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

PUBLIC MEETING MINUTES

January 16 - 17, 2025

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 January 16 - 17, 2025 at the Hilton Washington D.C., 1750 Rockville Pike, Rockville, MD 20852 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
Sunita Chandrasekaran, University of Delaware
Timothy Germann, Los Alamos National Laboratory (LANL)
Roscoe Giles (Vice Chair), Boston University
Anthony Hey, University of Washington
Alice Koniges, University of Hawaii
Satoshi Matsuoka, Rikagaku Kenkyusho (RIKEN) Center for Computational Science
Vanessa Lopez-Marrero, Brookhaven National Laboratory (BNL)
Irene Qualters, LANL
James Whitfield, Dartmouth College

ASCAC members present virtually
Richard Arthur, General Electric (GE)
John Negele, Massachusetts Institute of Technology (MIT)
Sameer Shende, ParaTools Inc.
Krysta Svore, Microsoft
David Torres, Northern New Mexico College
Juan Torres, National Renewable Energy Laboratory (NREL)
Theresa Windus, Iowa State University
Cristina Thomas, 3M Company (retired)

ASCAC members absent
Vinton G. Cerf, Google Inc.
Alexandra (Sandy) Landsberg, Office of Naval Research
Jill Mesirov, University of California, San Diego
Vivek Sarkar, Georgia Institute of Technology

Office of Science personnel present in person

Harriet Kung, Deputy Director for Science Programs (DDSP) Office of Science (SC)
Benjamin Brown, ASCR
Christine Chalk, ASCR
Tanner Crowder, ASCR
Hal Finkel, ASCR
Marco Fornari, ASCR
Carol Hawk, ASCR
Saswata Hier-Majumder, ASCR
Margaret Lentz, ASCR
Christopher Myers, ASCR
Kalyan Perumalla, ASCR
Robinson Pino, ASCR

David Rabson, ASCR
Julie Stambaugh, ASCR
Ceren Susut, ASCR
Jordan Thomas, ASCR
Abre Valencia, ASCR

Presenters

Giuseppe Barca, University of Melbourne
Susan Clark, Sandia National Laboratories (SNL)
Michael Cooke, SC
Sudip Dosanjh, National Energy Research Scientific Computing Center (NERSC)
Ana Gainaru, Oak Ridge National Laboratory (ORNL)
Cindy Rubio González, University of California, Davis
Andrey Lokhov, LANL
Joseph Lukens, ORNL
Victor Mateevitsi, Argonne National Laboratory (ANL)

Attending in-person

Gina Adam
Lura Biven
Susan K. Gregurick
Susannah V. Howieson
Earl Joseph
Arthur Maccabe
Esmond Ng
Michael L Parks
Griffin Reinecke
Ed Seidel
Valerie Taylor
Yuping Zeng

There were approximately 260 individuals, including those that attended virtually, present for all or part of the meeting.

Thursday, January 16, 2025

OPENING REMARKS, Martin Berzins, Chair of ASCAC

Berzins convened the meeting at 10:02 a.m. Eastern Time (ET), welcomed attendees, and briefly reviewed agenda.

VIEW FROM WASHINGTON, Harriet Kung, DOE Office of Science

Kung is currently the Deputy Director for Science Programs (DDSP), as stipulated by The Federal Vacancies Reform Act of 1998, and will resume the role of Acting Director upon approval of the incoming presidential administration. The DDSP organizational chart and recent personnel changes were reviewed.

Several SC highlights from 2024 were reviewed. A sampling of the major scientific discoveries and breakthroughs include two-dimensional (2D) quantum materials for secure, energy-efficient communications; the release of the largest three-dimensional (3D) map of the universe by the Dark Energy Spectroscopic Instrument (DESI); the mitigation of disruptive tearing instability in fusion plasma by an artificial intelligence (AI) algorithm; and an innovative quantum gate for advanced quantum networking for the enhancement of error resilience. Examples of enabling research and technologies which facilitate breakthroughs include a new method to produce superheavy elements by the Lawrence Berkeley National Laboratory (LBNL); the transfer of beneficial genes into plants through the use of the clustered regularly interspaced short palindromic repeats (CRISPR) by scientists at Oak Ridge National Laboratory (ORNL); the use of light detection and ranging (LiDAR) by the Wind Forecast Improvement Project (WFIP-3) to improve wind turbine estimates; and the protein design and computational analysis done by David Baker using SC facilities, which resulted in a Nobel Prize in Chemistry. Regarding tools in the SC portfolio, the Legacy Survey of Space and Time (LSST) camera was completed; Aurora, a new exascale computer was delivered; and the Advanced Photon Source was upgraded with new capabilities. Key workforce developments include new partnerships in fusion through the Milestone-Based Fusion Development Program and Fusion Innovation Research Engine (FIRE) Collaboratives, new apprenticeship programs, and continued internships.

The upcoming Basic2Breakthrough (B2B) initiative will communicate the value of DOE-funded research and national laboratories through short stories and videos. Communications will target the general public, legislators, and partners interested in DOE programs, objectives, and funding directions.

The 2024 Enrico Fermi Award winners include Héctor D. Abruña, for revolutionizing the fundamental understanding of electroanalytical chemistry and development of energy storage materials; Paul Alivisatos for developing the foundational materials and physical chemistry to produce beneficial nanocrystals and polymers; and John H. Nuckolls for leadership in inertial confinement fusion and high energy density physics, contributions to national security, and leadership of Lawrence Livermore National Laboratory (LLNL) at the end of the Cold War. The 2024 SC Distinguished Scientist Fellows were Mary Raafat Mikhail Bishai (BNL), Lois Curfman McInnes (Argonne National Laboratory), Kristin Persson (LBNL), and Gerald A. Tuskan (ORNL). The White House named 55 Presidential Early Career Award for Scientists and Engineers (PECASE) awardees, including 35 from SC.

A recent Deputy Secretary for Energy memorandum outlined a framework for Research, Technology, and Economic Security (RTES) policy and is available online.[1] The memorandum details RTES goals and commitment to the due diligence review process for financial assistance, and the mitigation and clarification / reconsideration processes. The RTES working group is responsible for policy development. The RTES office conducts due diligence reviews, and the SC program offices serve as the procurement and decision-making authority for financial assistance awards. SC engagement with RTES due diligence is driven by three principles: actions must be informed by the science, be risk-based, and done on a case-by-case basis; actions must focus on behaviors and timeliness; and actions must be characterized by fairness, transparency and non-discrimination. SC recommends universal disclosure and the use of the Science Experts Network Curriculum Vitae (SciENcv), abstains from imposing any citizenship or residency requirement on personnel working under its awards, and does not require university principal investigators (PIs) to seek prior approval for foreign travel.

DISCUSSION

Giles questioned how changes resulting from the transitioning presidential administration would be communicated to the community. **Kung** committed to the sharing of relevant information. Initial communications with the transition team indicated ASCRS's work is important and must be accelerated.

Koniges asked whether SC will aid with the creation of Basic2Breakthrough communications. **Kung** confirmed staff would provide assistance, and only details of the desired communication is required for submission.

VIEW FROM GERMANTOWN, Ceren Susut, Associate Director of ASCR

Susut reviewed staff and organizational changes in ASCR.

DOE exascale supercomputers led in performance and efficiency in 2024 and received top recognitions in the TOP500 Project's list of the world's most powerful supercomputers, the AI High Performance Linear Equation Package (LINPACK) for Accelerator Introspection Mixed Precision benchmark (HPL-MxP), and the Green500 bi-annual ranking of energy efficiency in supercomputers. Additionally, researchers using Frontier produced the largest astrophysical simulation of the universe and quantum chemistry calculations to earn the latest Association for Computing Machinery Gordon Bell Prize. Summit, which earned multiple Gordon Bell Prizes in the past, was decommissioned on November 15th, 2024. The National Energy Research Scientific Computing Center (NERSC) turned 50 in 2024 and is currently the largest SC user facility with over 10,000 annual users.

ASCR research investments aim to follow critical technology trends and seek to seamlessly integrate experiments, computations, and simulations. DOE announced $179M for Microelectronics Science Research Centers (MSRCs). MSRCs include the Microelectronics Energy Efficiency Research Center for Advanced Technologies (MEERCAT), which is committed to revolutionizing energy-efficient microelectronics by advancing integrated innovations, the Co-design and Heterogeneous Integration in Microelectronics for Extreme Environments (CHIME) Center, which aims to drive transformative advancements in extreme environment electronics, and the Extreme Lithography & Materials Innovation Center (ELMIC), which aims to advance the fundamental science driving the integration of new materials and

[1] https://www.energy.gov/sites/default/files/2024-11/DOE%20RTES%20Framework%20Memorandum%2011.26.2024.pdf

processes into future microelectronic systems.

FY25 solicitations include calls from Scientific Discovery through Advanced Computing (SciDAC) Institutes, Competitive Portfolios for Advanced Scientific Computing Research: Data Management and Visualization, and the National Quantum Information Science Research Centers.

Research highlights and community news include creation of the first AI-generated and portable basic linear algebra subroutines (BLAS) library, named ChatBLAS, to enhance productivity in the high-performance computing (HPC) software ecosystem, and the partial derivative Finite Element Method (∂FEM) algorithms for differentiating large-scale finite element applications. The DOE quantum information science (QIS) applications roadmap has been published online.[2] Two recent DOE-SC roundtables focused on Transformational Science Enabled by Artificial Intelligence, and a DOE BER/ASCR workshop on Envisioning Frontiers in AI and Computing for Biological Research will be held on February 4–6, 2025. Workshop details can be found online.[3] SC[4] and ASCR[5] will hold virtual office hours to broaden awareness of existing programs.

QUANTUM INFORMATION SCIENCE ROADMAP BRIEFING, Susan Clark, Sandia National Laboratories

Clark explained the QIS application roadmap was based on the input of 35 quantum experts, 19 of which served on the committee as authors. The final report consists of three roadmaps involving computing, sensing, and networking, and is available online.[2] The Quantum Computing to Extend the Frontiers of Computation roadmap includes simulating the strong nuclear force, chemical and material simulations, enhancing scientific computing applications and enabling more efficient management of infrastructure, logistics, and commerce. Quantum computing applications rely on the future development of a fault-tolerant quantum computer with many quantum bits (qubits). Technology milestones, including 10,000 physical qubits with low error rates, quantum interconnects for modular architecture, and large-scale architecture solutions, would enable development of very large fault-tolerant quantum computers in the next 20+ years.

The roadmap for Quantum Sensing to Enable Unprecedented Precision and New Discoveries includes precision measurements for new physics, sensitive and high-resolution probes, robust and deployed sensors, and biological, chemical, and biomedical sensing. Quantum sensing is a near-term realizable application from QIS. Classical engineering will be required to make the sensors useful, and investigations involving new quantum materials as sensors must be deployed.

Finally, the Quantum Networks to Harness the Power of Linked Quantum Resources roadmap includes linking sensors for increased resolution, linking sensors to quantum processors for finding rare or small signals, and linking quantum computers to increase processing power. Realizing the vision of large-scale quantum networks requires the development of specific quantum networking tools, including repeaters and transducers, to send fragile signals over long distances and manipulate them at the nodes.

[2] https://science.osti.gov/Initiatives/QIS/Community-Resources/SC-Sponsored-Reports
[3] https://www.orau.gov/BER-ASCR-Workshop
[4] https://science.osti.gov/officehours
[5] https://science.osti.gov/ascr/officehours

DISCUSSION

Hey recalled the mention of "many" fault-tolerant qubits required for physics applications and sought a more precise value. **Clark** explained the resource estimates vary and did not have an actual number.

Whitfield asked whether the roadmap produced any suggestions for the establishment of publicly owned resources for quantum devices. **Clark** responded commercial entities may lack flexibility or transparency. Widely accessible user facilities and test beds, and fundamental research at universities would mitigate potential access problems. However, no official recommendations were made.

Torres questioned whether quantum technology would be more energy efficient and sensitive to power quality than current HPC approaches, and whether quantum-supporting infrastructure is being researched. **Clark** noted power quality from the existing infrastructure can be improved at the endpoint. Whether quantum technology will be more power efficient is an open question.

Arthur expressed concern over a figure which implies the replacement of classical computing with quantum computing and noted the importance of correcting the implication for non-technical stakeholders. **Clark** agreed and explained the implication was unintentional.

Qualters asked whether hybrid classical and quantum systems were considered. **Clark** noted the roadmap discussed the use of classical computers to control quantum computers.

Chandrasekaran noted the lack of discussion around programming, the extensive time requirements of offloading tasks from classical to quantum systems, and the need to begin basic research on hybrid infrastructures.

SUBCOMMITTEE ON THE COMPUTATIONAL SCIENCE GRADUATE FELLOWSHIP (CSGF), Irene Qualters, ASCAC and CSGF Subcommittee

Qualters identified the goals of CSGF, which are to meet workforce needs, promote careers in computational science, provide practical work experience, and facilitate relationships between academia and national laboratories. Kung's charge to ASCAC consisted of eight questions to examine the effectiveness and impact of CSGF and the quality and breadth of the program over the past decade.

Individual charge questions and abridged subcommittee responses were presented. Question one: Does CSGF provide students with an effective and impactful program of appropriate quality and breadth? The program's value is supported by an abundance of data, including publications, careers and participant feedback. Question two: Is there a unique role for CSGF in the landscape of federal graduate fellowship programs? Of the 38 graduate fellowships encompassing computer science fields, none offer the full set of defining benefits of the CSGF pool. Question three: Is the program appropriately tailored to support the computational scientist workforce needed at the DOE laboratories? Support is successfully achieved through an interdisciplinary program of study, a 12-week practicum at a national laboratory, community building, an Annual Program Review meeting, and strong mentoring. Question four: What is the most effective governance model for the program? The governance is well-defined but lacks a visible strategic plan, measurable goals to achieve objectives, and there is limited evidence of strategic engagement with non-stakeholders. Question five: How should the CSGF evolve to ensure the best experience for students? Six areas of evolution to consider include integration of emerging technologies, professional development, community building, expansion of academic institutional reach, enhancement and coordination of student support services, and program evaluation practices. Questions six through eight dealt with topics such as the programs support of students at institutions previously with less representation in the federal research landscape, and attracting a broad applicant pool. The subcommittee recommends CSFG develop and communicate a comprehensive national talent outreach strategy to expand the fellowship pipeline.

Overall, the subcommittee found CSGF to be of high quality with broad and enduring impact and possessing multiple areas of excellence. Recommendations include maintaining areas of excellence, efficiently growing the program, and translating objectives into a strategic plan with five-to-ten-year measurable goals and priorities.

DISCUSSION

Berzins questioned whether the program's talent pool could be increased by including additional academic backgrounds. A process for mentoring students with application preparation is also needed. **Qualters** explained increasing the pool will require engagement with a broader set of schools, including engagement at the faculty level.

Koniges asked whether feeder programs existed to prepare students for CSGF and sought a comparison of CSGF to similar fellowships. **Qualters** reiterated the existence of defining benefits in CSGF not found in 38 other fellowships.

Berzins called a vote to accept the CSGF report. The report was accepted unanimously.

Berzins dismissed the meeting at 12:16 p.m. for lunch and reconvened the meeting at 1:47 p.m.

NEW CHARGE TO ASCAC, Ceren Susut, Associate Director of ASCR

Susut described the latest ASAC charge, requiring the formation of a Committee of Visitors (COV). The COV will have two tasks. The first is to assess the efficacy and quality of the processes used during FY20-23 for soliciting, reviewing, and monitoring program actions; the second is to consider which resources or technologies would help the program staff assess and manage research portfolios, and which metrics would be useful for the program to monitor. Additionally, within the boundaries defined by the DOE mission and available funding, ASCAC is to comment on how the award process has affected the breadth and depth of portfolio elements, the overall scientific and technical quality of the portfolio elements, and the different types of organizations in the research enterprise.

DISCUSSION

Berzins requested clarification on the charge's scope. **Susut** responded the charge involves all ASCR programs from the Research Division and the Advanced Competing Technology Division over the designated years.

Berzins dismissed the meeting at 1:52 p.m. for a break and reconvened the meeting at 1:59 p.m.

EARLY CAREER PANEL, Moderator: Cindy Rubio González, University of California, Davis, Panelists: Ana Gainaru, ORNL, Andrey Lokhov, LANL, Joseph Lukens, ORNL, and Victor Mateevitsi, ANL

González presented the panelists, recipients of the Early Career Research Program (ECRP) award, and facilitated introductions.

Lokhov researches statistical learning funded by ASCR, quantum computing funded by the Defense Advanced Research Projects Agency (DARPA), power grid funded by the DOE Office of Electricity (OE), and graph algorithms funded by LANL Laboratory Directed Research and Development (LDRD). Statistical learning work involves Markov Random Fields (MRFs) as a dominant paradigm for Scientific Machine Learning, and an influence in applications, such as many-body quantum physics, power grids, statistical physics, and field theories. DARPA's Quantum Benchmarking Initiative (QBI) aims to assess the feasibility of improving the timeline required to build an industrially useful quantum computer. The power grid project seeks to establish and benchmark a high-performance online sensor valuation method based on a novel idea for reliability assessment of measurements. The metric would then inform optimal sensor placement tools for enhanced grid reliability. Finally, graph algorithms research involves development of novel solutions for out-of-equilibrium dynamics of disordered systems.

Lukens conducts research on frequency-bin quantum information, quantum networking, and Bayesian quantum state tomography. Accomplishments facilitated by the ECRP include the development of the first deployed flex-grid quantum network, picosecond timing with White Rabbit, and research into secure classical communications with quantum key distribution. A Bell state analyzer for frequency-bin qubits was demonstrated, which could enable the construction of frequency-based quantum repeaters. In addition, an all-complementary metal-oxide semiconductor (CMOS) design for frequency polarization entanglement, with over 100 channels of high-fidelity polarization states, was achieved.

Gainaru conducts research on the optimization of HPC application performance and explained HPC application workflows are becoming proxies for experiments. Workflows combine high fidelity simulations, instruments, visualization services, analysis, AI surrogate models and digital twins. Interests include data fusion with extreme and small data, coupling between traditional HPC and AI, the extraction of knowledge from data, and smart querying capabilities. ASCR project experience includes SciDAC for fusion energy sciences, data management, and SciDAC Resource and Application Productivity through computation, Information, and Data Science (RAPIDS). Work highlights include complex querying for combustion simulations and optimizing SciDAC fusion applications on Frontier.

Mateevitsi works on the visualization team at ANL. One aspect of the research involves producing visualizations of large data outputs from supercomputers, which may entail writing custom rendering programs, additional data processing, or developing new visualization algorithms. Another aspect is research involving *in situ* visualizations, bidirectional steering, and concurrent central processing unit (CPU). *In situ* visualization enables the rendering of simulation data while still in memory and removes the need to stop the simulation during the lengthy data saving process. Bidirectional steering allows for the access of partial simulation results and the manipulation of parameters while the simulation is running. Concurrent CPU is the simultaneous copying of CPU data while analyzing the data. Additional work involves participation on a multidisciplinary team to apply digital twins, robotics, and haptics to isotope production, and Digital twins and AI-Enabled and Immersive Environments for Automated Scientific Laboratories (DAIMSL) to create automated, self-driving laboratories.

González asked what inspired the panelists to submit an ECRP to DOE. **Gainaru** appreciated the emphasis on collaborative research, which grants insight to the landscape outside of the personal domain. **Mateevitsi** saw an opportunity to work on big problems and have access to the necessary funds. ASCR funding has enabled many scientists to bring personal visions to reality. **Lukens** liked the focus on cutting-edge research and support for forward-thinking ideas.

Lokhov explained the timeline and level of funding provided by the ECRP award facilitated work on more challenging problems.

González sought the impact of ASCR funding on the work and research career of the panelists. **Lukens** noted the high level of research productivity and has a curriculum vitae (CV) on which a minimum of two-third of the publications are from ASCR-funded research. **Lokhov** mentioned the level of attention garnered, both internally and externally, when working on bigger projects grants valuable connections and influence in the field. **Mateevitsi** appreciates the opportunity to translate research ideas into reality and the resulting networking which leads to additional funding. **Gainaru** explained research ideas increase in applicability and impact through interactions with scientists from multiple domains.

González asked what the panelists wanted to work on and found exciting about the future. **Mateevitsi** is excited about digital twins and the utilization of AI and large language models (LLM) for visualizations. **Gainaru** is excited about digital twins, AI / machine learning, and reshaping the way computer science middleware handles and builds workflows, with the goal of delivering knowledge instead of data. **Lokhov** looks forward to the increase in computational capabilities and new ways of conducting science which will result from AI and quantum computing. **Lukens** is excited to witness how the growing interest in QIS, which continues to attract new people, skills and capabilities, will advance the field and result in novel collaborations.

DISCUSSION

Berzins inquired how the panelists currently use exascale machines and sought insight into the type of machines which will be needed to conduct research a decade from now. **Lokhov** will use Aurora to test and demonstrate the impact of new algorithms, and future research will benefit from clusters of quantum computers. **Gainaru** questioned whether the clusters will include HPCs. **Lokhov** confirmed. **Gainaru** uses Frontier to detect bottlenecks in workflow scaling. Future machines will need better storage capabilities, possibly by using deoxyribonucleic acid (DNA).

Matsuoka sought the extent to which panelists are using AI in current research. **Gainaru** noted AI is not used extensively, but usage will likely increase in the future. **Mateevitsi** frequently uses Chat Generative Pre-Trained Transformer (ChatGPT) to interface the various codes and tools used by big interdisciplinary teams, write programs for graphing data, and optimize codes.

Rannow (chat message read by **Gonzalez**) sought how correctness in research is confirmed. **Gainaru** consults domain science to establish correctness before performing optimizations.

Hey asked for the level of freedom the panelists have in choosing projects. **Lokhov** explained theoretical scientists at LANL can be fully funded through soft money, allowing freedom in project selection. **Gainaru** noted the lack of soft money at ORNL. Project choice depends in part on current ASCR solicitations.

Berzins dismissed the meeting at 3:16 p.m. for a break and reconvened the meeting at 3:46 p.m.

BREAKING THE MILLION-ELECTRON AND 1 EFLOP/S BARRIERS, Giuseppe Barca, University of Melbourne

Barca explained the presented work was recently awarded a Gordon Bell Prize in HPC and is a cornerstone for the digital chemical discovery technologies currently advanced at the University of Melbourne and by his co-founded company, QDX Technologies. Current processes/technologies for drug discovery can require10-15 years, costs up to $2.6B, and passes clinical trials only 12% of the time. Eighty percent of disease-driving proteins are “undruggable” with non-covalent therapeutics, leaving diseases like Alzheimer’s, cancers, and multidrug-resistant infections largely incurable. Covalent inhibitors, which form irreversible bonds with proteins, can target undruggable proteins. The presented work paves the way for the first accurate *in silico* software to design and model covalent binders using Møller-Plesset perturbation theory at the second order (MP2) potentials.

Results from current drug discovery simulations do not sufficiently correlate with experimental results, due to the inability to simulate atomistic models with thousands of atoms in complex environments. Simulations relying on quantum potentials yield results with improved experimental alignment and can model bond changes and chemical reactions. Outstanding challenges for quantum simulations include: a lack of scalability, as the amount of computation required increases greatly with the number of atoms in the simulation; the accuracy required to model non-covalent interactions; time evolution, as dynamic simulations rely on complex quantum mechanical gradients; and computational efficiency, due to the inability of many quantum chemistry techniques to efficiently utilize novel massively parallel processors and computer architectures.

The Extreme-scale Electronic Structure System (EXESS) was designed to solve the outstanding challenges in quantum scalability and performance. The system employs reduced computational complexity, retains the required accuracy, and efficiently exploits the computational capabilities of throughput-oriented massively parallel hardware. EXESS uses molecular fragmentation methods to break the chemical system into smaller monomer units. Compared to central processing unit state-of-the-art code (CPU SOTA), resolution of identity (RI) results were 95 times faster and 19 times more energy efficient with no reduction in accuracy.

EXESS performs biomolecular-scale quantum molecular dynamics using MP2 potentials through the combination of five technologies: molecular fragmentation, a multi-layer distributed graphics processing unit (GPU) memory and workload manager, asynchronous timesteps, a fragment-level RI-Hartree-Fock plus RI-MP2 algorithm, and runtime Double-precision, General Matrix-Matrix multiplication (DGEMM) autotuning. EXESS was three orders of magnitude faster than SOTA quantum chemistry codes. Leveraging additional GPUs would improve performance further. Applying the technique to a polymer with 232,000 atoms resulted in the first breakthrough of the million-electron barrier at the MP2 level. Applying the technique to simulations of amyloid fibril folding, implicated in Alzheimer’s Disease pathology, lead to a minimum of 1,000-fold speedup over SOTA techniques. Applying the technique to the prediction of polymorphic forms of therapeutics and organic compounds with 9,400 nodes resulted in a performance of 1.007 exaflops per second (ELOPS/s), breaking the known EFLOPS/s barrier for the first time. Efforts are underway to make the code faster.

EXESS performs quantum chemistry at scale with record-breaking performance and near-perfect scaling across thousands of GPUs. The system enables work on the grand challenges in drug discovery, enzymatic catalysis, and biomolecular science, from polymorphism and Alzheimer’s disease to the design of covalent therapeutics. EXESS sets the stage for the next generation of quantum-AI simulations. As a service to the academic community, the software is available at no cost.[6]

DISCUSSION

An attendee questioned the possibility of adapting EXESS for the analysis of plane waves, useful for material discovery. **Barca** replied that the team's current investments and expertise are in biosystems, but the system could be potentially adapted to materials science.

Chandrasekaran noted the multiple systems and GPUs which ran the software, and questioned how code portability is managed and will be maintained in the future. **Barca** explained code portability is facilitated using the Heterogeneous-computing Interface for Portability (HIP) programming model.

Hey asked if performing the presented calculations on a quantum computer would result in significantly different results. **Barca** noted the importance of the question, but the answer is currently unknown.

PUBLIC COMMENT

Seidel appreciated the day's focus on future-forward topics, such as graduate fellowships and early career awards. The DOE, SC and ASCR are stewards of the Computational and Data Sciences (CDS) community. CDS is still working to gain traction and requires support from all community members.

Berzins adjourned the meeting at 4:42 p.m.

Friday, January 17, 2025

OPENING REMARKS, Martin Berzins, Chair of ASCAC

Berzins convened the meeting at 10:00 a.m. ET.

RIKEN FUTURE COMPUTING PLAN, Satoshi Matsuoka, RIKEN

Matsuoka recounted the collaborative relationship between the U.S. and Japan and noted the Project Arrangement on High-Performance Computing and AI between DOE and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) was signed in April 2024.

The RIKEN Center for Computational Science (R-CCS) is a research institution with hundreds of personnel. Fugaku is the institution's flagship machine; it possesses exascale capabilities and is likely the largest machine in the world. Fugaku achievements include the top spot in major benchmark rankings: from June 2020-November 2021, TOP500 and High Performance LINPACK for Accelerator Introspection (HPL-AI); Graph500 and High-Performance Conjugate Gradients (HPCG) since June 2020; and ML performance (MLPerf) HPC since November 2021. In addition, the work performed on the system won a Gordon Bell Special Prize for HPC based research of the coronavirus disease of 2019 (COVID-19) in 2021 and 2022, and the system was used for weather forecasting in the Tokyo 2020 Olympic / Paralympic games. More recently, research enabled by computations using Fugaku won the Gordon Bell Prize for climate modeling in 2023. Hyperion Research assessed key Fugaku metrics and concluded the system will likely have 68-90-fold return on investment. Researchers

[6] https://exess.qdx.co

are pleased with the system's design and operations since there is a focus on high value sustainable development goals, and a focus on creating industrial economic growth.

R-CCS will have funded Fugaku until ~ 2031. A feasibility study for Fugaku-NEXT, the next-generation supercomputer, was in place from FY22-24. Research and development (R&D) for Fugaku-NEXT will occur from FY25-29, followed by deployment in FY29 and in operation in FY30. The machine is expected to achieve zettascale computing capabilities in AI. Fugaku-NEXT will utilize international hardware vendors, unlike its predecessor. The Next-Generation HPC Infrastructure Development Division was created to collaborate with vendors and research teams. R-CCS started an $700M AI for Science initiative in FY24 to utilize AI and HPC for solving the most difficult science problems. The initiative will involve the Transformative Research Innovation Platform-Advanced General Intelligence for Science (TRIP-AGIS) Program from FY24 to roughly FY31. TRIP-AGIS will develop generative AI models for scientific research and facilitate the social impact of advanced science. Quantum and HPC machines have been combined in a software, algorithm, and application ecosystem known as the Japan HPC-Quantum Infrastructure Project, with $150M of funding for operations during FY23 to roughly FY27. The project aims to create quantum HPC hybrid software, modular quantum software libraries, and cloud computing technology for a quantum supercomputer hybrid platform. Quantum machines and simulators are being installed to provide testbed capabilities at Riken, the University of Tokyo, and Osaka University. Systems will go online early FY25.

A memorandum of understanding (MoU) was signed between R-CCS and Amazon Web Services in January 2023 to expand the scientific platforms of Fugaku to a cloud computing network. The expansion will virtualize Fugaku and facilitate development of a software stack independent of vendors. Virtual Fugaku version 1 was released in August 2024 and contains Satellite Fugaku, a test environment for Virtual Fugaku for Fugaku users and Private Fugaku, a singularity container for AWS users. Roughly two-thirds of Fugaku users have industry involvement, and the virtual environment provides a secure and private workspace to protect trade secrets. Additional MoUs were signed between DOE and MEXT on HPC and AI, and between ANL and Riken on AI for science.

Thirty percent of R-CCS energy come from renewables, which mitigates electricity price fluctuations. Other methods for reducing the carbon footprint include reusing waste heat and energy-efficient supercomputer operation. Fugaku users also have the capability to select the level of power usage required, with an incentivizing point system favoring lower usages. Overall, measures have increased energy efficiency by 15%.

The *AI For Science Roadmap in Japan* was released on May 31, 2024, and summarizes efforts to drive future AI-for-science research in Japan. The roadmap includes examples, guidelines, and new challenges on the application of cutting-edge technologies, as well as an estimation of the AI computational performance required for the next-generation supercomputer. An update to the 2017 *Computational Science Roadmap* listed the key scientific fields which would experience an expansion of AI applications. The fields are nanoscience devices; energy and resources; elementary particles and nuclei; life sciences; drug discovery and medical care; design and manufacturing; social sciences; brain science and artificial intelligence; tsunamis, weather, and climate; and space and astronomy. AI for science will innovate new methods of generating AI data. Data for training AI models by scraping the internet will be exhausted by approximately 2028, as reported by Villalobos *et al.* Currently, AI models are pre-trained on traditional AI applications data and then tuned on science data. Future methods could have models continually pretrained on scientific data and then tuned on traditional AI applications data.

DISCUSSION

Koniges sought information on geothermal energy applications. **Matsuoka** discussed the existence of various teams researching seismic, energy, and sequestration applications incorporating AI but was unsure of geothermal applications.

Chandrasekaran requested information regarding in situ analysis applications. **Matsuoka** provided an example from the Institute of Weather involving in situ analyses of extreme weather, with the goal of sending high resolution local forecasts to citizen cellphones. Various other examples exist.

Hey questioned whether U.S. supercomputers have undergone anything analogous to Fugaku's assessment by Hyperion Research. **Joseph** informed Hyperion Research recently completed an assessment for NERSC.

NERSC 50 YEAR ANNIVERSARY TALK, Sudip Dosanjh, NERSC

Dosanjh noted NERSC's high energy efficiency has produced three sustainability awards. The next decade will contain many changes, including an expansion of the mission space; incorporation of complex workflows from observational science, AI, and quantum computing; and workforce development. As the mission HPC center for SC, computing time at NERSC is allocated in part by SC program managers. Deployment of new technologies must account for the needs of a wide range of dynamic users, 60% of which are students and early career scientists.

Seven Nobel Prizes have been awarded for research that relied on computations performed at NERSC, and NERSC has contributed to several scientific breakthroughs, resulting in 11,173 acknowledgments in peer reviewed publications and high-profile journals since 2020. Hyperion Research's recent assessment of the center's impact, value and scientific return on investment resulted in the following key findings: the value and impact of research performed at NERSC is among the best of global leadership computing sites; the center facilitates groundbreaking research, impacts the global scientific community, and is a leading force in enabling high-value computational research; the infrastructure is indispensable for addressing societal grand challenges in scientific research; and superior processes and personnel are key to accelerating researchers' time to science. In addition, extensive training programs designate the center as a catalyst for HPC workforce development. Collaborations within the ASCR ecosystem include Leadership Computing Facilities (LCFs), the Energy Sciences Network (ESnet), the High-Performance Data Facility (HPDF) collaboration with the Thomas Jefferson National Accelerator Facility (JLAB) and LBNL, and all ASCR facilities working together on Integrated Research Infrastructure (IRI).

The NERSC workload is diverse, with growing emphasis on integrated research workflows. The center's three advanced capability thrusts include large-scale applications for simulation, modeling and data analysis; complex experimental and AI driven workflows; and time-sensitive and interactive computing. The center's system roadmap contains NERSC-8: Cori from 2016; NERSC-9: Perlmutter from 2020; NERSC-10 from 2026, with the goal of providing an order of magnitude increase in performance over Perlmutter and supporting diverse and complex workflows; and NERSC-11 from 2030+, for science beyond Moore's Law and HPC workflows running seamlessly in an IRI. Each system on the roadmap was designed with increasing emphasis on energy efficiency.

NERSC Exascale Science Applications Program (NESAP) was founded to transition the broad science user base by preparing users for current and future systems. For NERSC-10, the program is focusing on workflows and is in collaboration with strategic science teams and technology vendors. The Superfacility Project at LBNL was designed to address the evolving HPC workload and connects experiment and compute facilities with the expertise and community needed for success. The project currently supports multiple science teams using automated pipelines to analyze data from remote facilities at large scale, without routine human intervention.

The AI strategy consists of deploying optimized hardware and software systems, applying AI for science using cutting-edge techniques, and empowering and developing the workforce through seminars, training and schools, as well as staff, student intern and postdoctoral programs. Since 2017, the center has driven the emergence of AI for science on HPC through research, training, and the founding of the HPC benchmarks in MLPerf. The increasing pervasiveness of AI will transform science and revolutionize operations. The end of Moore's Law, which explained order of magnitude improvements in new systems, facilitates new approaches for computing. NERSC aims to maximize energy efficiency, deploy best-in-class specialized computing elements to accelerate workflows, continue non-recurring engineering (NRE), and co-design efforts to develop representative benchmarks.

The NERSC quantum roadmap outlines the following tasks for 2024-2026: enable access to various quantum hardware, engage with vendors, develop and evaluate quantum and hybrid algorithms, identify opportunities for quantum-accelerated HPC codes, benchmark quantum hardware, and perform resource analysis for executing useful quantum algorithms. During 2026-2029, the roadmap includes the availability of near-term quantum hardware, user requests of both classical and quantum resources, workforce development, and assessment of requirements for quantum hardware onsite. Finally, 2030-2034 events and tasks include availability of high-performance quantum hardware, potential integration with traditional HPC, and users routinely solving problems using both quantum and classical hardware.

DISCUSSION

Koniges inquired how the introduction of a quantum computer would affect NERSC's cooling facility and asked about the threat of earthquakes. **Dosanjh** explained the use of dry coolers will maintain efficiency by reducing water evaporation. The center contains a seismic floor designed to mitigate the damage caused by earthquakes.

Hey requested additional information on the JLAB collaboration. **Dosanjh** noted the HPDF collaboration includes both JLAB and LBNL. **Brown** added the project is active and in its early days. The goal is to achieve critical decision-1 (CD-1) later in 2025.

DOE PUBLIC ACCESS PLAN, Michael Cooke, DOE SC

Cooke recalled the history of DOE Public Access Plans (PAPs). The 2013 Holdren Memo required results of federally-funded scientific research to be made publicly available and allowed a 12-month embargo of peer-reviewed articles. The 2014 PAP required author submission of accepted manuscript for peer-reviewed publications to DOE within one year of publication. The 2022 Nelson Memo required federal science agencies to develop new PAPs and installed the following changes: emphasis on the use and re-use of data and related products;

machine readability; equitable access; immediate public access to publications, removing the 12-month embargo; immediate public access to data displayed in or underlying publications; and expanded use of persistent identifiers (PIDs). The Nelson memo included timelines for implementing publication, data, and PID requirements. Implementation of policies regarding data and publications must begin by the end of calendar year 2025 (CY25).

The 2023 PAP is a product of broad agency engagement and community input. Regarding publications, the PAP stipulates: immediate access upon publication, earlier submission of accepted manuscripts through Energy Link (E-Link), access through Public Access Gateway for Energy and Science (PAGES), and emphasis on author deposits of accepted manuscripts through Green open access (OA). Regarding data, the PAP stipulates: Data Management Plans should become Data Management and Sharing Plans (DMSPs); the type of data being shared was changed to "Scientific Data" to validate and replicate research findings; data should be made available at time of publication; asks for a timeline for sharing other scientific data; and repository selection should align with the National Science and Technology Council (NSTC) Desirable Characteristics of Data Repositories guidance. The three scientific data management principles are to increase the pace of scientific discovery, protect integrity and enhance the value of science, and maximize appropriate data sharing. The White House defined a PID as a digital identifier that is globally unique, persistent, machine resolvable and processable, and has an associated metadata schema. Regarding PIDs, the PAP stipulates the collection of metadata associated with publications and data; metadata should include authors, affiliations and funding with associated PIDs, publication date, and PID for output; researchers must obtain a PID for use when publishing and reporting R&D outputs; and researcher PIDs must meet common or core standards. The PAP also stipulates obtaining PIDS for awards.

DOE Order 241.1C: Scientific and Technical Information Management ensures DOE fulfills its scientific and technical information (STI) management responsibilities and was updated in October 2024. The order requires identification, collection, review, control, preservation, and dissemination of R&D outputs and STI, and establishes requirements and roles for federal staff and contractors. The order is available online.[7]

The Public Reusable Research (PuRe) Data initiative aims to make data publicly available to advance scientific or technical knowledge. Resources are available online.[8] Jupyter, a DOE and National Science Foundation (NSF) funded project, won an award at the White House Office of Science and Technology Policy (OSTP) recognition challenge. .

A new NSTC report *Framework for Considering Data Infrastructure and Interconnectivity* serves as a high-level reference for planning, developing, operating, upgrading, and assessing R&D Infrastructures (RDI). In addition, the report identifies challenge areas faced when developing and upgrading data infrastructure and proposes collective agency and community actions to disseminate practices and approaches for planning data infrastructure and addressing common challenge areas. Recommendation areas for collective action include communication and exchange of practices, coordination, and workforce development.

The SC vision for implementing the PAP involves enabling the development and implementation of DMSPs in partnership with data, computing, and networking experts, integrating PIDs into the scientific workflow, encouraging data citation and reuse, and enabling career-impacting recognition for contributing to open science. Outstanding challenges include

[7] https://www.directives.doe.gov/directives-documents/200-series/0241.1-border-c
[8] https://science.osti.gov/Initiatives/PuRe-Data

DMSP creation without interrupting science and lowering the barriers for participating in science.

DISCUSSION

Hey sought the recommended OA policy for DOE-funded research publication. **Cooke** explained the choice belongs to the authors, but DOE's right to collect accepted manuscripts through E-Link is embedded in the financial assistance agreements.

PUBLIC COMMENTS

Seidel questioned whether DOE would provide facilities for academic researchers to store data linked to publications. **Cooke** noted facility access is under discussion, but infrastructure costs are currently a barrier.

An attendee (online comment read by **Crowder**) asked whether the Office of Management and Budget (OMB) Memorandum M-25-05, "Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Data Access and Management," impacts DOE DMSP. **Cooke** recounted the focus of DMSPs to make publications and data available and connect publications and data through PIDs.

An attendee (online comment read by **Crowder**) sought guidance for visiting professors working at DOE laboratories, but not directly funded by DOE. **Cooke** clarified PAPs relate only to DOE-funded research.

An attendee (online comment read by **Berzins**) sought clarification from Matsuoka on the notion of exhausting the data that is currently available online for training AI by 2028. **Matsuoka** mentioned a link to the referenced publication by Villalobos *et al.* is listed in the presentation slides and readily found online.[9]

An attendee (online comment read by **Berzins**) commented Matsuoka's description of a digital twin workflow differs from the approach used by manufacturers. **Matsuoka** noted many different digital twin workflows exist.

Berzins thanked the members, speakers, and attendees.

Berzins adjourned the meeting at 12:15 p.m.

[9] https://arxiv.org/pdf/2211.04325