office's primary role is to ensure that DOE capabilities and work are effectively leveraged by interagency and external partners. A major focus has been the implementation of the AI Executive Order, which included numerous directives and deadlines for federal agencies. DOE has played a significant role in promoting AI innovation while managing its associated risks, and SC has also been very engaged in these efforts.

Highlighted initiatives included a pilot program to consolidate resources and share training opportunities across DOE and national laboratories, as well as a website to centralize AI tools, models, and partnerships redated to foundational models for science, energy, and national security. The office also supported the development of a global AI research agenda, launched during the United Nations (UN) General Assembly in September 2024.

Regarding AI risk management, the office has been involved in making testbeds available, working other agencies to evaluate safety of AI models, read teaming AI models for radiological and nuclear risks, and working with the U.S. AI Safety Institute at the National Institute of Standards and Technology (NIST).

The office has been working across DOE and the national laboratories to assist with scenario development and planning related to the FASST Initiative. Additionally, a recent RFI was issued to gather input from the community to inform planning around the FASST Initiative. Beyond AI, the office is similarly focused on ensuring DOE's contributions in biotechnology,

quantum, and microelectronics are leveraged in broader policy discussions, particularly in areas where DOE can make the most impact.

DISCUSSION

Giles asked how the Office of Critical and Emerging Technologies thinks about its budget and allocations. **Fu** explained that the office currently has no budget, although a FY25 budget request has been submitted. The office is positioned within DOE as a small leverage-focused entity with the goal of elevating DOE's crosscutting work in the four critical technology areas.

Herrera inquired about the office's work in microelectronics policy. **Fu** explained that there is a significant amount of interagency work happening in microelectronics, particularly including the National Science and Technology Council (NSTC), SC, and the National Security Agency (NSA). The office will assist where needed to bring appropriate perspectives together and coordinate DOE's efforts and expertise. The White House is involved in guiding policy regarding AI diffusion, export controls, and how AI is being powered.

Berzins asked how the office plans to manage relationships with hyperscalers and address the global landscape of innovation. **Fu** replied that the office has been engaging with hyperscalers and utilities, especially on energy issues related to powering AI. The DOE has mobilized resources to address the rising AI and datacenter driven energy demands through technical assistance, grants, and support through Loan Programs Office. The office is also involved in discussions around AI safety and security, with the AI Safety Institute leading engagement with companies. While industry-driven AI models are advancing their scientific reasoning capabilities, there will always be a need for new types of scientific data that leverages DOE's scientific infrastructure and workforce expertise. DOE is uniquely positioned to provide scientific data, infrastructure, and expertise to address solving real-world problems.

EXASCALE SCIENCE HIGHLIGHTS FROM FRONTIER, Bronson Messer, Oak Ridge National Laboratory

Frontier is a unique and groundbreaking scientific instrument which has two exaflops of peak power distributed across 74 compute racks. Frontier's large amount of memory has already enabled significant scientific advances. Additionally, the Crusher test and development system has been subsumed into Frontier.

With this computing power, memory, and storage, Frontier is able to address a variety of scientific and technical problems. In particular, Frontier excels in addressing problems that require resolution of a vast range of scales, problems where a lot of physics is going on at every spatial point (combustion chemistry, radiation transport, etc.), and problems which require many high-fidelity simulations to understand the effect of input parameters. A highlighted scientific impact includes the 2023 Gordon Bell Prize winner's simulations of a magnesium system of nearly 75,000 atoms to achieve near-quantum accuracy. In this example, the scale of computation matters greatly to achieving high fidelity answers.

A key area of research which takes advantage of Frontier's memory, compute speed, and multi-physics simulation abilities is turbulence modeling, which is a longstanding problem in classical physics. Frontier's computational power allows for more precise simulations and

approximations of turbulent systems. As demonstrated by the 2023 Gordon Bell Special Prize for Climate Modeling winners, Frontier's computing paired with a cloud-resolved model offered key insights into weather patterns and phenomena like mesoscale convective systems, which are critical for understanding events such as hurricanes and tropical storms. Frontier has also been pivotal in aerospace research, including a project with GE Aerospace aimed at reducing noise in jet engines through high-fidelity simulations of turbulence. Additionally, the National Aeronautics and Space Administration (NASA) utilized Frontier to simulate retropropulsion-based flight in the Martian atmosphere, which is a complex modeling task involving turbulent combustion and chemical reactions in the atmosphere. In the field of fusion energy, researchers used Frontier to simulate turbulence in plasmas to better understand how to stabilize and control plasmas in tokamak reactors.

Other applications of Frontier have included molecular dynamics simulations, including studies on the crystallization of the BC8 diamond, water molecule behavior, and the magnetic properties of Calcium-48. These simulations provide critical insight into atomic-level interactions, and the scale of these simulations has been useful for benchmarking and identifying simulation bottlenecks.

The high fidelity of exascale computing has opened up a wide range of possibilities and scientific demands. New demands stem from the need to accurately account for larger domains, increase throughput for high-fidelity simulations and experimental data analyses, and relax approximations that have been shown to break down as capability increases. These needs can be mapped to specific innovations and next-generation technologies, such as AI, ML, quantum computing, and increased memory bandwidth.

Vendor engagement is needed to meet these next-generation system demands and the challenges of post-exascale HPC deployments. As part of meeting these needs, the ASCAC Facilities Subcommittee recently recommended ASCR launch a comprehensive R&D program. Additionally, ORNL has recently launched the New Frontiers program to support a pipeline of perennial two-year programs focused on hardware emphasis on energy efficiency, software emphasis on sustainable software, and cross-cutting maturation of software-hardware ecosystems.

DISCUSSION

Qualters asked whether introducing specialization has been considered to overcome the limitations of Moore's Law, such as to address turbulence. **Messer** responded that specialization is being considered, especially with New Frontiers. However, there is a balance to maintain, as many multi-physics simulations have generalized needs and require the ability to perform disparate algorithms. A potential use of special-purpose hardware may look like a quantum computer attached to an HPC resource, though this approach requires further testing.

Herrera inquired if experimental validation has been conducted of diamond substrate growth from the simulations. **Messer** clarified that while the simulations explored the properties of BC8, it is believed that BC8 only occurs in the interior of giant extrasolar planets. No physical experiments had been conducted to grow the material, although it may be possible to manufacture the material under future conditions.

Thomas asked about the allocation of ORNL's budget across personnel, energy, maintenance, and software related to operations. **Messer** noted that the budget is roughly equally partitioned across these four categories.

Torres inquired about the scalability of physics simulations in parallel, for example whether using 100 processors result in a 100x speedup for the turbulence simulations. **Messer** indicated that all simulations discussed used at least 3,000 nodes, although this is in part due to the self-selection of problems which effectively utilize Frontier's large capacity.

Germann asked how concerns related to resilience and power for exascale computing have evolved over the last fifteen years. **Messer** said scientists in the early days relied upon a sophisticated checkpoint restart system. In current day, the system has gotten better with improved time between failures, and researchers are not complaining about runtimes. The machine is at the edge of capability by design, in order to keep the work progressing even if there are difficulties with the machine. **Germann** asked if any interesting trends have emerged from power usage monitoring efforts. **Messer** said disentangling energy usage at different points in the system is a good first step to understanding energy usage. A future direction would be to give users the tools to measure energy consumption.

Arthur questioned to what extent New Frontiers is an independent ORNL effort and if there were parallel efforts. **Messer** clarified New Frontiers is an ORNL effort, although parallel efforts may emerge.

Qualters asked about lessons learned that relate to storage. **Messer** answered that storage is often affected first because there is a strong predilection to use the machine to its fullest capacity. storage is hit first. Key lessons learned relate to Lustre striping and how to best use Orion, Frontier's large file system.

Taylor commented that much energy efficiency in hardware comes from reduced/mixed precision and asked about future directions for energy efficiency. **Messer** said there is evidence that has shown the drop from double precision to single precision can realize significant performance and energy use advantages. For practitioners, it is recommended to assess what level of precision is actually needed. Tools are likely needed to determine what data structures and code can be reduced precision and what need to remain double precision.

Berzins asked if the demand to simulate larger number of atoms is going to grow with the number of machines available. **Messer** said there is still room to move to larger systems by an order of magnitude, bring down the total time to solution, and obtain more precise solutions. Additionally, there is room to improve the quantum accuracy line. **Berzins** questioned how architecture could enable the use of a higher percentage of peak. **Messer** noted that memory capacity and bandwidth is one of the primary considerations for practitioners, and that it is important to be able to map between high dimensional manifolds in a controlled and explainable way. **Berzins** asked about the potential to include specialized AI accelerators and quantum accelerators. **Messer** said modern microprocessors are AI accelerators to an extent, and that quantum outriggers will be one of the first options considered because even non-quantum problems can contain embedded quantum problems. Lastly, **Berzins** commented that Moore's Law may not be entirely dead as new innovations arise. **Messer** agreed that there are still performance increases to be gained through the marketplace and ingenious methods.

CHARGE TO REVIEW THE COMPUTATIONAL SCIENCE GRADUATE FELLOWSHIP, Irene Qualters, LANL

On July 26, 2024, SC charged ASCAC with conducting a review of the effectiveness and impact of the CSGF program over the last decade. ASCAC last performed a review of the CSGF program in 2011. The charge identifies eight areas of interest related to 1) whether CSGF provides students with an effective and impactful program of appropriate quality and breadth, 2) whether there is a unique role for CSGF in the landscape of federal graduate fellowship programs, 3) whether the program is attracting diverse applicants and making awards to diverse cohorts, 4) how CSGF can reach a broader applicant pool, 5) whether the program is appropriately tailored to support the computational scientist workforce needed at DOE laboratories, 6) the most effective governance model for the program, 7) how the CSGF should evolve to ensure the best experience for students, and 8) whether the program appropriately supports students at institutions historically underrepresented in the federal research landscape.

A broadly representative charge committee assembled in September 2024, with Irene Qualters and Valerie Taylor serving as chair and co-chair, respectively. Subgroups have been established within the charge committee to address the eight charge questions. At time of presentation, the charge committee has met with CSGF leadership and provided them with a set of questions, met with DOE CSGF Project Managers (PMs), and is currently focused on synthesizing the available information and identifying any further data required. A report draft is due by end of October 2024, with the final report due by end of January 2025.

DISCUSSION

Giles commented that there has been substantial activity within the CSGF program since it was last reviewed in 2011, and data from various self-studies conducted over the past decade could be used to address the charge. **Qualters** agreed and noted the principal investigators (PIs), and PMs have been responsive in supplying data. However, although there is a significant amount of information available (such as a recently concluded longitudinal study), some charge questions do not have data readily available.

LEADERSHIP COMPUTING AT NNSA, Si Hammond, NNSA

NNSA's Advanced Simulation and Computing (ASC) program enables fully integrated multi-physics modeling and highly predictive simulation of the U.S. nuclear stockpile to ensure safety and security. NNSA's capabilities have developed alongside ASCR's capabilities, especially through joint initiatives such as the development of exascale platforms. NNSA's exascale platform, El Capitan, will be online shortly.

ASC's simulation tools deliver verified and validated physics and engineering codes to enable simulations and risk-informed decisions, support stockpile needs including design weapons modification, and conduct an annual assessment. Currently, ASC is in the process of revising its strategic guidance. Within ASC, the Capabilities for Nuclear Intelligence is a new program which underpins assessments for overseas intelligence and the Computational Systems & Software Environment program deploys HPC resources and supports users. ASC has a tri-lab computing environment with HPC platforms at LANL (*Crossroads*), LLNL (*El Capitan*), and

SNL (*Spectra*), as well as Commodity Technology Systems (CTS) at 3 NNSA sites and additional HPC and AI hardware and testbeds.

NNSA completed a National Academies of Sciences, Engineering, and Medicine (NASEM) review of its post-exascale strategy. Recommendations from the NASEM post-exascale study are grouped into three categories: HPC Procurements & Roadmaps for the NNSA; Investment in foundational and applied R&D; and Workforce, partnerships, and training. NNSA is in the process of adjusting in response to this feedback, and one area of readjusted focus is the ASC Academic Alliance Program which will provide additional opportunities for new research areas, including in AI and ML.

In regard to procurements, NNSA has just accepted Crossroads at LANL, is in the process of accepting El Capitan at LANL, is planning procurements for Advanced Technology System (ATS)-5 at LANL, is planning CTS-3, and is evaluating vendor engagements for Vanguard-II at SNL.

ASC has had a series of congressional earmarks to develop advanced memory technologies and is currently executing the start of these contracts. Additional R&D and prototyping activities include the Next Generation HPC Networks and the Advanced Architecture Protypes programs. ASC is investing heavily in quantum computing as well as neuromorphic computing, the latter of which is an exciting opportunity to potentially lower energy and integrate with HPC.

ASC is also investing heavily in AI and ML and is working with other offices in DOE. ASC's AI for Nuclear Deterrence Strategy focuses on four main areas: material discovery, stockpile design and analysis, manufacturing and experiments, and maintenance and surveillance. FY25 includes seven pilot projects utilizing AI for ASC's scientific goals.

Beyond exascale computing, ASC seeks continued engagement across the DOE programs, other agencies, industry, and academia. Additionally, ASC is trying to engage more vendors and change to more flexible procurement models.

DISCUSSION

Taylor asked if the prototyping activities include close collaborations with laboratories, explore specialized architectures, and consider the hardware / software aspect. **Hammond** said there are a range of activities at different levels, including a hardware simulation capability designed to be scalable and explore potential architectures, as well as coarse-grained simulation and field programmable gate array (FPGA) prototyping. Following these activities, the testbed program evaluates instances of potentially viable industry hardware. Promising prototypes are carried forward to the Vanguard program to de-risk the technology in preparation for potential deployment. NNSA has a reduction HPC stack and a prototype stack for finding issues and benchmarking. Through ECP and work with SC, NNSA has developed a programming model which allows the quick transition of benchmarks and libraries to new architectures without having to re-write code.

Berzins inquired about NNSA's approach to newer architectures, future generations of graphics processing units (GPUs), and changes in software. **Hammond** said NNSA is already working with GPU vendors on what future architecture balances should be. One challenge is the GPU vendors prefer floating point (FP) 4, FP8, and FP16 as opposed to FP64. The MFEM

library is cleanly designed, with the potential to adjust implementation behind the scenes. It is challenging for NNSA to use lower and mixed precision in many cases; however, it is important to lean into mixed precision to expand the available hardware options beyond FP64. **Berzins** added there are many different potential solutions to getting better performance out of a code base, including specialized hardware and / or software approaches. **Hammond** commented it would be challenging for NNSA to change software quickly because the current software is highly capable and fully validated. For this reason, NNSA is investing in the Advanced Memory Technology (AMT) project and prototyping to advance hardware upgrades. One possible strategy is to differentiate NNSA's HPC platforms, although this approach has limitations for the extent of optimization possible. Additionally, NNSA has recently finished working on a new undisclosed technology with vendors, which will begin to be included in vendor roadmaps.

PUBLIC COMMENT

An anonymous participant asked about ASCR's view on public-private partnership in AI, especially regarding leveraging private sector expertise in AI and generative AI (GenAI) to accelerate scientific research. **Finkel** noted both strong private sector interest and strong interest from ASCR to effectively leverage and improve upon private sector investments. ASCR has encouraged the research community to work with the private sector to address scientific challenges. For example, some national laboratories have engaged in innovative collaborations, such as the partnership between Pacific Northwest National Laboratory (PNNL) and Microsoft launched in 2024 focusing on AI and materials science. Additionally, many companies leading the development of AI technologies recognize that they may lack discipline-specific scientific expertise, such as drug discovery or materials design expertise. The ASCR portfolio currently includes collaborations between industry partners, laboratories, and academic partners to address basic scientific research challenges, and this type of multi-stakeholder engagement is likely to increase in the future.

An anonymous participant inquired how the increased availability of cloud-based AI and computing services impacts ASCR's investment plan on hybrid HPC-Cloud computing. **Finkel** described how the plan has many stakeholders. On the research side, ASCR invests in a software ecosystem which supports a variety of different architectures. Additionally, the research community is exploring how the exascale software ecosystem can run on cloud environments, works in containers, be modularized, etc. to support different use cases. The cloud offers unique AI capabilities, and many commercial AI services are also available via application programming interfaces (APIs). Because science occurs at all scales through a variety of diverse collaborations, ASCR's goal is to invest in a software ecosystem that allows transitioning to and from cloud environments.

An anonymous participant asked about the role for software stewardship going forward. **Finkel** acknowledged the substantial history of investment that has gone into the software ecosystem (including mathematical solvers, performance monitoring, computer science libraries, compilers, tools, and debugging). ASCR has invested in this ecosystem for many years and realizes that the broad scientific enterprise is increasingly dependent upon the software ecosystem. On the research side, ASCR's role is ensuring the software ecosystem works for next-generation computing systems. This requires investment to ensure it is understood how to move

various pieces of the software ecosystem to the next-generation, which can be a complex task as next-generation systems may have novel operating principles. ASCR understands the importance of engaging stakeholders and forming partnerships to support the needs and direction of the wider community.

An anonymous participant questioned how the Office of Critical and Emerging Technologies will use the input from the recent RFI, as well as how it relates to funding for FASST. **Fu** responded that the Office of Critical and Emerging Technologies plans to use the input from the RFI to inform DOE-wide (including SC) scenario planning for FASST. Additionally, the input may inform potential stakeholder partnerships.

An anonymous participant inquired if the CSGF could modify its requirements related to time fellows must physically be in a laboratory to better support students who are married. **Chalk** acknowledged that this is a challenge common to many fellowships, and that the CSGF is an educational fellowship rather than an employment relationship. The CSGF steering committee has thought about what flexibility they can provide to accommodate changing life situations, and the charge committee will likely also consider flexibility as well. As an example of flexibility, the CSGF has allowed virtual practicums. **Qualters** noted the CSGF requires only one summer of practicum at a laboratory, however the Science Graduate Student Research (SCGSR) program may have the physical requirement referred to in the question.

Neeraj Kumar asked how collaborations between hardware vendors or AI partners and national laboratories can be accelerated, specifically in exploring the effective integration of coarse-grained simulations, FPGA prototyping, spike-grouping, and GPU-based simulations to optimize the performance of specialized architecture. **Hammond** emphasized that pilot projects are an effective way to accelerate collaboration, demonstrate capability, clarify objectives, and align shared goals between vendors and academic partners. Currently, NNSA is working on technologies independently, but as these efforts mature, they will explore how to integrate and combine them in the future.

Berzins adjourned the meeting at 4:36 p.m. ET.

Friday, September 27, 2024

BASIC RESEARCH NEEDS WORKSHOP ON QUANTUM COMPUTING AND NETWORKING, Ojas Parekh, Sandia National Laboratories

Parekh thanked the Quantum Computing and Networking Basic Research Needs (BRN) Workshop organizing and report-writing team for their efforts.

The grand challenge identified through the workshop is to demonstrate an end-to-end rigorously quantifiable quantum performance improvement over classical analogs, especially for problems of practical value. While not surprising, this goal represents a significant challenge. The BRN workshop took the approach of identifying five Priority Research Directions (PRDs) across a general computing stack represented by the following layers: applications, programming models, algorithms, compilation, error resilience, and hardware architectures. It is essential to have synergy and tight integration across the computing stack in order to realize quantum advantages.

Quantum algorithms can be designed for ideal abstract quantum computers, for physically inspired abstract quantum computers, and for physical quantum computers. Each of these approaches has unique goals and challenges, and the current approach for physical quantum computers is to find empirical demonstration of quantum "wins." For quantum computing to show true advantages, the quantum algorithms must demonstrate unique quantum mechanical properties which would not be similarly capable via a classical alternative. If an advantage is discovered which is not based on a quantum mechanical property, then the advantage represents an inefficient allocation of classical compute.

Useful quantum advantages need to satisfy many conditions: they must be efficient; they must address a problem where there is an exponential advantage large enough to justify the overhead involved in quantum computing; there must be evidence that no comparable efficient classical algorithm exists; and there must be applications where the advantages can translate into impact. The ideal quantum advantage is an exponential advantage against the best possible classical algorithm. Achieving these types of advantages is possible when expanding areas of consideration beyond just quantum speedups to include possibilities related to communication, energy consumption, memory, space, and other resources. Additionally, classical problems may not be the best suited problems to achieve quantum advantages. For example, quantum algorithms excel on a quantum linear systems problem using succinct representations as opposed to a conventional linear systems problem using matrices, which has implications for factoring and physical simulations problems.

Space-efficient sublinear algorithms can minimize space and limit the number of qubits, which cuts down on cost. A model which achieves space advantages by not explicitly storing data is the streaming model, where the dataset is built by a stream of small updates and an answer is expected at the end of a stream. Applications for streaming algorithms include graph problems, such as communication networks, social networks, and sensor networks. It has been known since 2008 that there are exponential quantum streaming advantages for graph problems, however useful applications of this advantage to relevant problems have been limited. In 2021 the first natural problem was found to utilize this advantage, a polynomial advantage for triangle counting, although this advantage was not significant enough to justify the overhead. In 2022, it was found that there is no quantum advantage possible for the Max Cut partitioning problem. However, in 2023 an exponential advantage for the directed Max Cut problem was found, marking the discovery of the first such quantum approximation advantage. The directed Max Cut problem allows for modeling one way communication and is useful in modeling transportation networks. This discovery suggests the search for quantum advantages may benefit from focusing on adjacent problems rather than conventional problems. In addition, quantum advantages appear to be highly sensitive to the specific problem and inputs. Quantum advantages might be missed due to a bias towards solving certain kinds of problems.

Five PRDs were outlined as critical areas of focus moving forward. PRD 1 is the development of end-to-end software toolchains to program and control quantum systems and networks at scale. PRD 2 is the use of efficient algorithms to deliver quantum advantages. PRD 3 is the benchmarking, verification, and use of simulation methods to rigorously assess quantum advantages. PRD 4 is reducing noise and improving resilience through error detection, prevention, mitigation, and correction. PRD 5 is the development of hardware and protocols for

next-generation quantum networks, as well as the potential for quantum networks to provide distributed computing advantages.

DISCUSSION

Herrera asked if the pursuit of random circuit walks and boson sampling was helpful or harmful to the goal of trying to find useful algorithms. **Parekh** said that finding new avenues for quantum advantage is exciting scientifically. Additionally, some utility demonstrations will likely rely heavily on sampling tasks.

Leung inquired if classic nondeterministic polynomial (NP) problems are being considered, such as scheduling and configuration. **Parekh** noted the limitations of quantum computing to provide rigorous advantages to NP-hard problems and highlighted the promise of approximation methods. NP-hard problems should be considered carefully, especially regarding opportunity costs, scalability, and what can be solved classically.

Taylor asked if the software stack considered by PRD 1 is inclusive of HPC and quantum integrated systems. **Parekh** said PRD 1 is inclusive of these hybrid systems, and that hybrid methods are a theme throughout the PRDs. It may be advantageous to leverage classical resources while transitioning to quantum computing, although designing a software stack which incorporates classical and quantum methods has its own challenges.

Leung inquired if workforce development was considered as necessary to building the future quantum enterprise. **Parekh** stressed the importance of creating opportunities and pathways for people without prior quantum expertise to contribute to quantum-related problems. For example, applied mathematicians could be engaged through work on hybrid approaches or by contributing to a non-quantum-specific piece of a larger quantum-relevant problem. **Leung** asked about opportunities to engage and encourage students to enter the field of quantum computing. **Parekh** explained that students are generally interested in quantum and are regularly engaged through traditional means.

Shende asked if work being done in Europe to program quantum device using LLVM compilers and OpenMP would be feasible for domestic adoption or more readily integrate with the U.S. HPC ecosystem. **Parekh** replied that leveraging existing classical resources is an area of interest. Many projects use LLVM, and some projects are using Open Multi-Processing (OpenMP). With these approaches, additional work will need to be done to ensure that the quantum hardware is appropriately controlled.

In relation to PRD 1, **Windus** asked about the tension of maximizing the productivity of the emerging field of quantum computing software stacks while avoiding locking in a specific software stack before the appropriate time. **Parekh** said this tension is being considered, and that it is important for the community to converge deliberately and mindfully.

Berzins inquired about the realistic timeline for usability. **Parekh** responded the timeline depends on goals, with empirical evidence for promising quantum hardware is currently happening. The ultimate goal for a quantum utility based on a quantum mechanical advantage will likely require appropriately addressing fault tolerance, so a timeline is difficult to predict. Some recent theoretical work has shown that noise can make things classical.

UPDATE FROM ASCR FACILITIES DIVISION, Ben Brown, Advanced Scientific Computing Research

Highlights of ASCR's dynamic FY24 include: coordination of vendor engagements for major upgrade projects; engagement in the once-per-decade SC Facilities charge; launch of the Integrated Research Infrastructure (IRI) program; start-up of the HPDF project and kickoff of the IRI / HPDF coordination effort; engagement in the National Artificial Intelligence Research Resource (NAIRR) Pilot; advancement of strategic software stewardship through the ASCR Facilities Software Task Force; deliverance of the ASCR Leadership Computing Challenge allocations into the exascale era; and the continued work to deliver scientific impact across a range of efforts.

The ASCAC Facilities charge report is has important high-level recommendations, and the report is informing the thinking of the ASCR Facilities Division.

Regarding upgrade project timelines, the ASCR Facilities Division expects Aurora to be operational by CY 2025. The NERSC team is planning to extend Perlmutter through mid-2028 while the NERSC-10 system is targeted for early user access in 2027 until it begins operations in the first half of 2028.

In May 2024, the Aurora Supercomputer at ALCF achieved #1 ranking in the AI performance High-Performance Linpack Mixed-Precision (HPL-MxP) benchmark and #2 ranking overall at the International HPC Conference. The Early Science projects using Aurora are leveraging the scale of the system for AI in a compelling variety of applications. In addition, the ALCF AI testbed is a significant contribution for the community to explore the next-generation of AI-accelerator machines.

At OLCF, the Summit system's lifespan had been extended by one year. This extension unlocked potential to advance AI research and early IRI efforts. The SummitPLUS allocation call exemplified the commitment to deliver a strong return on taxpayer investment. Summit has been a powerful and reliable instrument for scientific discoveries, as evidenced by the strong record of Gordon Bell Prizes emerging from Summit. In addition, the New Frontiers Vendor R&D Program represents a new chapter in OLCF's approach to simulating vendor engagements of a variety of sizes and longevities. Lastly, the Quantum Computing User Program (QCUP) continues to be a leader in organizing thought and approaches to exploring the nexus of QIS and HPC systems and users.

ESnet is a research network of 15,000 miles of dedicated fiber cables. In 2023, 1.7 exabytes were transmitted across ESnet. Recently, ESnet has expanded its transatlantic network for the High-Luminosity Large Hadron Collider (HL-LHC) and ITER era by securing a 15 year lease linking New York, Dublin, and London. DOE's Advanced Research on Integrated Energy Systems (AIRES) project leverages ESnet for real-time power grid monitoring. ESnet is collaborating with other ASCR facilities to create a testbed for conceptual design of IRI technologies by allowing for coast-to-coast detector-to-compute real-time streaming load-balancing as well as the streaming of synthetic data. Lastly, ESnet's Culture & Engagement Program undertook a year-long all-staff effort to define core values.

Celebrating its 50th anniversary in 2024, NERSC annually serves apprixmately10,000 users across about 800 institutions and national laboratories. Close to half of the users served by NERSC are graduate students and postdoctoral researchers, which represents the essential role

NERSC plays in early career development. Highlighted research conducted using Perlmutter ranged from the discovery of a unique gravitational lens to the AI-based design of synthetic biology. A key IRI highlight is the automation of the Doublet III-D (DIII-D) Tokamak plasma reconstructions. Additionally, the AI@NERSC and Quantum@NERSC efforts provide structured approaches to engaging and training the NERSC community.

In FY25, the ASCR Facilities Division will approach strategic thinking as informed by the FY24 ASCAC Facilities charge report and by the vision of a collaborative and mature ASCR facilities ecosystem.

DISCUSSION

Leung appreciated the NERSC data regarding postdoctoral researchers and asked if the data includes more granular details such as demographics and institutions. **Brown** replied that the Office of the Deputy Director for Science Programs curates user statistics guidance, and over the last three years there has been a mature effort to achieve a finer grain resolution for these types of descriptive statistics. Additionally, the user statistics information is publicly available on the SC website.

Berzins stressed the importance of software as a key differentiator for AI and asked how DOE approaches participation in an AI landscape which already has dominant software companies. **Brown** recommended DOE and ASCR acknowledge and leverage the tools currently in the marketplace which users find appealing while also taking action to promote a diverse and healthy ecosystem of software options.

THE NATIONAL AI RESEARCH RESOURCE (NAIRR), Katie Antypas, National Science Foundation; Ben Brown, ASCR

NAIRR is the infrastructure component of NSF's initiatives in AI. A few years ago, a NAIRR taskforce was charged with investigating the feasibility and creating an implementation plan for a national AI research resource. The vision of the program was to provide national infrastructure which connects research communities to the necessary computing, datasets, software, and expertise to advance the AI ecosystem. The NAIRR taskforce report laid out four key goals: spur innovation, increase the diversity of talent in AI, improve U.S. capacity for AI R&D, and advance trustworthy AI.

The 2023 AI Executive Order directed NSF and collaborating partners to launch a NAIRR Pilot, which officially launched in January 2024. In order to pilot the large vision described in the NAIRR taskforce report, the NAIRR Pilot's strategy is to investigate all major components of the NAIRR vision at a limited scale or scope. Additionally, the pilot has sought to leverage partnerships and adopt a pioneering approach while adhering to trustworthy AI principles. In addition to NSF, there are twelve partnering agencies and twenty-six partnering non-governmental organizations. The partners are making contributions to the NAIRR Pilot in-kind or with existing resources. The agencies are contributing to different aspects of the pilot based on their expertise: DOE and NIH are leading the NAIRR Secure thrust; many agencies are providing datasets and models; NSF, DOE, and private sector partners are providing computing resources, including DOE's Summit system; multiple agencies are providing software and training; and all of the agencies are contributing to governance and the steering committee.

Pilot users include AI researchers, domain scientists, students, and educators based at U.S. institutions. Through an online portal, users can access resources including computing resources, testbeds, datasets, models, software, user support, and training. Pilot development activities include community outreach, the NAIRR Secure thrust, the NAIRR Classroom thrust to equip educators, and demonstration projects.

The NAIRR Pilot aims to demonstrate the value of the larger NAIRR initiative by supporting novel AI research, reaching broad communities, and gaining experience advancing and refining the design of the full NAIRR. Many NAIRR Pilot activities are currently underway, including matching researchers to resources, operating the pilot portal, and providing user support and community building. The first awards have been announced for the NAIRR Classroom initiative, and the first projects have been announced for the NAIRR Secure and Open Demonstration Projects. Additional community efforts include expanded outreach to communities across the country, the creation of the NAIRR Advisory Subcommittee, a recent NSF / NIST-led workshop on trustworthy AI, and an upcoming NSF / DOE workshop on planning the NAIRR software stack. Furthermore, the steering committee is developing a process by which non-governmental datasets can be contributed to the NAIRR.

DOE facilities and ASCR have made substantial contributions to the NAIRR Pilot through access to leadership, computing resources, AI testbeds, and data assets. Specific contributions from OLCF include access to the Summit supercomputer and the CITADEL secure computational environment. Additionally, ALCF has contributed AI testbeds as well as the *Intro to AI-driven Science on Supercomputers* training series.

DOE is partnering with the NIH on the NAIRR Secure Pilot, which seeks to enable research with protections and sensitive data as well as refine the requirements and infrastructure design patterns for the future NAIRR Secure resources. The NAIRR Secure Pilot has three demonstration projects related to synthetic data generation, metrics for assessing the fidelity of synthetic data, and the use of large language models to assist with biomedical question answering.

With regard to NAIRR's role in managing data, a number of agency datasets are available for analysis through the pilot. Additionally, the NAIRR Steering subcommittee on Data and Models is working on a process for non-governmental datasets to be contributed to the Pilot. At a high level, NAIRR is not envisioned to fund the collection or creation of specific community datasets or to define dataset standards, which evolve and are best defined by communities. NAIRR is envisioned to set guidelines and criteria for dataset inclusion, support data search and discovery, incentivize contribution of analysis-ready datasets, provide technical expertise and user support for data communities and community driven curation efforts, and provide access to computing, data platforms, and restricted datasets. Example data demonstration projects include: Pelican Platform, which co-locates data and compute; the National Data Platform, a data discovery system prototype connected to compute; and Sage National Discovery Cloud for Climate (NDC-C), a testbed supporting AI explorations at the edge with the connections to larger scale computing, networking, and data infrastructure.

The NAIRR Pilot currently offers awards and resources tailored for educators, researchers, and those seeking to use data and models. Computational resources, cloud-based

services, API access to closed models, and a variety of additional resources are also available to the community via partner contributions.

The approximately 150 NAIRR Pilot projects represent a range of compute request sizes. Of the ten projects with a compute size greater than 100K Ampere 100 (A100) GPU normalized hours, five were supported by Summit, four were supported by Nvidia, and one was supported by NSF. In total, NSF is supporting approximately 100 of the small- and mid-sized compute projects, as well as the projects not requiring compute including NAIRR Classroom projects, model access, and novel architecture hardware projects. Overall, the majority (69%) of projects are focused on core and fundamental computer science and AI research without a particular application. When analyzing the projects by computational hours, 33% are allocated to computer science and AI, 17% are allocated to ecology, 14% are allocated to geology and Earth sciences, 12% are allocated to astronomy, 10% are allocated to materials engineering, and the remaining hours are allocated to biochemistry and molecular biology, health and clinical sciences, performance evaluation and benchmarking, and other topics. This breakdown represents the reach the NAIRR Pilot has into different research communities.

Highlighted DOE projects include a foundational model for aquatic sciences, an Earth digital twin, and an AI project on unified representation learning. Highlighted NSF-supported AI projects include training foundation models from private federated client data, enhancing language model safety and trustworthiness, evaluating accuracy and bias of compressed language models, developing reliable agents for complex digital tasks, and designing neuro-inspired oversight for safe and trustworthy large language models.

In terms of geographic and funding diversity, current researchers represent thirty-four states and a variety of additional funding resources. The majority of researchers are associated with R1 research universities; however, there are a number of researchers from smaller universities and colleges. In addition, 10% of awards have gone to researchers in EPSCoR states, and 20% of awards have gone to minority serving institutions.

The first four NAIRR Classroom awards were recently announced, and the recipients are educators at four public universities with class sizes ranging from 25 to 800.

NSF announced an opportunity (NSF 24-109) to expand the NAIRR Pilot community and promote and diversify AI education, workforce development, and broaden access to AI resources. In July 2024, NSF hosted a workshop on Community-Informed Policies and Best Practices for NAIRR. A future NAIRR Software Workshop will be hosted December 2024. In addition, the NAIRR Pilot is hosting several upcoming events and training series. Lastly, over the next 18 months the NAIRR Pilot will integrate partner contributed resources, mature operations, deploy and support demonstration projects, sponsor an annual meeting in 2025, support the Pilot user community, build on community outreach and education, develop metrics for assessment, identify gaps, and provide an interim report on Year 1 experiences.

DISCUSSION

Qualters asked about the biggest surprise encountered during the pilot. **Antypas** highlighted the unexpectedly large number of biomedical and health researchers, including those looking for secure resources that can handle sensitive data like Personally Identifiable Information (PII). Additionally, both the interagency steering committee and industry partners

have been highly engaged in the project, which has underscored the importance of strong public-private partnerships.

Qualters inquired about the biggest barriers to accessing these resources. **Antypas** emphasized that the primary challenge lies in the expertise required to use the resources effectively. The challenge is not in allocating computation hours, but in training and partnering with communities new to AI in a sustained way. **Qualters** asked for the percentage of users who are new to using these computing resources. While a precise figure was not immediately available, **Antypas** recognized the importance of this question and noted to follow up on the matter.

Brower-Thomas inquired if the training opportunities were limited to undergraduate and graduate students, or if postdoctoral and early career researchers could participate as well. **Antypas** clarified the training is open to all, including senior researchers who are interested in incorporating AI into their research. Additionally, the NAIRR Classroom resources are not limited to undergraduates and can be available for a variety of different types of training or community workshops.

Arthur asked if NAIRR and HPDF may limit cross-disciplinary or cross-agency collaborations by avoiding the standardization of data sets. **Antypas** responded that while standardization is important, standards should emerge from the specific science communities themselves rather than being dictated by infrastructure providers. Platforms can incentivize standardization, but the impetus must come from within the scientific disciplines

An anonymous participant questioned how outcomes of the NAIRR investments in universities and classrooms are evaluated. **Antypas** explained the pilot aims to assess whether communities which are new to incorporating AI continue to engage with federal grants and the broader research community. While long-term impacts may be difficult to measure during the pilot, resource usage and engagement metrics will be tracked.

BUILDING ON SUCCESS: ADVANCING PRIVACY-PRESERVING FEDERATED LEARNING WITH DISTRIBUTED OPTIMIZATION, Kibaek Kim, Argonne National Laboratory

Kim presented recent research and advancements in privacy-preserving federated learning (FL), highlighting key developments, challenges, and applications to the DOE. FL is a distributed learning paradigm which allow for training AI and ML models on distributed data sources collaboratively.

A FL approach has many benefits, particularly in areas of privacy, efficiency, and scalability. Regarding privacy, because FL trains models locally, data owners are not required to share data and are able to protect sensitive information. As for efficiency, only model updates are shared across clients, which reduces data transfer and saves storage and bandwidth. For scalability, FL can scale to handle large and complex training tasks by leveraging multiple computing resources. FL is a relatively new area of research, and it can be modeled as a distributed optimization problem.

Privacy-Preserving FL (PPFL) has several applications across DOE domains, such as facilitating collaborative experiments and research using multimodal data while preserving data privacy. This allows multiple institutions to collaborate without sharing experiment data. In the

example of climate science, research centers can securely work together on climate models without sharing raw data. As another example, electric grid data analysis can benefit from PPFL by maintaining consumer data privacy in analyses of electricity consumption patterns.

Although raw data remains on client devices, without proper safeguards like a privacy-preserving scheme, FL risks the potential for model parameters or updates to leak sensitive information. Attackers with access to model parameters could successfully reconstruct raw data. However, by employing differential privacy to add noise to model updates, attackers are not able to accurately reconstruct raw data.

A related challenge is the privacy-utility tradeoff, where the goal is to achieve an optimal balance of both privacy and performance in algorithm design. By inserting noise into the model at different injection points (before training, during training, at model output), Kim and collaborators demonstrated that adding noise during training outperforms the state-of-the-art approach of adding noise to the model output, and the combined approach of adding noise during training and having multiple local updates outperformed all approaches tested. This design achieves better performance without compromising privacy.

Other challenges of FL include the heterogeneous computing environment (where client machines can have widely varying capabilities) and client drift (where local models can drift into misalignment from other local models and the global modal). To address the challenge, Kim and collaborators designed a FedCompass algorithm to adaptively synchronize and manage clients with varying computing power and an Asynchronous Exact Averaging (AREA) algorithm to perform client drift correction in a secure manner. These approaches achieved faster training and higher accuracy than current state-of-the-art methods.

The open-source software package APPFL v1.0 (Advanced Privacy-Preserving Federated Learning v1.0) was released August 2024. APPFL v1.0 offers developers the ability to design, simulate, and evaluate new privacy and FL algorithms. For users, this software package supports the deployment of secure, scalable FL experiments across distributed clients.

Two DOE use cases of PPFL were highlighted: performing load forecasting based on heterogeneous building data and integrating complementary X-Ray tomography and fluorescence data for combined analysis of material properties. Both examples demonstrate how privacy can be retained amongst collaborators without losing significant model performance.

Future research directions include exploration of large AI and foundation models, incentive and fairness to FL clients, privacy preservation at scale, synthetic data generation, sustainable and robust workflows, and interdisciplinary collaborations across applied math, computer science, and DOE facilities.

DISCUSSION

Taylor asked about the potential for hierarchical PPFL at scale, such as privacy structured by geographically defined regions. **Kim** affirmed this possibility and noted the neutral network structure could incorporate geographic distinctions to preserve privacy at different levels.

Qualters inquired about standardized norms for evaluating FL models beyond binary evaluations, which could help evaluate risk. **Kim** replied that while this is a good idea, there are currently no established standardized metrics in place due to the novelty of the research area.

PUBLIC COMMENT

Harvey Newman asked how university-based and international groups involved in high throughput data intensive systems can become involved in the initiatives discussed by Brown. For example, one such effort by volunteers is the Global Network Advancement Group and its R&D working groups. **Brown** replied that the IRI leadership group has a plan to establish a community web presence and a Slack instance for virtual engagement. IRI and HPDFs start with a DOE-local context and expand outwards from there, so there are certain aspects of these efforts closely linked to DOE mission drivers. The IRI has an outreach and engagement subcommittee led by Eli Dart of ESnet and Rafael Da Silva of ORNL. Additionally, IRI representation will be present at the Innovating the Network for Data-Intensive Science (INDIS) workshop and at the DOE booth at the SC24 conference in Atlanta, GA in November 2024. Furthermore, the ESnet Confab25 meeting in the San Francisco Bay area in April 2025, will include a rich IRI technical element.

Suzanne Sincavage inquired how nonprofits can support NAIRR. **Susut** encouraged direct communication with NAIRR and recommended the NAIRR website as a resource for more information.

Berzins adjourned the meeting at 12:38 p.m. ET.

Respectfully submitted on October 21, 2024
By Aiden Layer, M.L.I.S.
Science Writers, Oak Ridge Institute for Science and Education.