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

MEETING MINUTES

September 28th, 2023

VIRTUAL MEETING

ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE

The U.S. Department of Energy (DOE) Advanced Scientific Computing Advisory Committee (ASCAC) convened a virtual meeting on Thursday, September 28, 2023 via Zoom. The meeting was open to the public and conducted in accordance with the requirements of the Federal Advisory Committee Act (FACA). Information about ASCAC and this meeting can be found at http://science.osti.gov/ascr/ascac.

ASCAC members present in virtual attendance

Daniel Reed (Chairperson)
Richard Arthur
Keren Bergman
Martin Berzins
Tina Brower-Thomas
Jacqueline Chen
Silvia Crivelli
John Dolbow
Mark Dean
Roscoe Giles
Susan Gregurick
Bruce Hendrickson
Gilbert Herrera
Anthony Hey
Mary Ann Leung
Vanessa Lopez-Marrero
Jill Mesirov
John Negele
Vivek Sarkar
Edward Seidel
Sameer Shende
Krysta Svore
Valerie Taylor
Cristina Thomas
Theresa Windus

ASCAC members absent

Vinton Cerf
Jack Dongarra
Timothy Germann
Alexandra Landsberg
Satoshi Matsuoka

Also attending virtually

Asmeret Berhe, DOE Office of Science (SC)
Luca Bertagna, Sandia National Laboratories (SNL)
Christine Chalk, ASCAC Designated Federal Officer, Advanced Scientific Computing Research (ASCR)
Xujing Davis, SC Biological and Environmental Research (BER)
Lori Diachin, Lawrence Livermore National Laboratory (LLNL)
Hal Finkel, Acting Research Division Director ASCR
Mike Heroux, SNL
Devinn Lambert, Deputy Director for Crosscuts and Earthshots, Office of the Undersecretary for Science and Energy
Eric K. Lin, Deputy Director for Creating Helpful Incentives to Produce Semiconductors (CHIPS) and Science R&D, U.S. Department of Commerce
Michelle K. Schwalbe, National Academies of Sciences, Engineering, and Medicine
Andrew Siegel, Argonne National Laboratory (ANL)
Ceren Susut, Associate Director of Science, ASCR

There were approximately 275 individuals present for all or part of the meeting.

Thursday, September 28, 2023

OPENING REMARKS, Daniel Reed, ASCAC Chair, convened the meeting at 10:08 a.m. Eastern Time and welcomed attendees.

The possibility of a government shutdown is being monitored.

The Office of Science and Technology Policy (OSTP) is preparing a quadrennial review, which will result in broad technology recommendations. Current changes in the technology landscape have left questions regarding the future of advanced computing. Recent reports highlight similar conclusions and challenges, including the need for a broad national strategy around the future of advanced computing.

VIEW FROM WASHINGTON, Asmeret Asefaw Berhe, Director of the Office of Science

Congratulations to Ceren Susut, now the Associate Director of ASCR. The community's work encompasses important topics that capture the public's attention and present the DOE with the opportunity to be a leader in advanced computing.

The Oak Ridge Leadership Computing Facility (OLCF) will enact a one-year extension of the Summit system. Frontier at the OLCF and Aurora at the Argonne Leadership Computing Facility (ALCF) will have increased research capacity.

There are four finalists for Gordon Bell prizes in high performance computing (HPC); one involves the climate under the exascale computing project (ECP) initiative, which advances the goals of the Biden and Harris administration for climate change research. The others leverage exascale capabilities and artificial intelligence (AI) to accelerate progress in small modular reactors, novel materials, and human immunodeficiency virus (HIV) research.

The budget is being closely monitored. Work is in progress to secure funds that will meet current expectations and formulate plans for the continued operation of the DOE Office of Science (SC) in the case of a government shutdown. Additional efforts are focused on supporting and expanding the computing workforce of the future. Concerns for the ECP workforce and software ecosystem are noted. Securing resources for SC in general will translate to support for all these areas.

A new science era is beginning which combines the power of exascale computing with the potential of AI. Data trained on limited segments of society can be detrimental. Therefore, the recruitment of a diverse workforce must be a priority.

DISCUSSION

Taylor asked for comments on the future of microelectronics and the application of AI under the National AI Research Resource (NAIRR) Taskforce. **Berhe** mentioned the creation of the new microelectronic centers, a general expansion of efforts, and the continuation of current co-design activities. The magnitude of all efforts will depend on the budget. DOE and ASCR have been active in NAIRR conversations, but questions remain regarding the governing structure and resource allocation.

ASCR RESEARCH PRIORITIES, Hal Finkel, Acting Research Division Director, Advanced Scientific Computing Research

Five critical technology trends motivate ASCR, and several have spurred workshops, roundtables, and funding opportunity announcements (FOAs) in recent years that build on ECP investments.

The first trend is Heterogeneous, Distributed, Co-Designed, and Energy-Efficient Computing. Related FY23 FOAs supported Accelerate Innovations in Emerging Technologies (Accelerate); three Exploratory Research for Extreme-Scale Science (EXPRESS) initiatives (i.e., Modeling Future Supercomputing Systems for EXPRESS, Programming Techniques for Computational Physical Systems in EXPRESS, and Quantum Algorithms across Models in EXPRESS); Quantum Testbed Pathfinder; and Scientific Enablers of Scalable Quantum Communications. The ASCR Basic Research Needs (BRN) Workshop in Quantum Computing and Networking and the Quantum Computing for Biomedical Computational and Data Sciences Joint DOE-National Institutes of Health (NIH) Quantum Roundtable were hosted in FY23.

The second and third trends are Data, Privacy, and Scientific Integrity and Exploding Software Complexity, respectively. While these trends generated several FY21 and FY22 FOAs, there were none in FY23. However, software complexity was addressed in an FY23 DOE/National Science Foundation (NSF) Workshop on Correctness in Scientific Computing.

The fourth trend is Scientific Computing and Networking: from Exascale to the Edge. Related FY23 FOAs supported the Energy Earthshots Research Centers (EERCs); Biopreparedness Research Virtual Environment (BraVE); Accelerate; Advanced Scientific Computing Research for DOE User Facilities; Scientific Discovery Through Advanced Computing (SciDAC) in partnership with DOE Fusion Energy Sciences (DOE FES); and Distributed Resilient Systems. A report, titled *Integrated Research Infrastructure Architecture Blueprint Activity* was completed in FY23.

The fifth and final trend is Creating Trustworthy and Efficient Artificial Intelligence for Science, which generated an FY23 FOA titled Scientific Machine Learning for Complex Systems.

Several FY23 solicitations were issued in FY23 to grow and diversify the ASCR community, including the Early Career Research Program (ECRP); Funding for Accelerated, Inclusive Research (FAIR); and Reaching a New Energy Sciences Workforce (RENEW). Many Phase I and II Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) grants were also awarded in FY23 for building capability in the ecosystem and transferring technology.

Anticipated FY24 solicitations are expected to award funds to a smaller number of larger, broadly scoped projects and seek opportunities to expand the number of institutions in ASCR's portfolio. There is a new location for FOAs on the ASCR website.

Science highlights featured six projects addressing the ASCR priority of trustworthy and efficient AI: 1. Scalable Transformers on Frontier for Real-Time Experiment Steering; 2. SuperNeuro: an Accelerated Neuromorphic Computing Simulator; 3. Privacy-Preserving Federated Learning as a Service using Argonne Privacy-Preserving Federated Learning framework (APPFL); 4. Correctness of Automatic Differentiation (Autodiff) on Machine-Representable Inputs; 5. Symbolic Representation for Neural Networks (SyReNN): A Tool for Analyzing Deep Neural Networks (DNNs); 6. Dehallucination of Large Language Models (LLMs) for High-Level Planning.

ASCR welcomes a new staff member on part-time detail from Argonne National Laboratory (ANL).

DISCUSSION

Herrera asked for clarification of the term "quantum at the edge". **Finkel** explained the portfolio looks at quantum computational properties in edge quantum systems. Starting with quantum data from quantum sensors, a determination is made on how to embed quantum computation directly into the process.

Gregurick mentioned the NIH is interested in quantum computing and asked for additional comments on the topic. **Finkel** had nothing to add.

Sarkar asked how research ideas are selected for transition after the research is done. **Finkel** explained the process is most often bottom-up, as people adopt the specific technology that suites their needs. Opportunities to enable and facilitate the process exist through small business awards and seeds and collaborations around software stewardship and technology.

Taylor questioned whether the description of FY24 solicitations being broadly scoped applies to multiple projects within ASCR or multiple groups across SC. **Finkel** explained the scope refers to ASCR's purview to have an inclusive posture across its portfolio.

Windus asked for more details on SciDAC. **Finkel** replied several calls were issued for SciDAC partnerships with different offices over the past few years, and all portfolios are on schedule.

Seidel expressed concern that the rush to AI has left traditional simulation and modeling activities without support. **Finkel** explained funding requires a balanced strategy. Many of the core areas invested in apply to both AI and other areas. This is the rationale for having broadly scoped solicitations. However, authorizers and appropriators have expressed interest in the heavy investment in AI.

Reed asked about DOE's strategy for prioritizing programs in the current period of transition and uncertainty. **Finkel** replied portfolios are structured with a spectrum of risk and timescale. DOE looks at both high-risk and high-reward transformative opportunities, and innovative work to advance core techniques that the research enterprise depends on.

ENERGY EARTHSHOTS INITIATIVE, Devinn Lambert, Deputy Director for Crosscuts and Earthshots, Office of the Undersecretary for Science and Energy

Created by President Biden two years ago, the Energy Earthshots Initiative targets the research, development and deployment (RD&D) breakthroughs required to solve the climate crisis and reach the 2050 net-zero carbon goal. Each of the seven Energy Earthshots announced thus far targets decadal performance in a key area: 1. Floating Offshore Wind (reduce the cost of floating offshore wind in deep waters to $45 per megawatt-hour); 2. Enhanced Geothermal (reduce the cost of systems to $45 per megawatt-hour); 3. Long Duration Storage (reduce the cost of grid-scale energy storage by 90% for systems that deliver 10+ hours of storage); 4. Industrial Heat (develop cost-competitive industrial heat decarbonization technologies with at least 85% lower greenhouse gas (GHG) emissions); 5. Hydrogen (reduce the cost to $1 per kilogram); 6. Clean Fuels and Products (develop cost-effective fuels and products from sustainable carbon sources to achieve 85% lower net GHG emissions); and 7. Carbon Negative (charge less than $100 per metric ton of stored CO_2 equivalent). Meeting these ambitious goals is challenging but achievable and will require an all-hands-on deck approach.

DOE resources, planning, and stakeholders form a feedback loop to guide completion of the targets. ASCR's participation in the Earthshots is evident through existing projects and research, FY24 budget requests, and involvement in SC-wide Earthshot activities. Two such activities are the EERC and Science Foundations for Energy Earthshots FOAs.

Three exemplary Earthshots projects are leveraging exascale computing, AI, and machine learning (ML) to make research safer, faster, and more economical. The first project in Enhanced Geothermal from Lawrence Berkeley National Laboratory (LBNL) focuses on overcoming well bore failures by increasing the accuracy of fracture evolution predictions through microscale resolution. The second project in Floating Offshore Wind, led by the PI for ECP's Exa-Wind, established new software to simulate a two-megawatt turbine with an 80-meter rotor, allowing the comparison of simulated data to field measurements. This both drives innovation and quantifies risk for project developers. The final project is in Clean Fuels and Products; the startup firm Pyran scaled their process for creating biomass-based 1,5-pentanediol by 1000X in a single step. This achievement was made possible with computational modeling and simulation provided by Oak Ridge National Laboratory (ORNL) and a collaboration with the Consortium for Computational Physics and Chemistry (CCPC).

DISCUSSION

Reed asked how the combination of shifting computing and a focus on renewables will affect the completion of Earthshot objectives. **Lambert** replied increasing efficiency in computing and grid modernization are critical spaces that require innovations. There must also be a willingness to purchase clean power at a higher price to support communities engaged in the early deployment of clean fuels.

OPPORTUNITIES AND CHALLENGES FOR DIGITAL TWINS, Michelle K. Schwalbe, National Academies of Sciences, Engineering, and Medicine

The National Academies of Sciences, Engineering, and Medicine (NAS) appointed an ad hoc committee to define digital twins and identify the needs and opportunities required to advance the mathematical, statistical, and computational foundations of digital twins in applications across science, medicine, engineering, and society. The committee has organized three large public workshops to address these tasks. Results will be summarized in a consensus report to be released this fall. The study was funded by DOE, NIH, NSF, and the Department of Defense (DOD). Within NAS, several divisions and boards are collaboratively working on this study. Though currently under debate, the definition of digital can be thought of as the coupling of a digital model with a physical asset, such that dynamic updates to the model occur with data from the physical twin throughout its lifecycle, thereby informing decisions and realizing values.

The Biomedical Sciences workshop identified the following future opportunities: developing interactions between patients and digital twins to inform clinical decisions; and conducting virtual trials. Challenges included: a need for improvements in modeling, computation, and data; managing data heterogeneity and disparate multimodal data; accounting for and communicating uncertainty; developing trust with patients and clinicians; addressing ethical privacy and security concerns; and improving data sharing mechanisms.

The Atmospheric, Climate, and Sustainability Sciences workshop identified opportunities in: interactive Earth simulations to inform policy decision; community engagement; and higher resolution models. Challenges included: the need for increased observational, computational, and resource management abilities; expanded research on parametric sparsity and generalizing observational data; increasing the generation of training data and computation for the highest possible resolution; and improving uncertainty quantification and calibration.

The Engineering Domains workshop identified opportunities to provide information for asset management; enhance asset reliability and maintenance; and test, design, and apply prototype processes. Challenges included connecting the simulations across length scales and physical phenomena, advancing experimental design as well as sensor placement and scheduling; capturing sources of variation; managing uncertainty propagation and quantification; integrating data science and domain knowledge to enable decision-making; managing data and applying advanced analytics; and developing ontologies and harmonization among the digital twin user community.

DISCUSSION

Arthur asked for thoughts on a national infrastructure for hosting the execution of non-private digital twins and the corresponding data. **Schwalbe** cannot comment on details until the report is released. The report will describe such opportunities, and there is a lot of interest in this area for digital twins.

Mesirov asked for clarification on the difference between digital twins and traditional modeling and simulation. Cross-cutting modeling issues are very domain specific. **Schwalbe** explained the difference lies in the feedback cycle. The digital twin continually co-evolves with an individual physical twin, such as a single airplane. In contrast, updates in traditional modeling occur in discreet sessions and apply to many entities, such as all airplanes of a given model. There was agreement regarding the cross-cutting issues, and a comment that these issues were a key challenge tackled by the study.

Berzins asked if it is a fair assessment that digital twins respond to feedback loops more quickly than massive simulations. **Schwalbe** responded the timeframe is domain specific. For example, atmospheric modeling would be expected to take longer than modeling a single patient.

Reed commented the real-time feedback of digital twins is the critical enabler.

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

The FY24 President's Budget Request (PBR) of ~$1.13B represents an ~5.4% increase over the FY23 Enacted Budget. The FY24 House and Senate Marks are both ~$1.02B.

The FY24 House Mark for HPC and network facilities recommends not less than $219M for the ALCF, $255M for OLCF, $135M for the National Energy Research Scientific Computing Center (NERSC) at LBNL, ~$90.2M to support necessary infrastructure upgrades and operations for the Energy Sciences Network (ESnet), and $7M for the High Performance Data Facility (HPDF). Not less than $295M was recommended for Mathematical, Computational, and Computer Sciences Research, while not less than $245M was recommended for quantum information science (QIS). Additional House guidance seeks expansion of the quantum internet, networking, and communications testbeds, $15M for the Quantum User Expansion for Science and Technology (QUEST) program, a report of near-term application developments in quantum computing technologies, and the expansion of work with NIH. No further funding for RENEW or FAIR are included, although not less than $35M for the Established Program to Stimulate Competitive Research (EPSCoR) across SC programs is recommended. Up to $5M from ASCR is requested for the Energy Earthshots. Finally, the House requested a table of user facility funding levels from the previous five fiscal years showing optimal operations using the new determination as well as information on all funding that is contingent on the future availability of funds.

The FY24 Senate Mark for HPC and network facilities recommends not less than $219M for the ALCF, ~$248M for OLCF, $135M for NERSC, and ~$91M for ESnet. The Senate supports The Committee supports the PBR for the HPDF's continued planning and design. Not less than $280M was recommended for Mathematical, Computational, and Computer Sciences Research, including $20M for the Computational Science Graduate Fellowship (CSGF). The

Senate advised not less than $255M for QIS and $135M for AI/ ML. The Senate supports RENEW, FAIR, and recommends $35M for EPSCoR. Among other activities, the Senate encourages ASCR to support microelectronics research and the new Microelectronics Science Research Centers, advises issuing up to $18M for the Energy Earthshots, and recommends supporting user facilities as a top priority.

The full Perlmutter system was accepted in July 2023. Graphics processing unit (GPU) nodes have been available to NERSC users since May 2022 and have seen ~ 10,000 annual users from ~800 institutions and national labs. Cori was retired in May 2023. Aurora is on track for deployment. Frontier is open for science, and all allocation programs have been enabled since April 2023. Frontier has over 1,500 users, and 240 active projects, including three Gordon Bell Prize finalists traversing the topical areas of new alloy discovery, advanced reactor design, and atmospheric modeling. Summit has been stable and is being extended for an additional year under the name Summit Plus which is currently accepting proposals.

Instead of focusing on the end of ECP, a celebration is warranted for the beginning of a new science era brought on by exascale computing. Three themes are in focus: 1. Leveraging the exascale ecosystem, especially in combination with AI; 2. Leveraging heterogeneous architectures; and 3. Finding a balance between agility and long-term sustained funding. The latter will involve investments through the EXPRESS program and through bigger research teams and center-scale activities that involve multiple institutions and more sustained, larger funding.

DISCUSSION

Berzins asked about strategies for retaining traditional advantages in photolithography, energy saving architecture, and hypersonic science which do not seem to be areas of priority, although the U.S. is falling behind in these areas relative to other countries and may require interagency collaborations. **Susut** explained this is being addressed through efforts in microelectronics and the already active interagency partnerships in advancing the field. EXPRESS also gives the flexibility to pursue these topics. **Berzins** commented these are good steps, but there is an urgent need to consider what the machines will be after exascale, so the architectural advantages gained are not lost over the next five years. **Susut** agreed.

Reed shares in Berzins' concerns and noted maintaining the U.S.'s traditional advantages is not just an SC or DOE challenge, but a trillion-dollar national strategy question. **Susut** replied ASCR is well prepared to work with industry partners. ECP not only created an ecosystem but also granted experience in advancing different fields and working with industry partners. This experience must be leveraged while acknowledging budget realities.

UPDATE FROM THE EXASCALE COMPUTING PROJECT – HOME STRETCH
Lori Diachin, Lawrence Livermore National Laboratory

ECP has reached threshold in the key performance parameters (KPPs). ECP anticipates the following completion rates: for KPP-1, nine to 10 of 11 applications will complete their base challenge problem with performance exceeding 50X; for KPP-2, seven to eight of 10 SC applications, and three of four National Nuclear Security Administration (NNSA) applications will meet their base challenge problem; KPP-3 will likely see an integration score of 55+ points out of 68; and KPP-4 has already delivered all 267 PathForward milestones.

Three months of technical work remain, and the ECP leadership team is increasingly focused on project closeout, communications, outreach, and training the future DOE HPC workforce. The ECP's legacy, which is being documented for a wide audience, includes a suite of applications that will impact problems of national importance for future decades, a widely available integrated software stack for GPU-accelerated computing , best practices and lessons learned for thinking about how to program GPUs, over 1,000 researchers trained and ready for accelerator-based computing, and best practices for running a large-scale software development RD&D 413.3b project. ASCR HPC facilities were integral to meeting KPPs. All ECP teams have run on Frontier, using ~5M node hours. Six ECP teams have Innovative and Novel Computational Impact on Theory and Experiment (INCITE) allocations, using ~3.6M node hours. Sixty-one ECP teams (186 users) have run on Sunspot, the Aurora test and development system (TDS), using ~ 9K node hours, and 12 ECP teams (14 users) have run on Aurora, for ~24K node hours.

A highlighted effort from the hardware and integration team involved implementing AMD's HIP and NVIDIA's CUDA programming models on Aurora's Intel GPUs. This simplifies the movement of applications to Aurora and supports future work with AMD to push HIP as an open standard. A highlighted Gordon Bell finalist project, ExaSMR, delivers experiment-quality simulations, advancing DOE's goal of building an operational small modular reactor (SMR) in 10 years, which will help develop safe, clean, and affordable nuclear power options.

ECP-funded technologies and participants have won 13 R&D 100 Awards since 2016 and many other notable awards and distinctions. The industry and agency council (IAC) has been active with several events including quarterly calls, and technical meetings and collaborations with the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA). IAC will hold two final meetings in the fall of 2023 and winter of 2024, and a virtual workshop in October 2023 to share best practices and lessons learned in applying the principles of continuous integration/ continuous delivery to HPC software development and change management.

ECP has a significant focus on training the next generation to create a pipeline for the DOE HPC workforce. ECP fostered sustainable research pathways for HPC by funding faculty and students across multiple labs, and by holding workshops and bootcamps.

ECP is moving toward project closeout activities, with the end of technical work anticipated by January 2024 and a project completion date of April 2024. Other closeout activities include: 1. Communication and outreach; 2. Project closeout and archive; and 3. Review preparation and completion of the Critical Decision (CD)-4 items (delivering KPPs, open-source publishing of ECP software and technologies, and working with vendors and other partners to transfer ECP technologies).

ECP has increased staff resources to inform the public of the project's accomplishments and lessons learned on technical, collaboration, and project management fronts. The communication strategy includes multiple publications, case studies, stakeholder briefings and a significant presence at the SuperComputing 2023 (SC23) conference. The community has realized the impact of the ECP to deliver the promise of next-generation computing capabilities, to enable researchers to move their science forward more quickly, and to provide a breakthrough resource for discovery and advancement.

ECP Application Development Update, Andrew Siegel, Director of Application Development, ECP, Senior Scientist, Argonne National Laboratory

KPP 1 aimed for 11 selected applications to demonstrate performance improvement on mission-critical problems, and KPP 2 aimed for 14 selected applications to broaden the reach of exascale science and mission capability. Defining and refining the KPP metrics was a significant effort that spanned three generations: 1. Creating general definitions; 2. Developing individual figures of merit (FOMs); and 3. Writing formal KPP verification contracts. Overall, a balance was needed to allow for the freedom and creativity to spur innovation while retaining quantitative metrics that indicated efficient use of the resource and significant advances.

The KPP verification workflow entails completion of the challenge problem, completion of the KPP verification report, an external review and principal investigator response period, signature of the external reviewer, and finally signature of the federal project director (FPD). The time from completion of a KPP challenge problem to the collection of a Federal Project Director's (FPD) signature, which formally denotes success, can take up to 12 weeks.

Of the 11 KPP-1 projects, two are still in progress, two are under review, two have reached the external reviewer signature stage, and five have been signed by the FPD. Of the 12 KPP-2 projects, four are in progress, three are under review, and three have been signed by the FPD. During the ECP, languages and programming models used to meet KPPs evolved; many applications began by using native GPU and loop pragma models before moving to C++ abstractions.

Three key lessons that characterized the ECP for KPPs 1 and 2 were 1. A holistic rethinking and restructuring of implementation, algorithm, and models; 2. Recognizing and utilizing the co-design loop between software and applications; and 3. The necessity of leveraging community tools and diverse expertise.

ECP Software Technology Update, Michael A. Heroux, Director of Software Technology, ECP, Senior Scientist, Sandia National Laboratories

ECP software technology works on products that applications need now and in the future, resulting in portable libraries and tools for accelerators. Software can be divided into three categories: 1. Next-generation established products that include widely used HPC products; 2. Robust products that address key new requirements; and 3. New products that enable exploration of emerging HPC requirements. Science highlights featured the R&D 100 Award-winning open-source library zfp and the Variorum application. Overall ECP has contributed to the exascale readiness of 70 libraries and tools.

The Extreme-scale Scientific Software Stack (E4S) is a community effort to provide open-source software packages for developing, deploying, and running scientific applications on HPC platforms. The E4S ecosystem has resulted in partnerships with other U.S. agencies, industry, and other key stakeholders. The latest release in August 2023 contained 115 directly supported products. Two highlights of the release are as follows: 1. Business relations have granted access to multiple vendor specific programming environments, which ultimately increase product stability, and 2. E4S also includes the most popular AI/ ML frameworks and industry standard libraries and tools to improve ease of utilization.

ECP's legacy includes: 1. A collection of portable GPU-enabled libraries and tools ready for today's platforms and adaptable to future systems; 2. A de-risked path for other communities to replicate ECP's successes by using the resulting libraries and tools, and leveraging the lessons learned; 3. A foundation for a 100X improvement across many scientific domains; and 4. A key to addressing HPC environmental impact concerns by migrating from CPU to GPU-based HPC systems.

DISCUSSION

Berzins asked whether the team encountered more impediments than expected on Aurora. Is there any perception the people who were involved in the project feel they have a good future, and has that perception been quantified? **Diachin** recalled seeing similar problems with Frontier and believes problems were as expected. While the long-term future of those involved in ECP is outside DOE's control, ECP has been diligently monitoring opportunities, encouraging participants to pursue these opportunities, and tracking success. **Heroux** added the fate of and opportunities for all program teams are being carefully tracked.

Hendrickson commented an additional legacy of ECP is the community-developed software stacks that will be the basis for the next 10+ years of developments in the field and have favorably positioned both the technology developers and application scientists.

Seidel requested more commentary on the outgoing staff's future. **Diachin** replied things look good for FY24. Most staff have found new positions, although not necessarily in their key research area. However, staff fate in FY25 is less certain and hard to predict. **Heroux** added DOE is putting a lot of energy into post-ECP efforts. The decreased budget is forcing innovation, deeper partnerships, and deeper levels of trust that the investments made in ECP will translate to further opportunities. **Siegal** added it is hardest to track staff who leave voluntarily, because they are not where they think they can make an impact. There is no natural follow-on to what these staff have been doing for seven years and in which they have deep expertise.

Bergman asked if more could be done to communicate ECP's accomplishments externally. **Diachin** explained such communication is currently an area of focus. There are currently six to seven writers working with leadership to craft stories suitable for lay audiences.

Giles lamented there were no opportunities to create a smoother path for transition post-ECP and asked for comments on the project's international impact. **Heroux** explained there are multiple international collaborations moving forward, and people outside the U.S. know about ECP and see the value in forming partnerships.

Reed commented that everyone wants to ensure what ECP created — including people, the sphere of collaboration, and resources required for the maintenance of accomplishments — is preserved and not lost in this transition. The community is united in gratefulness for what was accomplished and in concern for the team.

REPORT FROM THE SUBCOMMITTEE REVIEW OF THE DOE- NCI COLLABORATION, Tony Hey, ASCAC

ASCAC provides guidance to the activities of Joint Design of Advanced Computing Solutions for Cancer (JDACS4C), a collaboration working towards predictive oncology. Three DOE-National Cancer Institute (DOE-NCI) projects and a cross-cutting ECP project in support of HPC for cancer analytics took place at Frederick National Laboratory in June 2023. Reviewed DOE-NCI projects include: 1. Modeling Outcomes using Surveillance data and Scalable AI for Cancer (MOSSAIC); 2. AI-Driven Multi-scale Investigation of RAS/RAF Activation Lifecycle (ADMIRRAL); and 3. Innovative Methodologies and New Data for Predictive Oncology Model Evaluation (IMPROVE). The ECP project is titled the CANcer Distributed Learning Environment (CANDLE).

The committee found MOSSAIC to be the most mature and impactful of the three projects. The project aims to deliver advanced computational and informatics solutions needed to support a comprehensive, scalable, and cost-effective national cancer surveillance program and lay the foundation for an integrative data-driven approach to modeling cancer outcomes at scale and in real time. MOSSAIC's technology is already being used by 16 Surveillance, Epidemiology, and End Results (SEER) program sites and the Veteran Administration's (VA) registry. The project's exploration of the use of foundation models with SEER data was commended. ADMIRRAL ambitiously uses the Multiscale Machine-learned Modeling Infrastructure (MuMMI) computational model to study the RAS/RAF protein complex. Significant progress has been made, with more powerful AI-enhanced modeling capabilities introduced in conjunction with impressive experimental validation of model predictions. IMPROVE, only its second year, has already made significant progress towards creating a community-based framework for cross-comparison of AI models used to validate cancer drug response models. The codebase and documentation are available on GitHub with data and curated models.

CANDLE aims to deliver a viable exascale-optimized software framework for deep learning applied to cancer and other potential drug-discovery scenarios. CANDLE provides deep learning benchmarks for all three DOE-NCI projects. Additionally, this ECP project delivered all milestones on time, exceeded the KPP benchmarking performance improvement metric by a figure of 5x, and is used for 30+ cancer deep learning models. Furthermore, CANDLE demonstrated the potential to transfer technology to other domains during the COVID-19 pandemic; the paper *GenSLMs (Genome-Scale Language Models) Reveal SARS-CoV-2 Evolutionary Dynamics* was awarded a Gordon Bell special prize at SC22. The project also won an R&D 100 Award. A stretch goal is the development of a DOE Transformer-based modeling framework and has resulted in collaboration with the ExaLearn ECP project to investigate LLMs. LLM foundation models trained on scientific datasets were run on Frontier and Aurora protypes before the ChatGPT revolution. Teams are working towards an international Transition Processing Council (TPC) consortium to develop a 1-trillion parameter model using the NVIDIA Megatron and Microsoft DeepSpeed codebases for training.

DISCUSSION AND VOTE ON REPORT

Mesirov expressed concern over the lack of communication, outreach, and marketing to the NCI community and asked how this could be improved. **Chalk** explained the issue has been discussed between DOE and NCI. Communication is more of an NCI challenge as it pertains to their community, but DOE will share this feedback.

Berzins commented CANDLE is a mechanism for interacting with local cancer centers and has been downloaded many times. A formal tracking system to identify where and how CANDLE is being used would be beneficial.

All ASCAC members present voted to approve the report.

ENERGY EXASCALE EARTH SYSTEM MODEL (E3SM) PROJECT UPDATE

Xujing Davis, Office of Biological and Environmental Research

The Energy Exascale Earth System Model (E3SM) project aims to make actionable predictions of the changing Earth system. E3SM has been running for ~10 years and is currently in phase three of four. Unique capabilities are as follows: 1. Exascale Readiness: E3SM developed the first benchmark of its kind, by running ~3-km global simulation, Simple Cloud-

Resolving E3SM Atmosphere Model (SCREAM), on Frontier with the record setting performance of simulating a global year of climate in a day; 2. Regional-Refined Model (RRM): E3SM is the first Earth System Model (ESM) running fully coupled global simulations with a RRM in most components and completed climate production simulations; and 3. Coupled Earth-Human Feedback: E3SM is coupled with the Global Change Assessment Model (GCAM). The project is also a finalist for the 2023 Gordon Bell Prize.

Phase three, ending in 2025, focuses on GPU readiness. Some codes still run on CPUs, but new GPU capabilities include 1. cloud-resolving simulations via SCREAM v1; 2. a Multiscale Modeling Framework (MMF) that offers coupled climate simulations; 3. high-resolution land simulations; and 4. a Model for Prediction Across Scales (MPAS)-Albany Land Ice Model (MALI). Phase four, scheduled for 2026-2028, aims for the full ESM to run efficiently on GPUs.

E3SM has benefitted significantly from SciDAC and ECP. SciDAC contributions, for example, improved E3SM ice sheet modeling, performance and algorithms, and physics and accuracy. ECP contributed the E3SM-MMF, the model's GPU-enabled configuration, which allows for 100+ year-long simulations and made the model exascale ready. The traditional E3SM configuration also benefited from ECP through code portability and performance improvements.

The long-term goal for E3SM is to assert and maintain an international scientific leadership position in the development of ESMs while addressing DOE's mission. The project's current opportunities and challenges involve improving the model's capabilities, fully transitioning to GPUs, and ensuring teams are thriving. Within DOE, future efforts will focus on strengthening the core through the Office of Biological and Environmental Research (BER) as well as SciDAC partnerships while enhancing integration with other DOE offices and programs. E3SM will also seek to coordinate and collaborate across agencies while working with other science community initiatives.

E3SM C++ Atmosphere Model: Our Journey to Exascale, Luca Bertagna, Sandia National Laboratories

E3SM has several components (such as atmosphere, land, ocean, land ice, sea ice, and others) that can run at different resolutions and a variety of time/ space scales. The E3SM Atmosphere Model (EAM) was originally written in Fortran and highly performant on CPU-based HPC clusters. ECP advances in HPC architecture made exascale simulation of EAM dynamics and physics possible; however, code portability became necessary. To achieve portability, a general-purpose library, in this case the C++ library Kokkos, was chosen to handle architecture specific choices. The resulting portable code is referred to as EAMxx and was designed to expose parallelism, maximize vectorization, and minimize memory movement. Resulting kilometer-scale resolutions allow for predictions with reduced uncertainty and capture the extreme weather associated with climate change.

The transition from EAM to EAMxx entailed three phases which started with Climate Modeling Development and Validation (CMDV), continued with non-hydrostatic atmosphere dycore and physics porting, and ended with EAMxx having the full atmosphere component running end-to-end on the device. Running the full GPU system resulted in an ~6X speedup relative to CPU performance and consumed 1.7X more power, making it more efficient and 3.5X faster per watt. Several components of EAMxx were developed by computational scientists funded by ASCR in collaboration with BER.

EAMxx allows for analysis of historical weather events, detection and removal of model biases, and long-term simulations with various sea surface temperatures for climate studies. The model will serve as a reference point for all subsequent atmosphere code's exascale performance.

DISCUSSION

Reed asked if the system is still a long way from being able to do large-scale parametric studies, which will require even more performance. Performance improvements may also come from next-generation architectures. **Bertagna** confirmed and explained the system may be able to achieve a small increase in performance but not an order of magnitude, so the task will likely result from algorithmic improvements. **Davis** agreed and added that by 2025, the system will be able to run a global, three-kilometer resolution, decade-long simulation.

Spotz (via chat) asked about the status and strategy for integrating all the sub-grid physics into a C++ model. **Bertanga** explained parameterization for the sub-grid is already in C++. Other parametrization is required to run at lower resolutions, and that is on the to-do list.

CHIPS RESEARCH AND DEVELOPMENT OFFICE: A SUMMARY AND UPDATE

Eric K. Lin, Deputy Director for CHIPS R&D, U.S. Department of Commerce

The goals of the CHIPS R&D program are to promote U.S. technology leadership, accelerate ideas to the market, and generate a robust workforce. The CHIPS R&D program comprises four programs under development: the National Semiconductor Technology Center (NTSC), the National Advanced Packaging Manufacturing Program (NAPMP), up to three Manufacturing USA institutes, and the Metrology R&D Program overseen by the National Institute of Standards and Technology (NIST). These programs will operate in partnership with academia, industry, and government with guidance from the Industrial Advisory Committee.

The NTSC is the flagship program meant to be a public-private consortium established and run by a nonprofit operator and aimed to be viewed throughout the world as an essential resource with a network of respected scientists and engineers, state-of-the-art facilities, effective programs, and demonstrated technical achievements. The NSTC will include three programs, each focusing on a key area of development. These areas include: 1. Technology leadership – seeking community input into the grand challenges across multiple areas; 2. Community assets – physical infrastructure such as prototyping facilities and necessary tools; and 3. The workforce – the NTSC is envisioned to be a community convener across the whole range of semiconductor technology engaging in outreach, support, and training.

The NAPMP aims to use R&D tools to build a domestic ecosystem in advanced packaging. The program will be structured like a traditional funding program and will leverage public-private partnerships. As a space where performance improvements could be made within semiconductor technology, the program will implement a broad range of technologies such as heterogeneous integration, wafer and panel-based approaches, tooling and automation, and substrate technology. A key goal is to increase the level of advanced packaging done inside the U.S., which is currently under 5%. This will be approached through technology innovation and ecosystem support. The NAPMP will utilize the NSTC to support pilot packaging facilities.

The metrology program is foundational to all R&D programming and is centered around the grand challenges in the measurement science space. Goals include expanding measurement solutions for the semiconductor ecosystem, harnessing the diversity of people and ideas both inside and outside of NIST, and expanding education and workforce development opportunities. Industry input will be sought for addressing the measurement science involved in new materials and packaging, physical metrology for next-generation microelectronics, computation and data, virtualization and automation, reference materials, calibrations and standards for processes, cybersecurity, and test methods.

The Manufacturing USA program encompasses a range of advanced manufacturing areas, and the current network includes 16 institutes and nine partner federal agencies. A request for information (RFI) was issued to the community to guide investments and design of up to three new institutes. Areas of RFI consensus include the need for larger scales to impact the semiconductor space, the need for connection to other CHIPS initiatives and institutes, and the need to ensure sustainability through long-term funding. No consensus was reached on the best topic for the new institutes. Topic suggestions for cross-cutting technologies included productivity enhancement, smart manufacturing and automation, new and advanced materials, and metrology and testing. Topic suggestions for focused technologies included substrate manufacturing for advanced packaging, sensors and microelectromechanical systems, and infrastructure to support technology transition to manufacturing.

The CHIPS Research and Development Office's Standards Summit will be held as an in-person and virtual event on September 26 and 27, 2023. The Summit will bring together CHIPS for America leaders, standards setting organizations, and industry alliances, domestic and abroad, to identify community priorities for semiconductor and microelectronics standards activities. The Summit aims to foster collaboration, coordination, and innovation within the semiconductor industry's standards community.

DISCUSSION

Berzins asked how CHIPS fits into the photolithography space, a key area in which the U.S. is falling behind. **Lin** explained the business model market drivers for how technology is adopted must be understood before the space can be entered. However, there is a lot of room for improvements across that process. The range of places to make a difference, either by margin or to improve yield or to other areas, is open.

Dean asked about investments in AI technology. **Lin** explained the semiconductor industry is interested in incorporating AI into many areas, such as yield and process optimization, and the technology will be utilized across all CHIPS programs.

Bergman asked if there is opportunity to have testbed-type systems. **Lin** replied there is certainly room and space for testbed systems.

Reed asked how to address the fact not all the country's needs will be solved by economic forces and will need government help to be driven forward. **Lin** replied there is space for DOE in the NTSC to have such conversations. However, the solution to justify the required investments is to present these needs as being important for national or economic security, either due to enhanced capacity or capability.

Reed posed a question to the committee: How can ASCAC shape the direction taken by the community in this time of uncertainty to support national needs, infrastructure, and the associated restrictions? **Giles** responded the answer is not easy, given the failure of the committee's past suggestions to make it up far enough up the chain to make an impact, such as the ECP transition report. Improving that pathway or identifying alternate pathways may help. **Reed** asked for Susut's perspective. **Susut** responded actionable items in terms of a timeline would be helpful and should be considered as a community. **Berzins** commented people respond best to threats and things are looking good now, but in two to three years the U.S. may fall behind China's technology. The current levels of complacency may be due to a general lack of awareness to the challenges and threats to the U.S.'s current position. **Svore** added there is both a risk and an opportunity for this group to identify and shape the "ECP 2.0" - an integration of exascale, HPC, AI, and quantum computing. **Reed** added food for thought: AI, government, and academia all think and move on different timescales. How can the committee move thoughtfully, but with enough alacrity to be able to embrace the opportunities and the challenges as they evolve?

PUBLIC COMMENT

None.

Reed adjourned the meeting at 4:03 p.m.

Respectfully submitted on November 2, 2023,
by Patrick Cosme, Ph.D.,
Science Writer, Oak Ridge Institute for Science and Education.