ASCR/ECP Transition Report

A report by the Advanced Scientific Computing Advisory Committee

Subcommittee members:

Dr. Jay Bardhan
Director
Outreach and Strategic Partnerships
Scientific Computing
GlaxoSmithKline

Dr. Alan Edelman
Professor, Applied Mathematics,
Massachusetts Institute of Technology

Dr. Fred Johnson
Computer Science Program Manager (Retired)
Advanced Scientific Computing Research
Office of Science
U.S. Department of Energy

Alexandra Landsberg (ASCAC)
Division Director, Mathematics, Computer and Information Sciences
Office of Naval Research

Dr. Richard Lethin (ASCAC)
President, Reservoir Labs
Senior Lecturer, Yale University

Dr. David Levermore (ASCAC)
Professor, Mathematics
University of Maryland

Dr. Juan C. Meza
Division Director
Division of Mathematical Sciences
National Science Foundation

Dr. Daniel A. Reed (ASCAC)
Senior Vice President for Academic Affairs
Professor, Computer Science and Electrical & Computer Engineering
University of Utah

Dr. Roscoe Giles (Chair)
Professor, Electrical and Computer Engineering
Boston University

Table of Contents

Executive Summary

In September of 2018, the DOE Office of Science charged ASCR's Advanced Scientific Computing Advisory Committee (ASCAC) to "…identify the key elements of the Exascale Computing Project (ECP) that need to be transitioned into ASCR's research program or other new SC/ASCR initiatives after the end of the project [and] to address the opportunities and challenges for future high performance computing capabilities." This report has been prepared as ASCAC's response to this charge.

By design and in its evolved execution, the Exascale Computing Project has organized activities of a large community of people – some directly part of the formal project, many more who are stakeholders in related scientific, technical, research or business efforts. To understand the impact of ECP and assess next directions, we have made an effort to speak directly with a wide spectrum of the ECP community and its stakeholders. Our findings and recommendations reflect both our understanding and appreciation of the technical and scientific achievements of ECP as well as the engagement of the DOE lab communities, researchers and stakeholders in its activities.

(A) Advancing and Building on ECP

We find that ECP is a successful project in multiple dimensions including its primary objective of organizing large complex resources to design and develop exascale computing systems that will be deployed to satisfy DOE mission needs. At its conclusion, ECP will have created artifacts (evaluation systems, software libraries, demonstration applications) and adopted practices (in software engineering, project management, co-design, stakeholder collaboration) that should be expanded and built upon to realize the full impact of exascale computing.

We recommend that ASCR build a shared software stewardship program to leverage and build on the ECP developed ecosystem to develop, curate, harden, and distribute software essential for effective use of HPC systems. ASCR should collaborate with other DOE offices and select outside entities to support development of key applications, especially those which continue to defy attempts to address them at the exascale level of computing performance and problems involving edge computing. We recommend that the ECP collaboration models be extended as appropriate to hardware and independent software vendors to engage them early and substantively in new directions and that similar collaboration with university groups should be explored.

We recommend that ASCR adopt and incorporate modern project management practices and tools into its programs to facilitate collaborative work between labs and programs.

(B) Advancing ASCR Research

Applied mathematics and computer science research are essential for future progress in advanced scientific computing. In addition to research supporting new emerging themes such as quantum computing and machine learning, there are remain many open research issues in traditional Applied Mathematics and Computer Science for HPC that were set aside to fund the ECP.

For the U.S. to maintain an international edge, in the post-exascale era, Moore's law will not save us. This research, the responsibility of ASCR, cannot continue to be neglected.

We recommend a significant expansion of ASCR research investments and creation of a more stable funding environment that can support research efforts based both at the DOE Labs and by external researchers. Important areas include: algorithms, programming languages, compilers, optimization, productivity, networking, streaming, edge computing, correctness and formal verification, computer architecture, specialization, devices, heterogeneity, modeling and simulation, workflows, security, visualization, automation, distributed computing, and cloud computing.

We also recommend that ASCR create pathways to wider distribution and uptake of research results that make it to the threshold of distribution. This should be an ongoing continuous effort within the research programs.

(C) Current and Future Workforce

The strength and vitality of ASCR activities are defined by the enthusiasm, engagement, and creativity of its talented workforce. We find that ASCR has a skilled and motivated workforce that collaborated to deliver high quality results in the context of ECP. ECP built ties and trust between the DOE/SC/ASCR communities. Diversity, equity and inclusion initiatives are supported by leaders and the researchers and staff we interviewed.

Addressing the workforce issues of retention, diversity, and opportunities for innovation will be critical during the transition to the post-ECP environment. The committee heard loud and clear from leadership and researchers that there are concerns about what funding and programs will be available after the ECP ends, and the need for ASCR to put forward the plan for engaging laboratory talent after ECP ends.

We recommend that ASCR craft programs that will develop the diverse, multi-generational workforce as researchers shift focus back to basic/fundamental problems. Strengthening ties to the universities and broader ecosystem will help sustain a pipeline of diverse, talented, well-trained professionals. Explicit recognition and cultivation of scientific software professional career paths will be important for development of shared, usable, scientific software.

(D) National and International Leadership

DOE's investments and vision for scientific computing continue to lead national and international efforts to advance the scale and impact of scientific computing. ECP represents the latest chapter of DOE leadership in this space.

At the same time, the rise of a large private-sector market for large-scale computing (as well as the proliferation of applications and the diversification of an international supply chain) necessitate a new strategy for maintaining leadership. These new horizons challenge the traditional operating structure of ASCR. The subcommittee heard repeatedly that ASCR research should be 5 – 10 years ahead of the facilities and should anticipate the needs of facilities.

We believe that it is essential for DOE/ASCR to maintain national and international leadership in advanced computing. The recommendations we have made for the transition lay the foundation for ASCR research and ASCR technologies to impact future computing and DOE's mission. To maintain world leadership in scientific computing, DOE and ASCR need to be able to connect to stakeholders across US universities, industry, laboratories, and agencies.

For ASCR to maximize the impact of its research budget, it should leverage the investments of the larger ecosystem wherever doing so makes sense. This requires both situational awareness and mutual understanding of how the associated communities' long-term priorities and commitments align with those of DOE/SC/ASCR.

Acronyms

Context DOE:

ALCF	Argonne Leadership Computing Facility
ASC	Advanced Simulation and Computing
ASCAC	Advanced Scientific Computing Advisory Committee
ASCR	Advanced Scientific Computing Research
BRN	Basic Research Needs
ECI	Exascale Computing Initiative
ECP	Exascale Computing Project
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
LCF	Leadership Computing Facilities
NERSC	National Energy Research Scientific Computing Center
NNSA	National Nuclear Security Administration
OLCF	Oak Ridge Leadership Computing Facility
OSTI	Office of Scientific and Technical Information
SC	Office of Science

Context: ECP

AD	Application Development
HI	Hardware Integration
IAC	Industry Advisory Council
L1	Level 1
L2	Level 2
L3	Level 3
L4	Level 4
ST	Software Technology

Context: General

AI/ML	Artificial Intelligence/Machine Learning
BSSw	Better Scientific Software
BoF	Birds of a Feather
CREATE	Computational Research and Engineering Acquisition Tools and Environments
CS	Computer Science
E4S	Extreme-Scale Scientific Software Stack
ECR	Early Career Researcher
EDA	Electronic Design Automation
EVM	Earned Value Management
FFT	Fast Fourier Transform

FMM	Fast Multipole Method
GPU	Graphics Processing Unit
GRP	Global Research Platform
HPC	High Performance Computing
HPCMP	High Performance Computing Modernization Program
ISV	Independent Software Vendor
LLVM	Low Level Virtual Machine
ONR	Office of Naval Research
OS	Operating System
PDE	Partial Differential Equations
PM	Program Manager
SDK	Software Development Kits
TACC	Texas Advanced Computing Center
UQ	Uncertainty Quantification

Introduction

The nation is approaching a critical point in the evolution of computing. The threshold we are crossing to exascale computing, the rapid development of artificial intelligence (AI), and the realization of quantum devices signal the dawn of a new era for computing and data science - an era of increased prevalence of, and need for, multidisciplinary work. Department of Energy (DOE) mission science requires world class computing capabilities and facilities. Through Advanced Scientific Computing Research (ASCR), DOE has led in computer science and applied math research, which enables advanced scientific computing and has created and maintained world class computing facilities[1]. The challenge now is for the U.S., DOE, and ASCR to continue leading in advanced computing throughout the transition to a new era.

The DOE Exascale Computing Initiative (ECI) is pioneering the design, development, and effective deployment and use of a new generation of advanced scientific computers. The comprehensive DOE Exascale Computing Project (ECP), jointly supported by the Office of Advanced Scientific Computing Research, in the DOE Office of Science (SC) and the Advanced Scientific Computing (ASC) program in the National Nuclear Security Administration (NNSA), is creating the software, system designs, and prototype applications for these computers.

This transition report focuses on the next steps ASCR should take beyond ECP's conclusion in 2024. By design and in its evolved execution, the Exascale Computing Project has organized activities of a large community of people – some directly part of the formal project, many more who are stakeholders in related scientific, technical, research or business efforts. To understand the impact of ECP and assess next directions, we have made an effort to speak directly with a wide spectrum of the ECP community and its stakeholders. Our findings and recommendations reflect both our understanding and appreciation of the technical and scientific achievements of ECP as well as the engagement of the DOE lab communities, researchers and stakeholders in its activities.

We also discuss the relationships and impacts of the ECP activities on broader ASCR research activities both their synergies and stresses placed on the research programs. The DOE Laboratories and laboratory staff have been critical to ECP, and the health and diversity of the workforce is a concern during the transition.

Our broad recommendations are that ASCR: (A) sustain and build on the fruits of ECP, (B) invigorate and broaden its base research programs in computing and mathematics, (C) build and grow the workforce needed to realize its goals, and (D) maintain and extend its national and international leadership in advanced computing. The following chapters discuss our findings and recommendations in each of these four areas.

Note Success requires ASCR to continue its close collaborations with other DOE program offices as well as with the scientific community in universities and industry. The very existence of the exascale program is predicated on the longstanding collaboration between SC/ASCR and NNSA/ASC in addressing shared advanced computing challenges. The continued collaboration

[1] ASCR@40: Highlights and Impacts of ASCR's Programs. https://doi.org/10.2172/1631812

of these DOE offices is essential to maintaining leadership, attracting new talent, and providing the foundation for long-term DOE mission success.

The scope of the charge for this report is to consider ASCR's transition to the post-ECP environment. Our findings reflect the successes and challenges of both ASCR and ASC, while our recommendations necessarily are focused on ASCR.

About this report

In September of 2018, the DOE Office of Science charged ASCR's Advanced Scientific Computing Advisory Committee (ASCAC) to

> "…assemble a subcommittee to identify the key elements of the Exascale Computing Project (ECP) that need to be transitioned into ASCR's research program or other new SC/ASCR initiatives after the end of the project to address the opportunities and challenges for future high performance computing capabilities." [See the full charge in Appendix A]

In response, ASCAC formed a subcommittee to prepare this report. The subcommittee worked throughout 2019 and early 2020 on this report as ASCAC's response to the charge.

A key activity of the subcommittee was to meet with and interview leaders in ASCR, DOE, and the ECP as well as stakeholders in the community. The subcommittee held 3 Community meetings (2 special workshops and 1 "Birds of a Feather session (BoF)" at the SC2019 Conference), 13 smaller interview sessions, and about 40 subcommittee discussion meetings.

We have highlighted some of the voices from the community in this report – their eloquence is often more powerful and passionate than ours. We are grateful to all who participated and shared their experiences and views with us.

A: Advancing and Building on ECP

ECP's stated mission [https://www.exascaleproject.org/about/] is to

- **Develop exascale-ready applications** and solutions that address currently intractable problems of strategic importance and national interest.
- **Create and deploy an expanded and vertically integrated software stack** on DOE HPC pre-exascale and exascale systems.
- **Deliver US HPC vendor technology advances and deploy ECP products** to DOE HPC pre-exascale and exascale systems.

Based on our review of ECP's progress so far, their plans for the future, and their external reviews, we find that ECP has been highly successful to date, and we expect the project to succeed in this mission. In addition, there are vital lessons learned that can inform ASCR in its future efforts to advance DOE computing and advance research in applied math and computer science.

There is considerable documentation describing the evolution, execution, practices, and achievements of ECP available from ASCR and on the ECP website[2]. These include broadly accessible documents such as the ECP software technology[3], application development[4], and hardware integration[5] update reports.

This section of the report discusses our assessment of ECP with particular attention to the lessons learned from ECP and the activities needed to sustain and build on ECP's successes in the future. We present findings regarding ECP's activities and accomplishments and recommendations to ASCR for building on ECP's outcomes as well as ECP's organizational and management models.

Finding A.1: ECP has been successful overall

ECP is a successful project in multiple dimensions including its primary objective of organizing large complex resources to design and develop exascale computing systems that will be deployed to satisfy DOE mission needs. At its conclusion, ECP will have created artifacts (evaluation systems, software libraries, demonstration applications) and adopted practices (in software engineering, project management, co-design, stakeholder collaboration) that should be expanded and built upon to realize the full impact of exascale computing.

The new exascale computing architectures all embody the extreme heterogeneity of modern advanced computers.[6] They are composed of thousands of nodes, each of which contains multiple

[2] Exascale Computing Project. https://www.exascaleproject.org/
[3] http://exascaleproject.org/wp-content/uploads/2019/11/ECP_ST_update_2019_11_06-am-spreads.pdf
[4] Exascale Computing Project. Addressing a National Imperative. Application Development Update. September 2019. http://exascaleproject.org/wp-content/uploads/2019/11/ECP_AD_update_2019_11_25_spreads.pdf
[5] Exascale Computing Project. Better Scientific Productivity through Better Scientific Software: The IDEAS Report. January 30, 2020. https://www.exascaleproject.org/better-scientific-productivity-through-better-scientific-software-the-ideas-report/
[6] Productive Computational Science in the Era of Extreme Heterogeneity. DOE ASCR Report for Basic Research Needs Workshop on Extreme Heterogeneity, January 23-25, 2018. https://orau.gov/exheterogeneity2018/2018-Extreme-Heterogeneity-BRN-report-final.pdf

Graphic Processing Units (GPUs). The memory and communication systems reflect the complexity of these multiple levels of organization.

Designing, characterizing, optimizing, programming and ultimately getting useful science out of such systems is a very large team effort. It requires co-design of hardware, software, and applications. It requires incorporation and integration of many years of engineering and science application experience, previous accomplishments in mathematical and computer science research, and a clear vision of how to progress towards the goal of a usable and transformative system for a broad range of key DOE applications. Finally, it requires dedicated teams of people, inspired by the vision of success of the project, together with effective organization and leadership.

ECP has worked very hard and succeeded in all these dimensions. ECP's major thrust areas – Software Technologies (ST), Applications Development (AD), and Hardware and Integration (HI) – are well coordinated and effective in the individual and interconnected roles.

The ECI and ECP allow for a scale of operations, coordination, collaboration, and focus on applied computing that transcends anything that preceded it. It has revealed new potential for enhanced ASCR activities and improved effectiveness of traditional ASCR activities.

Beyond the progress towards exascale computing, ECP has taken significant steps towards a *new paradigm* for ASCR's strategy to translate advances from its basic research investments into deployment and practice at scale. The increased coordination has raised awareness throughout DOE and its network of partners regarding expertise in different teams, as well as software.

The sections that follow detail some aspects of ECP's activities which are particularly important as we transition out of ECP and consider what ASCR must do in the future, especially in light of the strong consensus that much of the kind of work done in ECP is similar to what will need to be done in the future.

Finding A.2: ECP successfully managed a large distributed collaboration

ECP's management practice contributed to the successes of the project.

As noted in Finding A.1, ECP shifts the overall paradigm from ASCR research happening in isolation as the default, with collaboration as a second stage, to collaboration as a default mode of operation, with accompanying activities and "hooks" built in from the start.

At the SC/ASCR level, this meant (1) separating functionalities in innovative ways (ECP was separate from hardware), (2) establishing processes and plans to build and maintain alignment across labs, and (3) taking care to avoid unfunded mandates, especially in the translation from research to facilities. ECP has adapted rigorous project management requirements (413.b)[7] to the exigencies of a large computer system development project incorporating parallel co-design and software development activities.

[7] DOE O 413.3B Chg 5 (MinChg), Program and Project Management for the Acquisition of Capital Assets, https://www.directives.doe.gov/directives-documents/400-series/0413.3-BOrder-B-chg5-minchg

At smaller scales, this meant learning how to use (and adapt) project management methodologies, as well as progress towards standardizing version control (Gitlab) and supporting more automated testing (continuous integration). Cybersecurity is an essential component of distributed collaboration, and ECP has opened new pathways to improve collaboration while maintaining security.

Adaptation of project management for research/development activities is challenging; ECP's lessons learned include ways to balance lightweight PM with supporting delivery by adding the right process at the right time.

"The ECP is managed, to the greatest extent feasible, as if it resided within a single institution. Lines of authority and responsibility follow the organization structure as documented in the ECP PEP. L2 leaders within the ECP report to the ECP project director, and the reporting within the management structures of the partner laboratories reflect and accommodate this arrangement. The overarching principle is to minimize the effects of boundaries between the core partner laboratories (Section 4.2) and the ECP project director, with reporting lines moving through the ECP organization and seamlessly across ECP."[8]

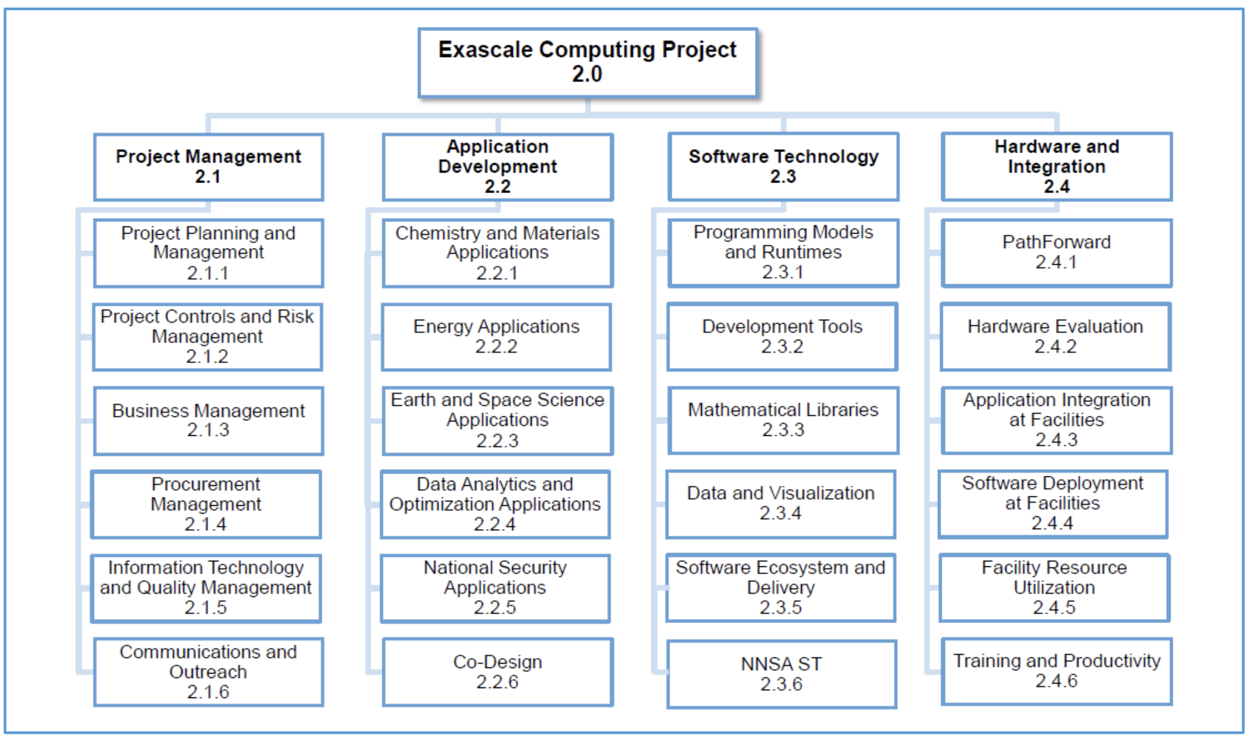

***Figure 1**. ECP WBS, Level 2. (Source: ECP Management Plan v1.1).*

The figure above shows the breakdown below Level 2. "The three technical focus areas and project management area are shown as L2 activities in the WBS and provide the overarching structure for the ECP. The WBS L3 categories within each technical focus area were developed to group subprojects with common themes and to balance the overall workload for each L3 lead.

[8] ECP Management plan v1.1, September 2019.

Within each L3 are individual projects at WBS L4." The L4 technical leads are principle investigators of individual subprojects at L4.

Agile management practices, lightly integrated with EVM project management, are of great help in effective planning, coordination, and reactive modification of software technology R&D activities. Deep expertise and years of experience of the ECP L2 and L3 software technology managers together with outstanding L4 technical leadership have proven valuable throughout planning and development of the ECP software ecosystem. The L4s and their teams made success possible by finding ways to adapt traditional research activities to support the overall shift towards integration and rapid translation from innovation to production and deployment at scale.

Building this ecosystem has also highlighted benefits of standardizing the R&D toolchain across sites, including engaging as customers (as when the six labs worked with GitLab, pooling their purchasing power) and as community partners (e.g., working with the LLVM community on compiler infrastructure; implementing modifications for Jupyter Hub, making the software fit-for-purpose (at the labs) and useful (to the larger community)).

> Some quotes from our interviews:
>
> "… collaborations that were formed under ECP are amazing. I'm proud of the fact that our project has unified a field in the US that is fractious and contentious at best."
>
> "ECP helped with collaborative teams. Many individual projects are cross labs, all-hands meetings. Move focus from lab boundary to concept boundary."
>
> "Should look at this [as the time] in history when DOE went to a distributed workforce."

Finding A.3: ECP created a software ecosystem focused on DOE science mission needs

ECP has created a well-designed software ecosystem for development, curation, and distribution of exascale systems and application software. This ecosystem integrates the fruits of years of basic research in: mathematics, computer science, applications, and systems software.

In particular, the ecosystem greatly reduces barriers for ASCR fundamental research maturation and impactful delivery at the facilities and with users. Several of our recommendations focus on realizing the potential of this new ecosystem.

ECP's creation of a software ecosystem has benefited from the project's unprecedented support for building relationships between applications, software technology, hardware integration, and facilities. Co-design (linking applications focused mathematicians and scientists, software developers, and hardware integrators at the facilities) has enabled relatively rapid progress in identifying new challenges of accelerated architectures and overcoming them. Continuous integration – systematically keeping software and packages up to date, tested and consistent across multiple facilities – was consistently and deeply supported. This included work on tools like Spack and GitLab that help enable continuous integration. It also supported the multiple software delivery vehicles mentioned above. The ecosystem integrated significant contributions from ASC as well as ASCR.

By contrast, in the previous ASCR software research and development paradigm, smaller teams tried new ideas largely in isolation, and there was no accepted process for maturing successful research results for end use. Success in these isolated, small-scale efforts would then lead to an extended evolution, at different times progressing on dimensions of scale, generality, or suitability for integration with other tools and technologies.

Several cracks in the old paradigm had emerged and become more and more of a hindrance to end-user delivery, especially at large scale:

- It was challenging to combine modules with different functionality into a larger single science application both because of poor interoperability and other differences in expected/assumed operating environments.
- Inconsistent environments on different computers at different labs hindered diffusion of awareness, and therefore leveraging, of powerful new tools; and
- Inconsistent environments at different labs also hindered ASCR's agility by making it difficult to identify methods whose strengths were highly specific to an application or computing platform.

The increasing complexity of the hardware stack risked driving development of ever-more specialized responses, which could not be capitalized upon by small teams acting largely in isolation. For such teams, there could not be reasonable expectations that they would, on their own, have the situational awareness of these capabilities, or the technical capability in-house, to leverage them.

> **E4S -- Extreme Scale Scientific Software Stack**
>
> E4S is the distribution mechanism for the full ECP ST software stack
>
> E4S is a community effort to provide open source software packages for developing, deploying, and running scientific applications on HPC platforms
>
> E4S addresses the need, within and beyond ECP, for correctly installing in the end users compute environment(s) a complex, integrated and interoperable stack of tools and libraries
>
> E4S provides both source builds and containers for a broad collection of HPC software packages.
>
> E4S goal is to accelerate the development, deployment and use of HPC software, lowering barriers for all HPC users.

The new paradigm is an ecosystem that includes multiple federal agencies, industry (companies from small to large, acting variously as vendors to DOE, as technology translators (e.g., Kitware), and as users, e.g., many members of the Industry Council), academia, and community organizations such as standards bodies. The ecosystem includes both tools (artifacts) and patterns of behavior (practices). The ECP artifacts demonstrate the power of the new approach, as do the practices. Delivering exascale capabilities to science and the nation will require expanding and building on both. They require different approaches, which we address in the following recommendations. In addition, Recommendation D.1 addresses the ecosystem role in maintaining ASCR's national and international leadership.

It is important to note that ECP has also created significant amounts of educational and training materials, which have been important to the project's successes (including building relationships between different communities) and will continue to be critical work products for sustaining ECP's benefits.

ECP integrated content development, training, and different modes of training delivery including webinars, seminars, and workshops. Highly attended monthly HPC webinars have attracted researchers outside ECP. The Better Scientific Software (BSSw) fellowship gives recognition and, in partnership with NSF, is being expanded and made broader.

Overall, ECP training has been created collaboratively and is effective. Previously, NERSC, ALCF, OLCF had excellent individual training programs; ECP created an umbrella for these programs and demonstrated the value of training one another, not just external researchers. Training includes more than simply technical content for example, ECP software libraries and tools, incorporating exercises on communication and teamwork practices.

Training and workshops at the ECP Annual Meeting, the SC Conference, and other such meetings are successes.

ECP's overall focus on high-quality software development practices has made numerous positive impacts in building the software ecosystem. ECP has exercised good judgment in choosing external tools that are available in the community and showing how to use them within a DOE application space.

Concrete outcomes of the software effort include:

> Spack is an exemplar of ECP infrastructure artifacts having wide impact beyond ECP
>
> Complex ECP and other HPC applications often contain hundreds of thousands or more lines of code and rely on large numbers of libraries or other packages. Spack is a high impact software infrastructure tool which greatly simplifies the management of such large, complex codes
>
> Funded by ASC prior to ECP, Spack was developed to facilitate the management of code interdependences through automating and facilitating complex code compilation via Python scripts
>
> Standardization on the use of Spack by ECP application codes has made numerous application communities aware of the value of Spack utilization and facilitated its wide-spread use
>
> Sustained funding for tools such as Spack is essential to enable them to adapt and evolve to support upcoming hardware and software environment changes

- Applications
  - Creation of a process for applications deployment to ASCR and ASC facilities.
  - Application Development (AD): 24 applications, 6 co-design centers, cross-platforms.
  - Broad buy-in from senior management in the DOE Office of Science and Applied Offices as well as NNSA.
- System software
  - ~70 software products organized into 6 Software Development Kits (SDKs): (Programming models and Runtime; Compilers and Support; Tools and Technology; Math Libraries; Visualization Analysis and Reduction; and Data Management, I/O Services, and Checkpoint Restart).
  - Dependence database to identify and track software interactions in the application and system software stacks.
  - Coordination of 6 principal DOE Labs: software development, engineering, hardening for transition to facilities and applications.
  - Effective use of modern software tools (Atlassian, Spack, continuous integration, Git Workflows, and containerization) for collaboration, delivery/deployment integration and coordination.

- Multiple software delivery vehicles: single products from source, SDK groupings for related products, and the Extreme-Scale Scientific Software Stack (E4S) for full suite installation.

Finding A.4: ECP supported collaboration with facilities and industry

ASCR and ECP/ECI have effectively collaborated with industry and the facilities to develop exascale computing technology and industry applications.

The ECP PathForward program, which engaged computer technology companies, followed on in the spirit of the earlier ASCR FastForward and DesignForward programs. Collectively we refer to these as "**xForward**" programs. The goal was to invest in collaboration with vendors in order to leverage their design and production capabilities to make available system components and systems more aligned with DOE's exascale needs.

One PathForward participant said

> "While AMD's DOE-directed research and development has focused on HPC capabilities, the performance, power, and end-user productivity benefits are far reaching and impact AMD's technologies for enterprise data centers, cloud computing, machine learning solutions, desktop and laptop processors, game consoles, and embedded systems. […] In addition, the program has enabled AMD to provide high-quality careers and training in the United States, feeding fresh expertise into DOE's and the nation's workforce."

The earlier xForward programs realized significant benefits from bringing industry engineers and designers into a situation to better understand the DOE needs in systems and to be able to design elements of systems to satisfy those needs. As part of ECP, there was also funding and commitment to support ECP project Hardware Integration team members and lab personnel spending substantial time on the interaction with industry counterparts in co-design activities.

> "One interesting thing for the first 15 years or so of [x]Forward, there's not a lot of funding given to lab people to interact with vendors. With ECP, there was actually money set aside to have someone go work collaboratively."

In ECP, the Hardware Integration teams collaborated closely with Facilities on applications integration, resource utilization, hardware evaluation, xForward, deployment of ECP software, training, and productivity. In addition, there was a culture of close interaction and coordination between AD and HI.

ECP's ability to coordinate teams and groups of teams with common goals and an evolved common culture bridged gaps that historically have challenged collaborations as well as gaps between vendors and the labs. ECP also supported more real collaboration with AD and software teams and the facilities; as one participant explained ECP's positive impact,

> "Facilities don't have the resources to support the software. ASCR is way out there. There's nothing in the middle. Always seen as a money problem. ECP supported bridging this gap".

ECP also has an Industrial Advisory Council (IAC), which we interviewed. The Industry Council provides a vehicle for engagement of industry leaders with HPC experience, needs, and insights to interact with and advise ECP. These perspectives are very valuable for insights on how the developing exascale systems and facilities might impact the wider computing ecosystem and benefit the nation. A particular insight we gained from our discussions with the IAC was a sense of the time scales and level of support for software and systems needed to attract industry adoption. The time scales desired are on the order of 5-10 years or more. This mirrors what many applications users would prefer before committing to adoption of new systems.

Recommendation A.1: Create a shared-software stewardship program within ASCR

ASCR should create a comprehensive program that leverages the ECP ecosystem to support and curate shared software. This should incorporate ASCR program office oversight while delegating operational control to a software engineering team of laboratory and academic experts.

Software Stewardship Vision

The focus of the proposed stewardship program is to support and extend DOE shared software products, starting with the ECP software stack and extending over time to other software that may be "productized," that is, made available to wide, shared use and "hardened" for reliable use by diverse users. The hub will provide a context for support of software packages in a coherent shared environment that enables packages to inter-operate smoothly, especially on the range of new heterogeneous architectures.

Stewardship is vital
Our community discussions indicated a strong consensus on the need for some form of software stewardship.

- The subcommittee heard strong consensus that software sustainability is a central concern and should be a top priority, and that the ECP transition is an opportune moment, because of its focus on software. Existing mechanisms for software distribution and support have shortcomings that few, if any individual projects, have been able to transcend. ECP's success in tracking, coordinating, and continually improving software in its domains provide a model that can be built on to sustain and broaden success.
- The successes and wider adoption of ECP artifacts, especially ST (e.g., at the Texas Advanced Computing Center (TACC), in the UK at Hartree Centre), provide additional motivation to solidify and build on this foundation. The work entailed does not exactly fall within research, nor within facilities. Establishing a program with a clear, decisive agenda to meet this need can sustain the next level of computing and bridge the gap between research success and end user success. A participant told us, "The problem is when software is used but not getting research funding."

Leadership for the long term
We believe the software hub should be led by DOE laboratory and academic software leaders who possess long-term experience with DOE software development and delivery.

The subcommittee heard strong consensus about the need for a stable (long-lived) path for productizing and deploying research software funded by ASCR during ECP, the transition, and beyond. The shared-software stewardship program provides such a path by supporting ECP products that drastically reduce barriers for development of performance-portable software and its production hardening and deployment.

The ASCR@40 Report (2020) highlights successes that represent long-term investments over multiple decades. Centers like the software hub should help these successes endure and have projects' outcomes last long enough to realize a long-term effect.

Responsibilities
Some Lab efforts are described as playing a key role between the work in academia and the work in industry. For example, industry sells software products (and support), whereas in academia, as one senior faculty member put it, "Doing software is not credit-worthy at universities." ASCR stewardship support for shared software provides a middle ground, a basis for multiple groups (academics, government, and industry) to adopt implementations of innovative work in research software.

Software libraries provide an important example, and Fast Fourier Transforms (FFTs) are a concrete case study in software libraries. Many applications depend on FFTs, which are well suited for distributed systems but the technical challenges are largely in implementation rather than in basic mathematics and computer science research. More generally, facilities do not have resources to support libraries, and managing facility resources to support libraries is complicated: for instance, hardware vendors are paying HDF5 ISVs using acquisition money.

> Comments from our interviews:
>
> "Across SC, there's a fear of adoption of [software components] because they don't know if the [software] will be [sustained] … [it is] often easier/perceived [to be] safer to pay vendor support."
>
> "[It] is hard to transition research after it's done. You should involve the people who it will go to. So, the question is, who will it go to?"
>
> "[It is] not clear people always think about sustainability when they start, [but] people should keep that in mind."

Tools
ECP ST has developed and deployed an array of new tools for managing aspects of the expected computing landscape over the coming decade, sometimes called the era of extreme heterogeneity. A post-ECP concern is how to focus on the coordinated deployment of technologies.

- These tools are successful due to a combination of artifacts (Kokkos, RAJA, Spack) and practices (working with OSTI and GitLab for continuous integration). To deliver continuing returns in DOE mission space, both the artifacts and practices must be expanded on.
- These tools build on "seed corn" from earlier investments in basic research, highlighting the importance of stable, long-term funding for basic research in the full area of disciplines stewarded by ASCR.
- There are numerous remaining challenges for this coming era, and they too will need both stable funding to grow new seed corn, and sustained support; the Basic Research Needs Workshop on Extreme Heterogeneity identified the following Priority Research

Directions: maintaining and improving programmer productivity; managing resources intelligently; modeling and predicting performance; enabling reproducible science despite non-determinism and asynchrony, facilitating data management; and analytics and workflows.

- The great opportunity for the post-ECP era is that supporting the artifacts and practices is a cost-effective way to meet the other challenges of extreme heterogeneity. In essence, the practices and ST tools funded by ECP become a “force multiplier” for more efficient translation of basic research into capabilities and delivery of the DOE/ASCR mission.

Models

There are models for the elements of such a program in the HPC community in labs and in industry as well ASCR and ECP itself.

One model for a stewardship program is the Computational Research and Engineering Acquisition Tools and Environments (CREATE) program at HPCMP, started by Doug Post, that takes best research into production codes, has regular releases, and a Quality Assurance Group. CREATE takes the important next step for software, but also takes care not to raid research.

It is also important to recognize that industry has substantial expertise in developing HPC tools. There are profitable U.S.-independent software companies developing and licensing specialized math libraries, compilers, operating systems, debuggers, and visualization tools. For example, large HPC users (e.g., General Electric and United Technologies) depend on independent software vendors (e.g., Cascade Technologies, Altair Engineering, ANSYS) for solutions and tools that have a long-term roadmap and support for many years. The stewardship program should leverage external expertise whenever possible and emphasize open competition to attract ideas and submissions from the widest possible array of experts in government, academia and industry.

The committee considered whether the stewardship program might be framed as a new kind of facility. Our opinion is that the development and software engineering in the Stewardship program is closer to research than the current leadership computing facilities, and so should stay closer to research. The discussion was that stewardship will *deliver* the software, while the facilities will *deploy* technology. That said, the software stack is like a beamline, and needs continued investment; a regular release cycle corresponds to “upgrading.” We recommend that the stewardship program go beyond simple sustainment. The software will need to evolve both in response to improvements in the hardware ecosystem and in response to changing application needs.

The Stewardship Program would focus particularly on the software stack. Support for Applications should be maintained by the other Office of Science and NNSA programs that use them (Rec. A.3).

Recommendation A.2: Engage current, and anticipate future, software needs

Important software and algorithms can originate outside of DOE and ASCR. ASCR should continue to monitor and anticipate external developments in critical areas and incorporate this information in planning the evolution and modernization of software. This activity overlaps with but extends beyond the scope of the Stewardship program.

Realizing the maximum potential of the investments in ECI and ECP will require ASCR to be consistently intentional about its efforts and mechanisms to track and predict the rapidly evolving landscape. ASCR has always tracked and anticipated the computational needs in DOE Science. Now the rate of technological change is increasing, in part due to enormous investments by industry in emerging areas such as AI and quantum computing. Increasing external rates of change necessitate correspondingly increased agility in ASCR's approaches to preparing for change, planning for change, building research programs, and adapting at pace.

Many of ECP's successes arise from recognizing that new, more integrated teams are required to deliver important research and software. ECP has provided its multi-institutional, highly distributed teams with the needed new tools. Successful import and adaptation from industry is a trend we expect to continue. Reducing barriers to translate research findings to deployment and production represents a major success and should be continued.

More broadly, the recommendations here aim to ensure that the high-risk research process is as lightweight as possible, while protecting the capabilities for application and integration that represent a signature accomplishment of ECP. New tools and ways of working support this goal by increasing capacity to introduce the right level of coordination at the right time. It is also important to provide improved and new mechanisms for increasing the degree to which world-class technological expertise informs better, timelier management decisions.

The ECP transition is a timely opportunity to consider the organizations and communities whose needs ASCR aims to meet. Such clarification will be helpful to define priorities in ASCR's technical work and its partnering strategy. As an example, the challenges and needs of non-ASCR user facilities can be quite different from ASCR's current portfolio: data acquisition rates and needs for persistent storage, are increasing exponentially, even as there is increasing interest in performing significant amounts of computation on the data, e.g., for AI. For meeting these needs, and others (e.g., tools for co-design), sustained and deep relationships with industry can be enormously valuable.

A significant opportunity in the post-ECP era is to enable an education and outreach program to other parts of the Office of Science, communicating the tools and models of ECP. Building such relationships will benefit from cautious and incremental approaches, as success requires sustained engagement and establishing shared understanding between communities before emerging needs can be communicated in a substantive, actionable way.

These recommendations are therefore more holistic than in Recommendation A.1, involving not only software but the landscape of software needs. Some of the efforts suggested are not technical research as in science, but rather "scanning the horizon" and building relationships (e.g., understanding the overall ecosystem better, see Recommendation D.1). As an immediate example, the subcommittee strongly recommends that ASCR support continued development and delivery of the training materials created during ECP.

We learned about several areas of current and future needs that should be highlighted. As breakthroughs in these areas emerge from ASCR or external research, we should identify which are ripe for broader effective deployment and directly or collaboratively work to bring them to the level that they become part of the shared software facility.

Legacy Code Transition

ECP has performed major work with key DOE applications which are still dependent on a Fortran legacy code base. We heard from developers and users about this issue:

- "ECP has allowed us to accelerate development by an order of magnitude, we are replacing code that took 20 years to develop and we are going to do it in 5 years. But I do not have any idea how we are going to keep that going. I am really scared from an effort standpoint, and if we do have to move away from Fortran, that is a much bigger investment".
- "[We] … have a 1M line application that certain industry partners use and [that] their Quality Assurance validated. It is a massive investment to even investigate the use of a new programming model into a code base. For example, when Kokkos was first introducing something, we dedicated 1 person for 2 months to put it through its paces before we fully invested in it. When we go to GPUs and suddenly we can't guarantee reproducibility beyond 6 or 7 significant digits, these are challenges. Then you start bringing in programming models and everything else. These are big investments in time and we tend to minimize how much those really are, especially for codes that have some level of legacy investment".

There may be promising ways of accelerating the transition of such codes. Flang (the Fortran complier integrated with LLVM), for example, might help by providing an essential pathway for working with Fortran code. Finding migration paths forward for key applications or replacements for them is a key concern.

Co-design and extreme heterogeneity

Research in systems software, algorithms, as well as aspects of heterogeneity will need to be translated rapidly to emerging hardware/software platforms. From a microelectronics perspective[9] we look forward to "deep co-design" where algorithms inform the structure. There should be mechanisms to assess these research developments and understand how to bring them to bear on systems addressing DOE's needs.

Deployment & Workflows

Productivity depends both on the underlying performance of systems and the high level tools and contexts that link them to applications and users. There are important emergent paradigms inside and external to DOE.

[9] Basic Research Needs Workshop for Microelectronics, Report of the Office of Science Workshop on Basic Research Needs for Microelectronics October 23 – 25, 2018. https://science.osti.gov/-/media/bes/pdf/reports/2019/BRN_Microelectronics_rpt.pdf

DOE has recognized the importance of edge computing and in-situ data management. It will be important to develop systems and systematic approaches to make these tools widely available.

Productivity is enhanced when there are high level tools and frameworks to ease user access. Effective and well distributed tools developed outside DOE/ASCR should be able to be brought into the ASCR community. Examples include Python, Julia and Go programming languages.

Finally, capabilities and tools for cloud computing are undergoing explosive growth. Careful attention to the elements of these systems that are applicable (and secure enough) to have a place in ASCR’s portfolio should be maintained.

Recommendation A.3: Collaboratively support applications

The Application Development (AD) activity within ECP brings together teams of applications scientists, software developers, and mathematics and algorithmic experts focusing on defining and demonstrating successful exascale applications across many fields. This activity is vital to the development of impactful exascale systems. We believe that key elements of it should be transitioned into post-ECP ASCR programs and activities.

The AD activity has some similarities to the long running and very successful ASCR SciDAC programs. It brings together multidisciplinary teams focused on well-defined computational science problems at extreme scale and directly engages relevant DOE science program offices that support the related disciplinary research. It builds on existing and evolving mathematical and algorithmic work in the computational and math institutes. It also has built an admirable level of trust between stakeholders and partners that can serve as a foundation for future activities.

As part of ECP, the AD program has additional focus on ECP related performance and science metrics and baselines and regular oversight and leadership's attention to (and support for) progress. From the perspective of DOE program managers we met with, the extra attention and feedback from ECP about applications projects’ challenges and successes were considered a positive aspect of the program. From researchers and developers we heard positive comments about the level of collaboration and the excitement of their creativity and successes. There was also a natural tension between the productivity goal focus and the need to support discovery in science, especially to be able to explore unanticipated science directions.

Because ECP is explicitly a multi-lab effort with support from lab leadership, the AD efforts as a whole involved many cross-lab collaborations and shared efforts. Of course, the focus on continuous integration and multi-platform system software is foundational for multi-lab deployment.

ECP also provides the funding that enables extensive collaboration and reporting with the AD activity. During ECP, this funding stream has been external relative to the applications’ offices.

The vision of the legacy of ECP in the application space is to build focused collaborative teams across disciplines to adapt and apply extreme computing to address significant problems. The experience of working together on ECP has enhanced the level of trust between program offices, labs, researchers, developers, and practitioners. This is a rich context for moving forward rapidly. The pace of technology change and improvement and the scope of computing needed to solve the most pressing DOE problems requires such speed and effectiveness.

We recommend that:

1. ASCR, in collaboration with other program offices, establish a new program that encourages coordinated interdisciplinary teams to address science problems amenable to extreme computing both in terms of computation and data sizes/volumes. The focus initially should be on problems that continue to defy attempts to address them at the exascale level of computing performance and problems involving edge computing. It is especially interesting to be able to view the interplay of architecture and application performance from the perspective of adapting the architecture to the algorithm. Careful attention should be paid to the development of appropriate science or performance metrics and baselines using lessons learned from the ECP application suite. ECP AD team participants would be eligible to recompete for such projects.

2. Scientific software for applications that is shareable among many research groups should have a pathway to be hardened, generalized if necessary, and "productized". It could become part of the spectrum of resources available from the Shared Software Hub.

3. SciDAC itself should be continued, engaging researchers in universities as well as the labs.

4. Consideration should be given, as was done in ECP, on how to work effectively on problems defined by the applied DOE offices. This includes mechanisms for joint support of projects.

5. All these programs should be jointly supported by ASCR and the relevant program offices.

6. ASCR should also welcome collaborations with non-DOE science entities on a shared support basis.

Recommendation A.4: Broaden industry and academic engagement

We recommend that the xForward model be extended as appropriate to hardware and independent software vendors to engage them early and substantively in new directions and that similar collaboration with university groups should be explored.

The current PathForward activities are winding down, but the supportive relationships with vendors and the DOE participants are still extant and can be built upon for future system needs. This is a time when new architectural models and systems are rapidly being created to meet new needs for edge computing and machine learning. These areas will remain on the forefront of DOE and ASCR interest and xForward programs can still serve an important role in engaging hardware vendors.

We also recommend that similar opportunities for co-design and collaborative work be offered to independent software vendors. There are important success stories of software based on DOE investments which have been adopted by the wider community with involvement of hardware and software vendors in "productizing" and pushing them out to wider communities and forming the support system for them. Examples include MPICH and HDF5.

ASCR and DOE will learn a lot from working on applications on extreme heterogeneous exascale architectures. Sharing lessons learned, SDKs, and best practices with industry (ISVs) will be important. Having ISVs that can extend the impact of DOE actions will benefit the community.

From an IC member: "… there are many areas within DOE and outside where exascale could be making a huge impact. Embracing a larger mission, exascale would be able to have a greater impact on the US economy and competitiveness and PathForward and broader participation, leadership by other entities (small business, universities)."

We also want to emphasize that university collaborations can play a unique role in broadening the impact of ASCR and in exploring the design space for new systems. Universities can turn things around more rapidly than labs. "Students take more risks than lab scientists". Universities can take deep dives in new areas.

Recommendation A.5: Adopt modern project management practices

The purpose of this recommendation is to enhance the effectiveness of ASCR program managers and to allow for consistent and clear communication between managers, project leaders, teams, and individuals. We were impressed with the level of communication in ECP and the aggregation and distribution of information within the project. An effect of this was the ability for expert technical leadership and decision making at multiple levels of the ECP management structure.

We recommend that ASCR adopt and incorporate modern project management practices and tools into its programs to facilitate collaborative work between labs and programs. ECP has demonstrated the importance of having strong expert technical management embedded at many levels of a large project. We encourage ASCR to follow such a model for the programs and program managers that build on ECP.

As ECP discovered, its development of exascale technology is not the same as a traditional (413.b) construction project (with predetermined goals and milestones that can be laid out in the beginning before the start of construction). ECP had to adapt the requirements of the (413.b) process to its circumstance where co-design and research and discovery could shift some of the goals and methods during the course of the project. Many ASCR programs are even less like a construction project (for example long-term research) and we expect that ASCR would need to adapt the use of the management tools to its circumstances.

Some important considerations are to use the tools to enhance communication and information sharing; to allow consistent and timely messaging throughout a program and across ASCR itself; to empower project and technical leadership to be more distributed through the levels of a project; and perhaps most challenging to ease rather than enhance the reporting burden throughout most of the hierarchy.

B: Advancing ASCR Research

Finding B.1: Applied mathematics and computer science research is essential for future progress in advanced scientific computing

There are still many open research issues in traditional Applied Mathematics and Computer Science for HPC that were set aside to fund the ECP. See Recommendation B.1 for specific topics, which fall in the areas of algorithms, programming languages, compilers, optimization, productivity, networking, streaming, edge computing, correctness and formal verification, computer architecture, specialization, devices, heterogeneity, modeling and simulation, workflows, security, visualization, automation, distributed computing, and cloud.

Prior to and during ECP, ASCR funded close to a dozen workshops outlining essential Applied Mathematics and Computer Science research. Several of these workshop reports[10] reflect input from hundreds of researchers; each report distills to compelling and vital research program designs. Each of these workshops should be "dusted off" and programs pursuing them should be initiated.

These traditional research questions are no less important than other high-profile topics of the day (quantum computing and AI/ML). The answer to "what comes after Exascale" is not just quantum computing. In fact, and this is important, traditional applied mathematics and computer science research are essential to progress in those high-profile fields (e.g., traditional compiler optimization for HPC has much to contribute to optimization of AI/ML).

But applied mathematics and computer science research for HPC are of course vital for more than those high-profile fields. They are vital for scientific computing and its central role in U.S. economic vitality, energy security, the environment, and national security. For the U.S. to maintain an international edge, in the post-exascale era, Moore's law will not save us. This research, the responsibility of ASCR, cannot continue to be neglected.

[10] Computational Materials Science and Chemistry: Accelerating Discovery and Innovation through Simulation-Based Engineering and Science (2010); Applied Mathematics Research for Exascale Computing (2014, https://doi.org/10.2172/1149042); STREAM2016: Streaming Requirements, Experience, Applications and Middleware Workshop (2016, https://doi.org/10.2172/1344785); ASCR Report on Quantum Computing for Science (2015, https://doi.org/10.13140/RG.2.1.3656.5200); DOE Network 2025: Network Research Problems and Challenges for DOE Scientists. Workshop Report (2016, https://doi.org/10.2172/1367529); The future of scientific workflows (2017, https://doi.org/10.1177/1094342017704893); ASCR Workshop on In Situ Data Management: Enabling Scientific Discovery from Diverse Data Sources (2019, https://doi.org/10.2172/1493245); Extreme Heterogeneity 2018 - Productive Computational Science in the Era of Extreme Heterogeneity: Report for DOE ASCR Workshop on Extreme Heterogeneity (2018, https://doi.org/10.2172/1473756); Basic Research Needs for Microelectronics: Report of the Office of Science Workshop on Basic Research Needs for Microelectronics, October 23 – 25, 2018, (https://doi.org/10.2172/1616249); Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence (2019, https://doi.org/10.2172/1478744); Report of the HPC Correctness Summit, January 25-26, 2017, Washington, DC (2017, https://doi.org/10.2172/1470989); Future High Performance Computing Capabilities: Summary Report of the Advanced Scientific Computing Advisory Committee (ASCAC) Subcommittee (2019, https://doi.org/10.2172/1570693)

Finding B.2: ASCR's base research program has been constrained during the ECP era

The funding for math and computer science research has suffered during the exascale project. Figure 2 shows that the overall ASCR budget has grown over the past 7 years. However, the breakout for research, illustrated in Figures 3 and 4, shows reduced math and computer science research funding (at approximately half their pre-ECP levels) while pre-ECP funding and the budget for facilities to prepare for and acquire pre-exascale and exascale machines grew. Initially, this was due to decisions to allocate the math and computer science research budget toward the ECI, prior to the budget appropriations for the ECP construction project. Even in FY20, while the math and computer science research budget has increased, it is still substantially below the levels of FY14-15.

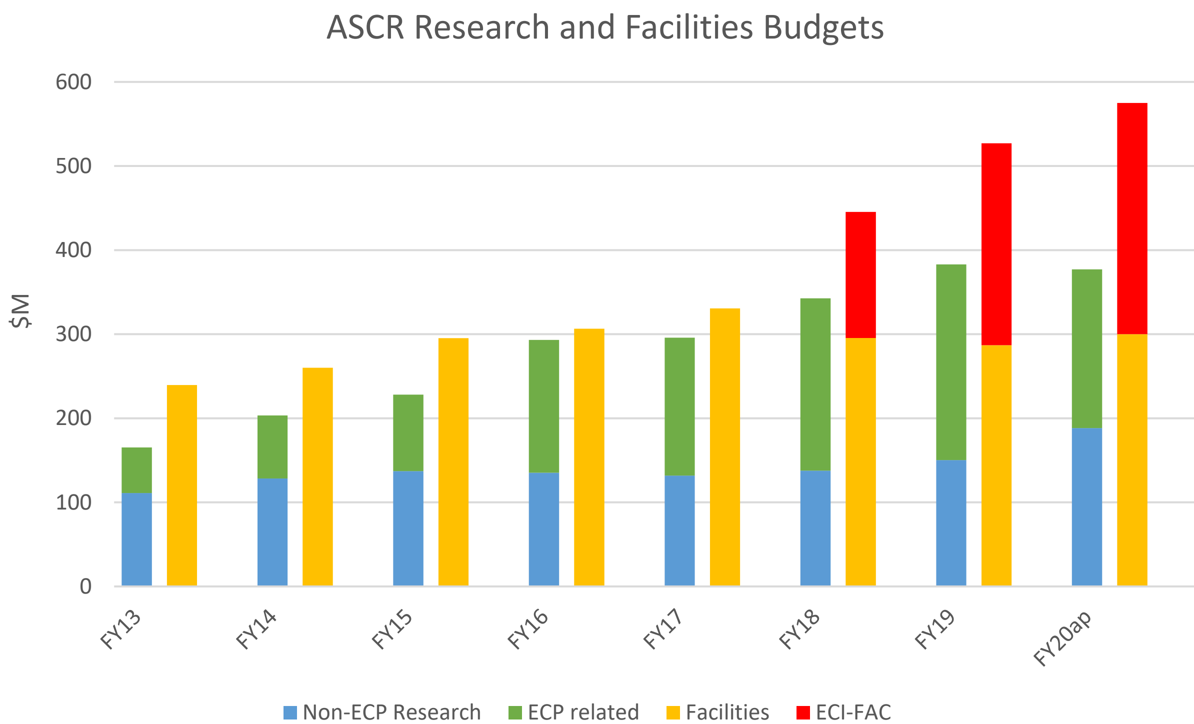

Figure 2. *Overall ASCR budget, excluding SBIR but including ECP and ECI-Facilities spending. (Source: DOE Office of Science, ASCR)*

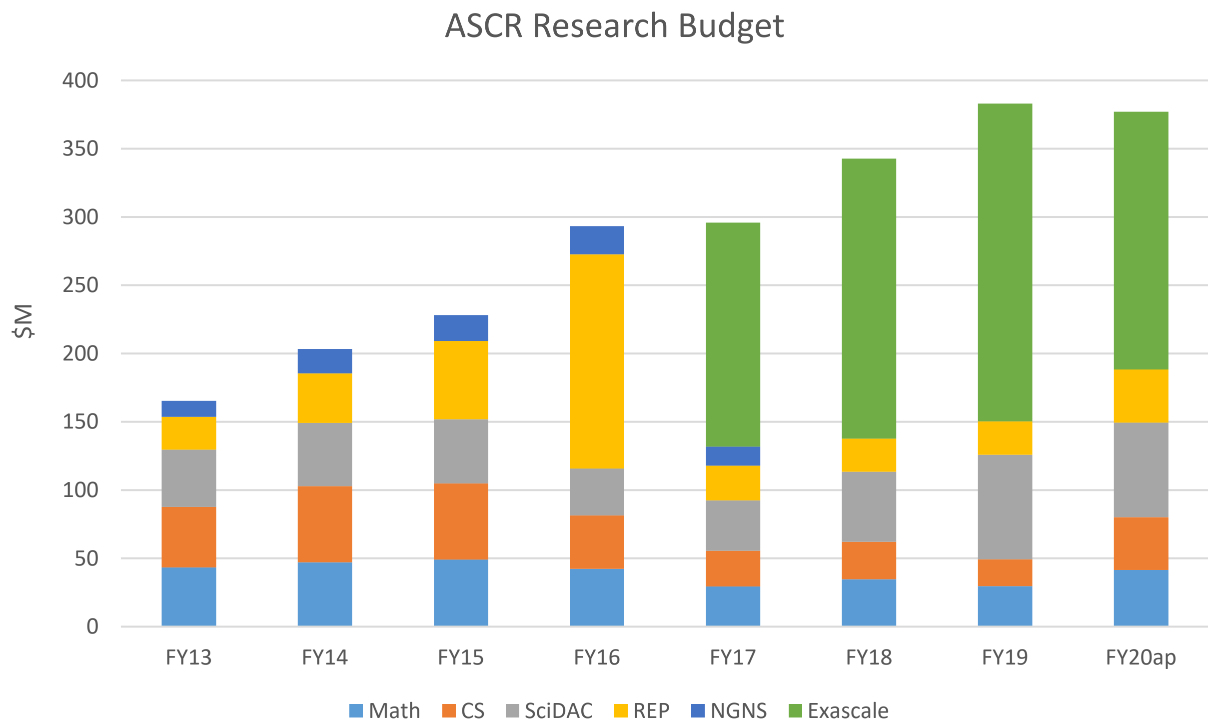

Figure 3. *Overall ASCR Budget FY13-FY20ap. Indicates that while the overall ASCR budget has grown in recent years, the funding for Math and Computer Science research has dropped. (Source: DOE Office of Science, ASCR).*

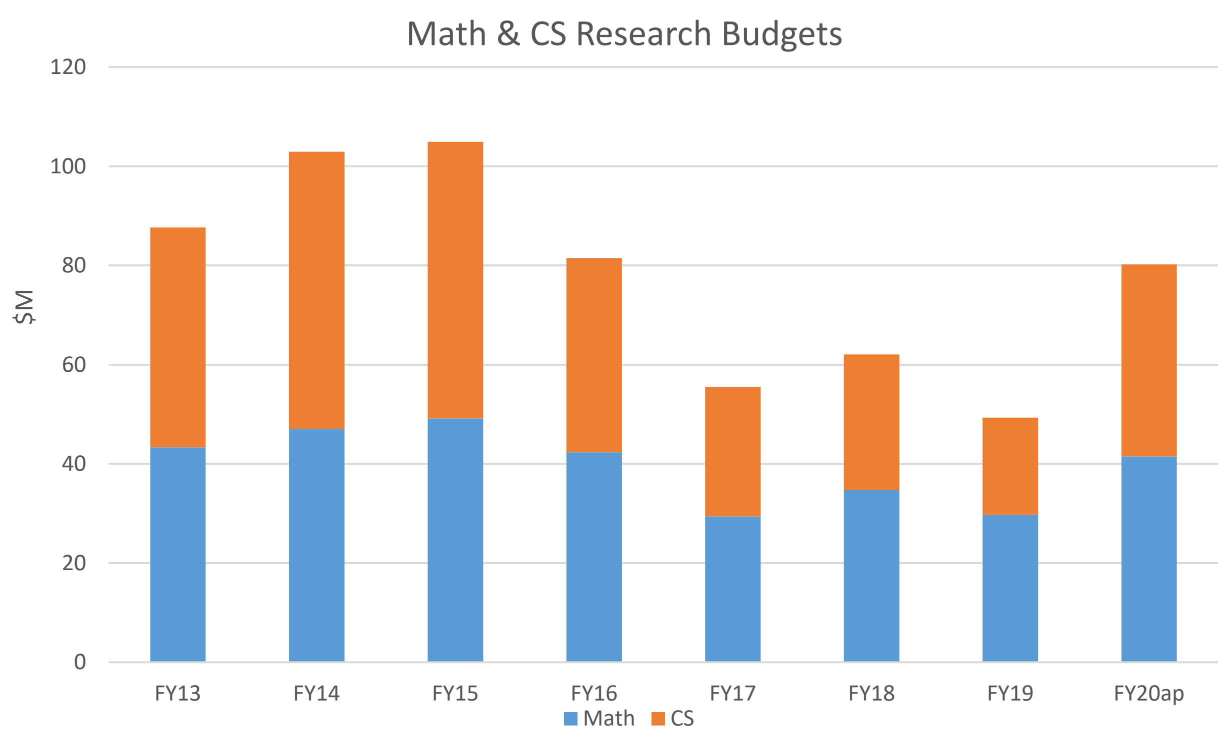

Figure 4. *Detail on Math and CS Research FY13-FY20ap. (Source: DOE Office of Science, ASCR).*

The change in funding levels has led to a decrease in the impact of ASCR on the fields of math and computer science. In our interviews, leading academics commented that the impact of ECP itself on math and computer science has been limited, as ECP has focused specifically on narrow development toward the ECP deployment, stacks, and applications under tight time frames. This has led to a decrease in fundamental research.

An aspect of this shift has also been to concentrate ASCR funding into the national laboratories, at the expense of academics, private research institutions, industry, and small business. The ECP does include a few such partners, but the vast majority of ECP funding has gone to the labs. As a result, the non-lab researchers have had to move to other funding sources, or completely out of research. The ripple effect is also on the talent pipeline, as academics can no longer train new graduate student research on ASCR areas of interest.

Even within the laboratories, in some areas, research in math and computer science has completely stopped, as the lab staff has moved over to ECP. This is taxing our national pool of people with broad, deep, understanding of math and computer science – forcing them to chase "flavor of the year" research topics. Such researchers are now being forced to "skill up" to make their research relevant to the AI and Quantum flavors.

Within the program committee of the annual Supercomputing conference, there is a general sense that the quality and breadth of research submissions has diminished.

Innovative projects such as Green Flash, a special purpose architecture for climate simulation, lost their base; unable to find private sources of funding, the project stopped (only to have the area picked up by European researchers). Reinvigorating ASCR research is essential for the effective evolution of the field of high performance computing and maintaining US leadership.

Recommendation B.1: Substantially reinvest in ASCR research

We recommend a significant expansion of the ASCR research investments in computer science and applied math.

In our committee's calls with members of the community, many topics in Applied Mathematics and Computer Science for HPC were identified as meriting additional research. The researchers we interviewed and met with put forward many specific topics and questions for research in computer science and mathematics that are itemized below. In some cases these re-iterated themes that have appeared in earlier ASCAC workshop reports (see footnote 10).

In general, the community was consistent in saying that high complexity, high reward topics represent the best research investments.

Computer Science

> **Computer architecture** post ECP is not just a supercomputer with a quantum computer bolted on. Post ECP, improvements will come from specialization, complex heterogeneous nodes, and new devices. An example of earlier and promising architecture specialization research to revive is Green Flash; this architecture would enable scaling the

precision of climate simulations to sub-km resolution, which is essential for water resource modeling and management. There is a need for deep co-design research, applications with architecture, all the way down to materials. These new architectures will have a breadth of implications for computer science, including algorithms and numerical methods that can take advantage of these new devices. Execution models also need study; the demands for power reduction and the adoption of near threshold computing will lead to the need to identify greater levels of concurrency and the ability to manage and load balance it dynamically, leading toward greater use of task-based dataflow execution models. (Dataflow execution models also have impetus from growing use of task based and asynchronous solvers). There is much potential from research in ways of rapidly generating HPC domain-specific architectures as systems on chip, building on new EDA tools from DARPA (e.g., chiplets) and associated software stacks.

Balanced Computer Architectures. There is a class of algorithms that presently is hard to change to make effective use of GPU accelerated architectures and some memory-starved CPU architectures, specifically applications that have a relatively low flops to data ratio. An example application is weather codes with complex atmospheric physics and data assimilation; even if the dynamical core can be made GPU friendly by the use of higher order methods, the vertical column physics are much less amenable to this. Other example applications are combustion codes in which preserving the physical correctness of the solution is important and extreme physics codes with very severe shock waves and complex chemistry. Manufacturing product design codes with very complex geometries are also typically facing this bandwidth wall. More generally it is applications using low-to-medium order methods for solving partial differential equations (PDEs) based on finite difference, finite element, and spectral methods. Iterative linear solvers based on conjugate gradient also suffer. In many cases these suffer from having roughly one flop per word accessed. An example architecture that is more balanced is the Fugaku (post-K) architecture, which is ARM based with vector extensions. The core chip has ⅓ of the flops of a Summit GPU but is served by a memory bus that is just as fast. For a fixed flop count, the ARM chip is thus a factor of 3 faster than Summit, of course the machine also has to be 3X as large for a given maximum flop level. Co-design research on balanced computer architectures, aimed at these types of applications and architectures, is needed. While flops may matter for applications: what matters for applications is time-to-insight; the full balanced ensemble of architectures, algorithm, coding, Uncertainty Quantification (UQ), data curation, etc.

Performance portability. As algorithms and code need to be customized for new architectures, extreme heterogeneity, and specialized hardware, how can the code be expressed in a way that enables automation and high-productivity human-guided mapping (placement, schedule, and resource management) from one architecture to the next, and from machine generation to generation? Research into new forms of compilers based on new geometric and new optimization algorithms will be required to achieve this. Performance portability can involve both adaptation to novel architectures as well as other abstractions such as libraries (whose adaptation might be automated) and languages that better support flexible implementation of key abstractions.

Related to this, there is a need for new performance modeling and prediction research that can encompass anticipated new algorithms and highly heterogeneous architectures.

Data management/workflows. How do we do more automated resource management? How do we manage data files across complex storage and memory hierarchies, now that we have NVM DIMM slots? How does the global management of data across facilities incorporate these new technologies? How do we facilitate data management, analytics and workflows? How can we do such data management in-situ, i.e., while the application is running? How do we expand the use of on-demand and cloud-based computing into the workflows? How do we manage authentication, access control, labeling, and provenance of data in large global research platforms? The community noted that the challenge of archiving and labeling data will grow with new demands from AI and Deep Learning, particularly with well curated very large "data" achieving a strategic role in the international AI race.

There is need for the **development and application of ontologies** for structuring and applying scientific and mathematical knowledge - ideally leveraging standards for ontologies such as W3C OWL, for provenance (e.g., The Underlay), and for integrating different ontologies and associated knowledge systems and formally reasoning within and across them (e.g., for science, for mathematics). Such systems can provide the knowledge foundation for hybrid symbolic (e.g., physics-model based) and connectionist (and other data-driven) AI/ML.

Research is needed for the challenge of **Distributed Computing and Data Ecosystem**: cross-facility/cloud-federated HPC, storage, and high-speed networking for implementation of scientific computing workflows. DOE research should adopt and advance the technology coming from the NSF FABRICs program and the international Global Research Platform (GRP) initiatives. There are substantial systems challenges in debugging and security when workflows cross administrative domains.

Edge computing. With the data explosion at facilities, and AI, there is a need for smart detectors and associated high performance *embedded* computing at the edge. This feeds into data management and advanced networking, e.g., real time distributed computing from detector directly to a facility for processing. Such research can also exploit opportunities coming from 5G based distributed sensing: e.g., distributed energy generation, storage, and utilization, and smart cities.

Complexity, Correctness, and Reproducibility. As the complexity of application, algorithms, systems, and hardware grows, new tools will be needed. Debugging systems at scale will go to a new level. But the complexity and resulting cost will make debugging less and less feasible; the state space of combinations will also make verifying software and components empirically (through extensive testing) incomplete, unreliable, impractical, and costly. New techniques for formally verifying and certifying the modules in our systems will be needed; ones that can address questions about how modules work together in a system, including formal methods for proving performance prediction, operation within bounds, stability, and numerical properties (accuracy/precision), so that systems work "straight off," including running with expected performance.

Enabling the use of mainstream languages and libraries, particularly emerging ones. Python, Scipy, Numpy, Matplotlib, Pandas, Anaconda, Julia, Ruby, and Go are examples of the languages preferred (with good reason) by the new generation of programmers, and very broadly used in industry and in communities such as those funded by NSF. Programming and interactive frameworks like Jupyter are increasingly the way in which these languages are used. And for AI, frameworks such as PyTorch, Caffe, TensorFlow and interchange systems such as ONNX are the rule.

Security Research. There is a need for research in making distributed systems and operating systems secure. Hardware bugs such as Spectre and Meltdown impact HPC "disproportionately." As the HPC community increasingly adopts containers and orchestration tools such as Kubernetes as a way to manage distributed resources; there is a need to do this securely.

Productivity research. It was suggested that researchers use science to study how people produce and use research software, to identify ways to increase productivity in scientific computing.

Networking. Continued research in networking is needed, especially in light of the expected rate of increase in experimental data, and data used for AI. A goal can be to "transfer one petabyte in one hour." This will require innovations through all levels of the networking stack (and perhaps new forms of networking stacks) from devices through systems, including transmission, routing, buffering, and flow control protocols.

Mathematics

Better models and algorithms trump brute force computing. ECP applied mathematics investments in linear algebra, PDE solvers, optimization, multi-scale simulation, and uncertainty quantification focused primarily on scalability of algorithms for the anticipated exascale architectures. This has resulted in substantial performance gains for certain applications. Equally important, these solvers are now packaged for broad distribution. These algorithms and architectures will bring unheard-of computational power to bear on many challenging problems. However, there is a tension and balance between architectures, algorithms, and modeling.

Now that Moore's law has ended, *new mathematical models and algorithms* is probably the most important area for ASCR's applied mathematics research investment to move beyond the end of the traditional exponential increase in computing performance. It is also important to highlight that one class of applications, and a very substantial class of problems, have a relatively low flops-to-data ratio – many have roughly one flop per word. For such problems it is hard to make effective use of the extreme floating point capability of today's exascale architectures. There are a number of promising areas for research in models and algorithms. The *Applied Mathematics Research for Exascale Computing* report (March 2014) authored by Dongarra, Hittinger, et al provides a comprehensive discussion on the value and need for investments in mathematical modeling. Much of this needed applied mathematics research fell outside the scope of ECP.

We recommend that ASCR revisit this report and focus on remaining and new challenges in mathematical modeling. In addition, we identify some specific ideas here:

Fast or low complexity algorithms. Capabilities of solving existing problems by simply using fewer flops would open doors in all areas of computational science, reducing the need for need for new hardware. We can do 10^{18} flops on an exascale platform. If we can bring an $O(N^3)$ algorithm to $O(N)$, that is a gain of 10^{12}, and if we can bring an $O(N^2)$ algorithm to $O(N)$, that's a gain of 10^9 in performance. New combinatorial methods, for example, allow inverting a sparse linear system in near-linear time, versus near-cubic time. There are also interesting research challenges in bringing these types of algorithms to new parallel and heterogeneous computing systems, e.g., applying 10^{18} flops to a low complexity algorithm. This is complex because the data structures, communication patterns, and computational patterns for low complexity algorithms differ substantially from traditional linear algebra that has been computational science's mainstay to date.

Sparsity. This area encompasses every aspect of sparsity – sparse solutions and sparse models – not just sparse linear algebra, for example the sparse multidimensional FFT and fast compressed sensing. Another aspect of sparsity is randomized linear algebra for low-rank approximations. Can we reduce the complexity of computing (making it sublinear) if what we're looking for in an answer is ultimately sparse? Can we (rapidly) find a "basis" or a "computational domain" where representing/computing a solution takes fewer resources in the spirit of reduced-order models? There are also interesting challenges here about "pre-computations," for instance, computing an optimal basis for a reduced-order model on the fly. This, of course, is not limited to scientific computing but applies to signal processing, data science, and computational science at large. There are a range of new models for reduced complexity (e.g., low-stretch spanning trees) that can be efficiently computed; perhaps there are new metrics to be discovered and applied to other classes of problems.

Non-convex optimization. The field of AI is littered with non-convex optimization problems: training of neural networks, approximations of high-dimensional probability densities (e.g., POMDP), Maximum Likelihood Estimators (e.g., Bayesian Networks). But advances in non-convex optimization go far beyond AI. Many computational problems can be formulated as nonconvex optimization problems. Examples include the solution to nonlinear PDEs in scientific computing, the computation of ground states in physics and molecular dynamics, the traveling salesperson problem in logistics and scheduling, graph algorithms, optimal control, optimal design, eigenvalue problems, and solving systems of nonlinear algebraic equations. It may be that it is always possible to reformulate a computational problem in a nonconvex optimization framework; this may not lead to the most efficient approach, but the question is worth investigating.

Quadrature and the efficient representation of functions/solutions. The goal here is to effectively reduce the size of the problem. For instance, performing the discretization of an integral equation that uses half the points of a traditional mesh reduces the cost by a factor of at least 50%. Similarly, if a complex function (high-dimensional probability distribution, high-dimensional interpolants) can be represented more efficiently, then one can increase performance as well. Two subcategories of research are: (a) Optimal discretization and optimal bases: this category includes Galerkin but it is broader, such as generalization of Gaussian quadratures to higher dimension (still an open problem), the idea is to discretize the whole domain at once (global quadrature) rather than using heuristic mesh (local quadrature) in the weak formulation. (b) Hierarchical representations: the goal is to use a (generally nonlinear) hierarchy to reduce the complexity of an approximation and make it computationally tractable, especially in high dimension. For example, can the meta-optimization of neural network structure be automated numerically? Is it possible to construct sparse representations of networks and data that maintain the critical essence of a problem, and can this be done for general problem sets?

Preconditioning in general. Preconditioners traditionally exploit problem structure to reduce the complexity of iterative numerical linear algebra, but they should not be considered solely as the provenance of solving large systems of linear equations. Problems involving nonconvex optimization of large neural nets could potentially have high-dimensional spaces preconditioned (or sparsified / regularized) in a way to make stochastic gradient descent (SGD) or other non-convex optimization faster. Better quadrature will result in better preconditioners as well; better hierarchical bases will allow for "smarter" preconditioning; and, low-complexity algorithms will be able to accelerate and take advantage of the benefits of preconditioning. Preconditioners that are smart – exploiting the underlying problem in a hierarchical sense and exploiting structural sparsity – adaptive, and can be constructed on the fly (in a problem-specific fashion), will have a large impact on the types and sizes of problems that computational scientists will be able to tackle.

New algorithms and numerical methods for new devices, e.g., forms of "physical" computing where the unique device characteristics (both quantum and classical) can be exploited directly for great computational efficiency. Such architectures could be based on using light or molecules. The architectures might be neuromorphic (e.g. reservoir computing) or other "analog" or "quantum analog."

New forms of PDE solving, e.g., when there are patches of data, approaching the PDE solving as a manifold completion problem, and in particular, incorporating knowledge of the physics in such data-oriented solving and data-driven model reduction.

There will be a substantial amount of research to take these new low-complexity algorithms to advanced architectures, but there is also a substantial amount of research and development to take old low-complexity algorithms (e.g., FFT, FMM) to new architectures.

There is substantial work to be done bringing **sparse linear algebra for classical methods** on finite elements to accelerators. Sparse linear algebra packages do not have adequate penetration because they are generally not optimized for architectures with accelerators. It is possible that new computer architectures, optimized for sparse linear algebraic approaches to graph computing, can facilitate more efficient use of sparse method libraries for scientific computing.

FFT is "perfect" for distributed systems, but the communication patterns (and intensity) make it hard to utilize accelerators. Low-communication FFT could efficiently use accelerators. A new challenge is developing implementations of sparse FFT.

The **Discontinuous Galerkin** method has been around for years. Accelerated computing can be extremely beneficial because it involves numerous small matrix operations. The research challenges lie in making sure that such methods respect underlying physical and geometric constraints, such as positivity and boundedness, and apply to problems with complex and evolving geometries. By building on this it becomes possible to obtain the advantages of high order with the same elapsed time as with low order methods on GPU architectures that are well-suited to the high computational complexity of high order methods.

Specialized solvers for Schrodinger Equation are particularly needed for applications coming from the DOE's Nuclear Physics and High Energy Physics programs. In general, there is an intense need for new technologies to support HEP data processing, where new experiments coming online will lead to orders-of-magnitude increases in data production.

In addition, there is an important class of problems with relatively low flops to data ratio that will not execute efficiently on an exascale architecture. Examples of such applications are weather codes with complex atmospheric physics and data assimilation, combustion in which preserving the physical correctness of the solution is important, and extreme physics codes with very severe shock waves and complex chemistry. For these applications and many others in the DOE landscape, the low flops to data ratio is a function of the underlying algorithm and not of the way it is coded. The algorithms are accurate and appropriate for the problem, the issue is the architecture. Understanding whether it is possible to have architectures that can help with these problems is the challenge. One possible architecture is the Fugaku (post K) architecture which is ARM based with Vector extensions (SVE). The question for the future is whether or not it is possible to design and build computer architectures through co-design that would address an even greater number of DOE applications in a way that complements the planned GPU machines with their ability to solve problems with high floating point requirements.

Traditionally, computer science and mathematics have been considered separate topics. However, post-Moore's law, the separation will be harder to sustain. Architecture will be harder to abstract, and the interaction between layers will increase. The ASCR research portfolio will need to be restructured to better integrate mathematics and computer science.

Recommendation B.2: Renew a stable environment for basic research

In order to create a stable environment that nurtures long term research, we recommend restoring the research budget of ASCR (as noted in Finding B.2 above, has substantially dropped during the ECP years), including funding for basic research in high performance computing. Such basic research fuels scientific innovation and will go directly and deeper after problems than was possible in the context of ECP. Several research staff at the national labs indicated that they are looking forward to "[getting] back to research" post ECP.

While AI and Quantum can be a part of restoring that research, those topics should not "suck all of the oxygen" out of other topics. There must be a balance that allows substantial research independent of AI and Quantum. This research should particularly expand beyond the national laboratories to include other entities, such as universities, to bring in fresh ideas. This also has the benefit of fueling training of the next generation of scientists. SciDAC and SciDAC models, with long term collaborations across laboratories and academic institutions, focusing deeply on innovative topics in mathematics, are considered ideal and could be applied to more than just mathematics, a sort of "SciDAC++."

The research program should re-establish collaborative, interdisciplinary/multidisciplinary research. Architecture will be harder to abstract, and the interaction between layers will increase. The choice of the mathematics – the algorithm – will be strongly influenced by the architecture. This drives the need for the ASCR research portfolio to be restructured to better integrate mathematics and computer science.

"Blue sky" research can support large breakthroughs. Such freedom to innovate is a strategic advantage to the U.S. Opportunities to do blue sky research at early career levels can attract new talent to the lab. Early career scientists will have the ability to do work that they might not otherwise be able to do in other institutions. Providing this opportunity at all levels can help retain talent. The Lab-Directed Research and Development (LDRD) programs are extremely competitive and not sufficient as evidenced by the fact of the labs having lost HPC talent because of the inability to support early career / innovative researchers.

A predictable cadence of FOAs for research is also essential. Dry spells disrupt progress and negatively affect retention. Bursts of FOAs create problems by forcing all staff to simultaneously write many proposals. As ECP ends, it is particularly important to regularly dovetail in new FOAs, easing the transition of staff from ECP to research.

Partnerships with other SC programs can also be done as basic research, e.g., ASCR working with BES programs in microelectronics, ideally in a long-term manner.

Recommendation B.3: Distribute research software

Building on some of the ECP experience with "hardening" research software for wide reliable use, we recommend that ASCR create pathways to wider distribution and uptake of research results that make it to the threshold of distribution. This should be an ongoing continuous effort within the research programs. DOE policy already directs the national labs to distribute the results of their research, subject to appropriate security and export considerations. Such distribution should be continued; open source can be one of the channels.

Open source does not by itself necessarily lead to uptake or sustainable "business models." It is also not clear that taking on the role of productization or customer support (vs. doing basic research) is the best use of the DOE research laboratory talent. A range of transition, sustainment, productization, and commercialization (e.g., SBIR) approaches should be pursued to move DOE research software products into use. These approaches can also mediate security and export considerations. Generous, but still limited or targeted distribution, can steer the assets created to our collaborators and not towards our competitors, to achieve national security and national economic advantage for the U.S. investment.

The ECP provides a model for strong transition-oriented research. The ECP's close coordination and collaboration among different lab software products to enable integration and interoperability, the intentional choice to avoid duplication of effort, centralized repositories, knowledge management, uniform build and test approaches, documentation standards, coding standards, and use of agile methodologies, maximize the return on ASCR investment by producing high quality research software. While this cannot be the only way that software is produced (rapid prototyping can produce innovation as well), these structures for development and release should be preserved.

C: Current and Future Workforce

The strength and vitality of ASCR research are defined by the enthusiasm, engagement, and creativity of the talented workforce that pursues it. This workforce is essential to maintain our lead in advanced computing.

ASCAC's 2014 study on the ASCR workforce[11] highlighted the "recruitment and retention challenges" in computer science expertise relevant to ASCR, the limited pipeline of new talent coming from universities, and the intense competition with industry for this talent. That report particularly brought forward recommendations related to the pipeline, including expanding the CSGF (Computational Science Graduate Fellowship) program, laboratory collaboration and recruiting on campuses, and strategic cross-laboratory and interagency efforts to collect data, coordinate recruiting, and particularly work with the dimension of diversity and inclusion as ways to improve the workforce.

This report, on ECP transition, reiterates those findings and goes beyond the earlier 2014 report. This report focuses on: 1) managing the discontinuity associated with the end of the ECP, 2) the importance of robustly funding academic research in ASCR disciplines to ensure there is a pipeline of talent with fresh ideas, and 3) the changes to the research agenda, mission communication, and workplace culture that ASCR and laboratories can undertake to make ASCR programs careers competitive against the high salaries that industry can offer. From postdoctoral researchers hired into Facilities, to L3/L4 leads, ECP team members are looking ahead to ask, "What's next?"

Finding C.1: ASCR has a skilled and motivated workforce

During ECP there have been stellar examples of great work. For example, ECP has built ties and trust between the DOE/SC/ASCR communities. This was enabled in large part through collaborative delivery of high-quality results. There have also been extraordinary efforts at the Facilities, Centers, and Labs.

However, there have also been challenges to the workforce pipeline at multiple stages.

Finding C.2: Retention, diversity, and opportunities are challenged beyond exascale

Addressing the workforce issues of retention, diversity, and opportunities for innovation will be critical during the transition to the post-ECP environment.

Several outstanding issues face ASCR in the future. As ECP winds down, the career paths for both researchers and developers supported by ASCR and the labs are uncertain.

[11] DOE Advanced Scientific Advisory Committee (ASCAC): Workforce Subcommittee Letter. https://doi.org/10.2172/1222711

Because of the uncertain transition from ECP to the next era for advanced computing, it is painfully obvious that retaining the talented workforce developed and nurtured during ECP will be problematic.

Likewise, the DOE is faced with the lack of diversity of the Lab workforce. According to current statistics,[12] the labs have about 20% women and 8-10% underrepresented minorities in technical research and leadership positions.

As mentioned above, there are opportunities for blue-sky research and development[13]. Making this a reality will be challenging for ASCR.

Finding C.3: Diversity, equity, and inclusion are valued

Diversity, equity, and inclusion questions were raised at every discussion between subcommittee members and external groups. Researchers uniformly expressed strong support for DEI and DEI initiatives.[14] Lab staff highlighted specific efforts at their labs and their engagement in them, and expressed their support for further expanding DEI efforts at the Labs (and the Nation's workforce more generally). "According to the Taulbee data, in 2014 women comprise a low and declining percentage of computing graduates, with 17.2% of Computer Science and 18% of all computing doctorates. Less than 2% of computational science doctorates are awarded to Hispanic or African-American students" [ASCAC Workforce Letter].[15]

Significant progress has been made during ECP and with SC leadership taking an active role supporting DEI. One workshop participant, a faculty member not funded under ECP, noted that in their experience, with an admittedly small sample of diverse students, the Labs have been very supportive of diversity – and in fact, more supportive than the broader academic community. SC/ASCR and the Labs should be proud of this. The DOE's Computational Science Graduate Fellowship (CSGF) program also received high marks.

Further illustrating ECP members' commitment to DEI, multiple participants shared news reports and studies that they had read recently on the subject, describing what they took away from those pieces and the implications for the Labs. Workshop participants noted that achieving DEI means increasing participation in activities like the workshops themselves. The workshops included robust discussions about DEI challenges. In the first workshop, researchers raised numerous opportunities to support DEI by fostering more inclusive culture and policies at the labs and highlighted the need for active support in recruiting diverse applicants. In the second workshop, participants discussed challenges that exist across all career stages, not just in the early-talent pipeline. Challenges in the early-talent pipeline are significant and well-known, motivating a variety of efforts to meet them, but it is also true that at all career stages, members of under-represented groups have fewer opportunities for advancement.

[12] Diversity and Inclusion, The National Laboratories. https://nationallabs.org/staff/diversity
[13] The ECP Management Plan v1.1, September 2019.
[14] David Rock and Heidi Grant. Why Diverse Teams Are Smarter. November 4, 2016. Harvard Business Review. https://hbr.org/2016/11/why-diverse-teams-are-smarter
[15] DOE Advanced Scientific Advisory Committee (ASCAC): Workforce Subcommittee Letter. https://doi.org/10.2172/1222711

The subcommittee was encouraged to see participants reinforcing positively and engaging with each other to improve their awareness and appreciation as a group of the scope and nature of DEI challenges and opportunities. Researchers at the Labs see multiple opportunities to improve DEI, which can be realized with appropriate support from leadership.

Recommendation C.1: Support researchers' re-engagement in "blue sky research"

In order for ASCR to sustain and grow its position of leadership in the range of disciplines it stewards, its research workforce must be supported to deliver fundamental, "blue sky" research through the transition and beyond – the seed corn for the steps beyond Exascale. The subcommittee recommends that ASCR capitalize on ECP's successes by crafting programs that will develop the diverse, multi-generational workforce as researchers shift focus back to basic/fundamental problems, while simultaneously sustaining ECP's positive outcomes such as improved software practices, as well as collaboration and trust across the community.

Many of ECP's successes and benefits accrue to ECP's significant differences from traditional expectations of a research program, many of which are associated with the requirements and deliverables necessitated by ECP's 413b project structure. At the same time, the subcommittee heard a strong consensus that meeting future ASCR needs will require that early career researchers (ECRs) gain substantial experience doing "big R" research outside of this kind of structure. The recommendations here reflect that consensus about meeting future needs.

Computing leadership at the Labs, the CRLC (Computational Leadership Research Council), formed as ECP began supporting ECP particularly as ASCR worked on developing the long-term plan for the workforce. The CRLC has been running workshops for early career researchers in applied math, and intend similar workshops for computer science. These workshops both create community amongst the ECRs and also develop the shared understanding of the most important problems in these areas. For these reasons, the subcommittee encourages ASCR to support these and similar activities.

Such road mapping and engagement-building activities are important, but they are not sufficient. For ECRs to build successful careers at Labs and in academia, ECP Program Directors and ECRs who focus on concrete deliverables need to be supported to transition to independent research and positions focused on independent research. From computer science and architecture to applied math and statistics, researchers across different fields highlighted the need for their field's ECRs to be supported in learning to do truly independent research. Although the individual Labs do have LDRD programs that support some ECRs, creating some of this type of opportunity, the programs are appropriately competitive and not of a scale suitable to address ECP transition needs. Researchers voiced concern how an inability to obtain research support is resulting in the loss of talented early career HPC researchers to academia and industry.

Attention to engaging directly with researchers at all levels about career paths is important. Early career researchers will benefit from mentorship in building independent blue sky research programs, and support in building their professional networks for fulfilling research careers. Mid-career and senior researchers can expand their horizons and motivate their continued engagement during the transition.

Recommendation C.2: Retain the current workforce

The committee heard loud and clear from leadership and researchers that there are concerns about what funding and programs will be available after the ECP ends, and the need for ASCR to put forward the plan for engaging laboratory talent after ECP ends. At the beginning of this study, the ECP budget plan had been for a slow decrease in funding to the end of the program. Recently, ECP announced a change in plans and ASCR is supporting sustained funding through the end of the program in 2024. However, this alone does not answer the question for researchers, but only defers it slightly.

Industry provides ample opportunities for researchers with the kind of computer science skills possessed by ASCR researchers. Demand for high performance computing, networking, algorithms, and mathematics is high across industries. A strong entry level graduate in computer science will regularly receive a compensation package of salary, bonus, and stock from a hyperscale internet company exceeding $300,000 per year; a principal engineer or research scientist's compensation can go to seven figures. These compensation rates are many times the amount this talent can earn at the national labs. The demand is driven by the impact that high performance computing has on the profitability and capabilities of these companies, in serving insatiable clickflows, data analytics, and particularly AI computing demands.

For entities that cannot compete for this talent with money, the strategy for successfully competing is offering a differentiated mission and culture, which ASCR can certainly do, particularly with mission: an opportunity to develop the essential tools for advancing science. Talent will be and is motivated by the opportunity to develop tools to understand our universe, our biology, the mind, our climate, and to develop new technologies applying that understanding. Attracting and retaining talented people requires that they feel valued as professionals, connected with their colleagues, engaged in contributing to the science mission goals of ASCR and DOE, and supported in their pursuit of career development opportunities.

It is important to retain the ECP workforce to ensure that it is available to support and utilize the new exascale machines when they arrive, and through their use. Evaluation of the KPPs (Key Performance Parameters) during the last six months of the project constitute a large burst of work, essential for overall success, at a time when staff may be seeking additional funding and also many researchers will be writing the final publications on ECP. The subcommittee strongly recommends providing a form of coordinated support that addresses ECP members' post-ECP needs and concerns, enabling them to focus on continued strong collaboration and delivery through the project end date.

Second, some elements of ECP are expected to trail off earlier than others, before the end of ECP – in particular, the hardware evaluation element. HI (Hardware and Integration) has collaborated with the hardware vendors on the design and evaluation of ECP hardware. The staff on this project will need to be redirected before the end of the program, perhaps toward next generation architecture research. Furthermore, in line with Recommendations A.4 and D.1, the capability and workforce hardware vendors are to a significant degree part of the ECP workforce and should be considered as such when planning retention strategies.

ASCR should articulate a vision for its exciting future that allows researchers to appreciate the opportunities and contribute effectively during the transition and beyond. Many researchers, especially those hired under ECP, will look to ASCR leadership for guidance on the future of advanced scientific computing efforts and their place in it. These scientists will seek stability and opportunities for professional growth. ASCR and the DOE need to define how their expertise can fit into the overall lab environment and future development.

Value as Professionals

A primary consideration for retaining the workforce through the ECP transition is that attracting and retaining talented people requires that they feel valued as professionals. The subcommittee recommends that ASCR take intentional steps to ensure that the workforce sees ASCR as a responsible steward of its intellectual resources, both during the transition and beyond. The transition is an important opportunity to reaffirm ASCR's commitment to responsible stewardship. The 413 project structure that scaffolded ECP is designed for projects in which tasks can be performed by any of a large number of professionals with the requisite competencies, an assumption which leads naturally to a particular kind of structure and processes, and thus work culture. By contrast, for tasks in ECP or in areas of future interest to ASCR, there are only a small number of professionals, who are highly sought after by other organizations. Workshop participants noted that Applied Math research has done comparatively better in this area, but the overall picture is that ASCR would benefit from better understanding workforce challenges, especially motivations for researchers who exit ECP or the Labs and their experiences of doing so. Leadership personnel, at many levels, share concerns about being responsible stewards of its talented workforce; the question is how to align these concerns and how to respond.

Stewarding takes multiple forms, especially in the larger and broader ecosystem framing that forms the basis for Recommendation D.1. The imperative to ensure the ASCR workforce (including the Labs as well as partners across academia and industry) feels valued as professionals arises in part from the deeper and more varied relationships that members of the ASCR workforce have across different parts of the ecosystem. For instance, ECP researchers have established deep and trusted relationships with communities that support key infrastructure for DOE/ASCR interests; abrupt or unjustified departures can and do rupture relationships with those organizations and communities, jeopardizing delivery on the ECP mission or others. In some areas, one cannot simply "turn on the spigot" to make progress, but rather must have existing relationships where trust and understanding of shared goals are already available. The subcommittee encourages ASCR to take a holistic view of the relationships ECP members have with external partners of all kinds.

Work policies and culture also contribute significantly to feeling valued as professionals, especially where such policies and culture lag, behind industry, that support work-life balance and the diversity of modern family life and structures. As one workshop participant commented, "DOE … is still talking about work-from-home, there are things where DOE has to self-reflect." Such sentiments were expressed concurrently with the understanding that DOE/ASCR mission goals necessitate careful assessment and design around issues such as work-from-home policies, but the overarching notion was that ASCR could do more and would benefit from doing more. The subcommittee encourages ASCR to harmonize its work policies to support workforce retention and future recruitment efforts.

A key aspect of a research retention policy is to encourage solid editorial policies related to authorship with respect to research produced with ASCR funding. Authorship is a key aspect of scientists building a research career. As it is a key incentive for building and retaining a workforce, authorship should be carefully apportioned – assuring appropriate credit to those who do the work. One model is the Science Magazine Authorship policy in which authorship on papers reflects contribution, responsibility, personal accountability, and integrity. ASCR program managers may wish to consider current lab policies and encourage formation and strengthening of them.

Connectedness and shared purpose

According to workshop participants staff retention will require nurturing interconnectedness, trust, and shared understanding; one participant citing Google's study of the five key factors in determining team success[16]. One of the greatest successes of ECP has been building relationships for coordination and collaboration. The subcommittee recommends that ASCR identify and support existing initiatives to build interconnectedness and trust across the ECP community, and sponsor new initiatives where gaps exist. For example, the CRLC has sponsored a workshop for mid-career researchers to build inter-lab relationships and identify shared directions of interest; CRLC intends to hold future workshops of this type for mathematics and for computer science. Such efforts should be supported where possible, including similar offerings for researchers at different stages of their careers. A senior faculty member involved in ECP stressed the importance of such activities, noting that early career researchers in academia are supported in finding future positions through invitations to give talks and to attend meetings, leading to introductions to potential colleagues. Because ECP postdocs may not have the same window into their communities it is important to identify the mechanisms in ECP that support this. The transition's change in focus and research interests, combined with DOE's unique capabilities and experience with large, interdisciplinary team efforts, furnish additional motivation for ASCR to increase support for relationship building across Labs and disciplines.

Engaged contributions to the science mission

Attracting and retaining talented people also requires that they feel engaged in contributing to the science mission goals of ASCR and the DOE. The subcommittee therefore recommends that ASCR take steps to assure that the ECP workforce is as engaged with these goals as possible. Where uncertainties cannot be addressed, the workforce needs to be confident that ASCR will take such steps when possible. The DOE mission goals, from next-generation clean energy to national security, are important factors in attracting and motivating the workforce, especially where the problems' scale and complexity cannot be matched anywhere else. "Exceptional scientists," said one senior PI involved in ECP, "are motivated by opportunities to work on important scientific problems – interdisciplinary ones, [and] exascale machines are just a tool for that … not just scientific problems, societal problems." Part of what has drawn people to ECP has been the scale and complexity of the challenge, and part has been the scope of possible impacts across the Nation and the world; but as the project comes to its end, people look for the next challenges, shifting

[16] Rozovsky J. The Five Keys to a Successful Google Team. re:Work. November 17, 2015. https://rework.withgoogle.com/blog/five-keys-to-a-successful-google-team/

focus – and employment – accordingly. The subcommittee encourages ASCR to identify science challenges enabled by ECP, that will engage the workforce for the transition and beyond, and to communicate the problems' importance and impact.

Funding structure and a stable funding base play important roles in maintaining researchers' engagement. ECP has been a stable funding base that enabled researchers to focus on the work rather than writing numerous proposals. However, researcher's engagement with the work has been impacted because of ECP's intense focus on coordinated delivery, making it impossible to follow their curiosity on new developments. One Lab member stated that summer interns express the same concern from an early career perspective: when considering career paths, research independence at the Labs is a major draw and counterbalances, to some degree, the drastic compensation differences seen in industry. Technical leaders in ECP have been able to redirect effort away from work that was not progressing, allowing more researchers to spend greater effort on directions making significant progress. An industry participant highlighted such capacity to "turn on a dime" as a significant positive at their workplace.

Source of career development opportunity

Attracting and retaining talented people also requires that they feel supported in their pursuit of career-development opportunities. As noted above, numerous participants addressed the importance of being responsible stewards of the intellectual talents and careers of ECP researchers. The subcommittee recommends that ASCR expand support in two areas of career development during the transition: career paths and mentorship. ECP has enabled numerous researchers to take on new kinds of roles, and their experiences in ECP have had powerful effects on their professional capabilities and interests. Some researchers have found they would like to continue developing managerial capabilities, while others appreciate the experience and wish to return to research. Industry participants highlighted that most technology companies have recognized the value of defining senior individual-contributor roles as an alternative to management. Lab staff highlighted the discrepancy as an important opportunity for the Labs to improve retention and culture: one commented, "Defining … exciting career paths that lead to retention still needs work."

A second area for supporting career-development opportunities arises in mentorship. All of the Labs now have postdoctoral associations, in part to increase the quantity and quality of mentorship for postdoctoral researchers. For example, Livermore's postdoctoral program is supported by an independent set of mentors who ensure that postdocs have access to the resources they need, and early conversations about their career paths are seen as helpful. Such programs require care in design, particularly for non-US citizens, and discussions about career goals and development need to be clearly separate from merit review. Senior Lab researchers suggested that such mentorship programs should be expanded to include researchers at all career stages. Furthermore, numerous workshop participants indicated that the experiences of Lab interns can be improved significantly through better mentorship, which highlights the multiple benefits that would accrue to investments in mentorship programs and training at all levels.

A sense of belonging post-ECP

Before and during the transition, many staff and researchers – especially those hired under ECP, and who do not have experience with DOE/ASCR outside of ECP – will look to the leadership of ASCR and ECP for indications about the future and their place in it. The subcommittee recognizes that in large organizations, major changes such as the ECP transition are often associated with some number of departures. However, as discussed throughout this report, career transitions take place within the overall ecosystem of computing professionals across industry-academia-government, and although this ecosystem continues to grow quickly, individuals within the ecosystem are communicating more, and with more parts of the ecosystem. The subcommittee therefore recommends that ASCR take steps to ensure, after the transition, that all project members understand that their contributions were valued and feel that they were afforded respect and transparency throughout.

For instance, one positive impact of ECP was that it led to the hiring of people with a diverse range of expertise, including many who would not have necessarily come to the Labs pre-ECP. Sustaining and growing ECP's benefits, and realizing the potential of ECI as a whole, will require strategic consideration of the paths of these diverse individuals, many of whom by design differ from the "big R" research focus pre-ECP. Retaining these individuals will be challenging in the absence of clear communication from leadership. As the Labs' expertise with the ECP software stack (in both research and Facilities) is vital to continuing to enable science throughout the Exascale era, the subcommittee recommends ASCR work to ensure that key enablers decide to continue supporting the DOE/ASCR mission. More senior Lab staff, especially those who took on significantly new roles to support ECP, face a different kind of challenge as the ECP transition approaches: deciding whether to return to some version of their previous role, to continue in the new direction, or to do something else entirely. Sabbaticals, perhaps in modified form, may offer a valuable mechanism for ASCR to support its science mission by offering these individuals opportunities to understand how their interests align with the new landscape.

The subcommittee heard vigorous discussion about mechanisms that might be useful to support retention and engagement through the transition. The subcommittee encourages ASCR to explore these mechanisms further. Corporations frequently end projects prematurely if they are non-performant or if priorities have changed. In some companies, redeployment experts (brought in well ahead of the project end date) work on behalf of project personnel to find suitable follow-on efforts – enabling team members to continue focusing on the project work until its actual termination, without uncertainty. There is precedent at DOE, as APS has hired an external organization to help place individuals with 6-12 months' notice. Furthermore, in projects, there can be a "toolkit" of human resource options for providing additional funding beyond salary to key individuals. This capacity has been written into ECP but not implemented. The mechanisms are not without their challenges for the ECP transition, however. For instance, coordinating these mechanisms fairly may be complicated because in some Labs different members of a single group may be funded by different sub projects of ECP; however, redeployment consultations do happen in large, matrixed corporations. There were additional concerns about whether retention efforts that occur at the ECP level could interfere with traditional, lab-level retention efforts, undercutting the positive inter-Lab cooperation and collaboration that the subcommittee found to be one of ECP's most universally valued outcomes.

Articulating a vision

In the business community, studies of large-scale organizational change emphasize the challenge of communicating the new vision and highlight that such communications are frequently drastically under resourced[17]. The subcommittee recommends that ASCR move swiftly to develop, articulate, and disseminate a new vision for the transition and beyond. As an example of the challenge: A workshop participant noted that multiple ECRs had left their group at the Lab during the transition into ECP, because ECRs have not yet experienced how uncertainties like that period resolve, and communication from senior leadership about the processes and likely outcomes is more appropriate than encouragement from their managers and mentors to stay despite uncertainty. Staff are uneasy about post-ECP funding, and the uncertainty impacts focus and therefore productivity. Reassuring staff that post-ECP will be exciting and important and stable is important. Fortunately, the scale of ECP, its successes, and the connectedness of the ECP project team provide ASCR with unprecedented capabilities to "energize the next generation," as "early career [researchers] see that they are a part of something big."

During the subcommittee's examination of the transition, the plans for the ECP related funding profile has evolved in a positive way. Senior ECP leadership shared that they believe a number of issues have been resolved, and that PIs have been encouraged to seek approval to pursue good ideas through the end of ECP. We encourage ASCR to reinforce this wherever possible and appropriate, and to coordinate with ECP leadership. The subcommittee believes that these kinds of efforts are important to ensuring delivery of the final ECP milestones, which occur during the final months. Key staff and leaders will seek to ensure stability for themselves, their teams, and other researchers for whom they are responsible, and therefore it is important that these leaders understand how ASCR's plans and actions support that.

The subcommittee's review of the ECP transition has coincided with growing excitement around AI and quantum computing, two computing technologies with revolutionary potential for science and DOE/ASCR missions. The subcommittee encourages ASCR to ensure that the science mission – which distinguishes ASCR from academia and industry, along with the long time frames – remains the central pillar of ASCR's vision, and that the science drivers in the vision are ones that support attracting and retaining top talent. The science mission is especially valuable as a unique factor because early career professionals see many examples where computing and data are used for private gain. The DOE mission, properly articulated, resonates deeply as a source of national well-being and security through ethical support of economic prosperity.

Recommendation C.3: Strengthen ties to universities and the ecosystem

Realizing the full potential of ASCR's contributions to advanced scientific computing over the coming decades will require a strong, diverse pipeline of talented and well-trained professionals. In workforce development, as in other areas already described, the world of high-performance and scientific computing has grown dramatically in size and complexity over the past few decades. Consequently, here as elsewhere, the subcommittee sees "ecosystem shaping" as the

[17] Kotter, J.P. Leading Change. Harvard Business Review Press, Boston, 1996

optimal framework for achieving cost-effective leadership in the fields stewarded by ASCR (Recommendation D.1).

Predictable and flexible funding

The subcommittee recommends that ASCR work to ensure that funding opportunities arrive on a predictable, regular basis and be administered predictably. Funding opportunities in recent years have arrived unpredictably and intermittently. Especially outside of the Labs, researchers whose expertise can benefit DOE have alternative sources of research funding and are frequently balancing multiple workstreams (e.g. the academic calendar). Under such circumstances, it is very possible for excellent researchers to miss an FOA or to be unable to provide the kind of responsive proposal that would be reviewed favorably and selected for funding. Therefore, to maximize the quantity and quality of truly responsive proposals from faculty and industry, as emphasized in Recommendation B.2, ASCR should, wherever possible, aim to achieve a regular, predictable cadence for FOAs and due dates. Such a cadence also supports DEI effort, as it helps ensure that all interested parties are on equal footing with respect to awareness and time for preparation. Furthermore, due to growth in the disciplines stewarded by ASCR, ASCR's relationships with the communities carry additional weight, and therefore argue for increased attentiveness to predictability. For instance, due to decreased support for applied math research at the universities, ASCR's efforts to re-invest in these relationships will benefit from predictability in the form of increased trust and confidence. Administration of awards also needs to be conducted uniformly because, for instance, faculty incur specific costs at different times during the calendar year, such as paying tuition for graduate students.

The subcommittee also recommends that ASCR work to increase funding flexibility to increase agility and research productivity. One of ECP's strengths has been the ability to quickly move funding support within institutions as appropriate. In contrast, however, one of the weaknesses has been the much longer delays associated with moving funds between institutions (these delays can stretch to months). This is especially challenging for industry partners, and future ASCR activities would be enhanced greatly by addressing these delays; this challenge also affects university researchers funding graduate students, as noted above. As one participant said, "The mechanism of getting money to the people is something we need to fix within the DOE."

Interaction and human development

A leading university researcher (and PI on DOE awards) commented that historically, "one of DOE's great strengths has been interaction and human-capital development across the DOE/university interface." This researcher highlighted that during ECP, ASCR's model appears to have shifted away from the traditional model in which programs had two solicitations, one for Labs and one for universities, with the latter targeting smaller awards funding a single PI. In recent years, and especially with ECP's focus and project requirements, the model has shifted towards large, multi-institutional collaborations which have led to the perception that the role of the Labs is much larger proportionally, and the role of universities greatly reduced. The subcommittee recommends ASCR provide strong support for individual university awards at the single PI level. Such support will strengthen the pipeline by increasing the pool of possible faculty applicants, increasing opportunities for DEI, and increasing the diversity of ideas and approaches.

The perception graduate students have of the vitality and strength of the DOE ASCR research program is critical to attracting new talent to the DOE workforce. ECP's significant and successful investments in "moving the needle" on best practices, coordination, and deployment cannot yield their full dividends over the coming decades unless the next generation of researchers aligns to the DOE mission and ASCR research programs. An important motivation for substantially reinvesting in ASCR's core fundamental research, and for creating a stable environment for basic research, is that DOE needs to be visible in important and emerging areas and exciting early career researchers through leading papers and participation in conferences. Investments in graduate students, through a variety of programs, provide important ways to support increasing DEI in the DOE workforce.

Ties to universities and beyond

Universities and industry partners have significantly greater flexibility in beginning new projects quickly, and in changing directions swiftly. Strengthening ties to universities and the broader system will increase ASCR's ability to capitalize on this agility, which will be important as technological evolutions occur more quickly in the era of extreme heterogeneity. Faculty are able to change direction significantly with each new student, and similarly industry is able to pivot nimbly as their research and development teams are already operating at professional levels. Research at the Labs naturally focuses on work on the "critical path" for the DOE mission, and this should remain the case even as additional resources support increased agility at the Labs.

To gain insights into opportunities and challenges in engaging universities regarding HPC education, the subcommittee spoke with alumni of the CSGF program, including current faculty members. The subcommittee heard about the need to increase DEI, and about the distribution of computational skills in introductory HPC courses, even at the graduate level. Frequently, students in such courses arrive with only moderate experiences in computing and in non-HPC interpreted languages such as Java and Matlab. Outside of computer science, only the most exceptional graduate students arrive with experience beyond the fundamentals in compiled languages, software engineering, and parallelism. Furthermore, although there is an emerging loose-knit community of faculty who teach introductory HPC, the faculty typically have to find their own way to that community. Relatively modest investments by ASCR to nucleate this community through ECP educational materials could yield large impacts, especially if coupled to support for the use of leadership facilities in courses. Such investments could also yield increased alignment among ASCR and the faculty who train the workforce. Participants highlighted that in many departments tenure and promotion committees focus on the top journals for that field, e.g. Nature, and that the challenges and serious contributions required to deliver high-performance simulations at DOE leadership facilities are not yet as widely recognized as they need to be.

For many reasons, it is of national interest that students are as prepared as possible when they enter the workforce, whenever and wherever that might be. The rise of HPC, including on-chip parallelism and accelerators, creates important opportunities and motivations for ASCR to support strengthening the university talent pipeline, especially at the earliest stages. This support is especially important in the Exascale era, as HPC can support and drive national competitiveness in many industries. Many university curricula and degree programs have not kept pace with changes in architectures, and at many academic institutions there is a dearth of parallel/high-performance computing courses; this includes both computer science (CS) departments and non-CS departments. This is a long-standing challenge, and the ECP transition is an opportune moment to address it, especially considering the scale and quality of educational materials prepared as part of ECP. Furthermore, there are some universities that do contain significant HPC curricula, including for example those with NNSA's PSAAP (Predictive Science Academic Alliance Program). The subcommittee recommends that as ASCR reinvests in relationships with universities, it take steps to capture best practices and lessons learned from these universities, and work to export these across the broader academic community.

Reduce barriers

The ECP ST (software technology) products significantly reduce barriers to entry for researchers to begin engaging on work in HPC. Compared to ECP, previous generations of ASCR supercomputing have not had the kind of software stack that would scale up on different machines. Furthermore, if ASCR decides to provide guidance for other SC ADs (Associate Directors) or third parties regarding hardware standardization, then university HPC facilities may be able to cost-effectively obtain platforms allowing their faculty to do DOE-relevant research with confidence in the performance at Leadership facilities.

Where the steps above can be taken, they support students' early engagement with DOE HPC facilities, improving both workforce training and support, delivering maximum scientific value from ECI (not just ECP), by broadening participation in INCITE (Innovative and Novel Computational Impact on Theory and Experiment) and at the LCFs (Leadership Computing Facilities) and NERSC. Support university recruitment of students in applied math, engineering, and key HPC CS (OS research, programming models); this is important from a national competitiveness standpoint, not just for the DOE mission. These activities are also opportunities to increase awareness of, and enthusiasm for, the DOE mission space. The subcommittee heard numerous comments about the importance of increasing awareness and exciting early career researchers around the DOE mission. A senior faculty member noted that students understand references to quantum computing and machine learning, but not to Exascale, due to the small number of classes and the relatively small cohort of experts in Exascale.

Funding for graduate student training is necessary but not sufficient. That is, funding can ensure that PhD students in strategically important areas graduate with the necessary skills, but funding alone cannot ensure that the new graduates build their careers at Labs or work as DOE partners in industry or academia. Multiple project participants stressed the importance of growing internships and related programs (e.g., PSAAP, where students spend significant time working at a Lab alongside a PI or team). It is also not enough to ensure that students have experiences at the labs; the experiences must be positive ones. Multiple participants highlighted the importance of improving mentorship and engagement for Lab internship programs. Uninspiring internships exacerbate recruitment imbalances compared to industry, especially as such experiences percolate further in the increasingly networked ecosystem of early career professionals. Investing in Lab mentorship skills, and in aligning Labs around metrics for internships, are viewed as important goals. Internship experiences create meaningful opportunities for students to appreciate the Labs' culture and mission, especially around team-driven, interdisciplinary problems. From the ecosystem perspective discussed in Recommendation D.1, DOE can complement these efforts by building long-term relationships with open-source software communities for key technologies, as talented students often engage early in the open-source community.

Even as ASCR invests in strengthening the university pipeline along traditional HPC lines, it must also include research areas outside the main. For each new generation of HPC systems or architectures, there is some group of researchers for whom it is already obsolete. As one example, GPU-based HPC was a considerably active research area in the academic community before it became an active area in the DOE HPC community. Similarly, for some of today's students, the accelerators that are foundational to the Exascale era are not as exciting or promising as FPGAs, ASICs, or even neuromorphic devices. Maintaining situational awareness and agility requires engaging innovative students at all levels of computing research, from architecture to systems design, to systems software, to parallel algorithms and computing.

Recommendation C.4: Create career paths for scientific software professionals

The subcommittee heard from many communities that ECP's increased focus on development was a positive shift for their community. This included other offices within SC [AD meetings]. The subcommittee heard that it was not just the increased focus on development, but that *software was a funded mandate*, such that delivery of high-quality, interoperable software was an explicit, articulated metric of success.

Sustaining and growing the benefits of this focus will require increasing the ability to recruit and retain top talent in software development. Shortcomings in recruiting and retention appear as complementary challenges in the facilities and on the research side. Staff at facilities report that recruitment is not the primary challenge, but rather retention. Specifically, although they are able to find people who want to work on ASCR's bleeding-edge HPC platforms, the positions are not research positions in nature, and there is not a clear career path for most people hired into non-research positions, so they tend to leave after a few years. Conversely, research staff report skills gaps in recruiting, including working with GitLab, containers, and specific languages that underpin enabling technologies (e.g., Go underlies Singularity). In other words, from their organizations, finding researchers with these skills who want to come to the Labs is a challenge.

In contrast, private industry (those segments with which the subcommittee spoke and has familiarity) offers long-term career paths that address, or at least mitigate, these issues for recruitment and retention. For example, the subcommittee's meetings included employees of companies who hire computational subject-matter experts whose primary activities focus on software associated with specific applications or technologies, but who also remain engaged with their main scientific domain(s). The company's support for their doing so adds value to the company, sometimes directly, and other times indirectly (including simply by facilitating retention). Such flexibility provides advantages, including significantly the opportunity to provide sustainable, growth-oriented career paths for the talented individuals associated with ECP who possess complex, often non-traditional combinations of skills, expertise, and judgment.

Just as important as creating career paths focused on software is where these career paths sit within the research enterprise. As detailed in other sections, the subcommittee heard numerous robust discussions around the appropriate balance of support for the shared software, especially key infrastructure. One key consequence of the lack of clear ownership (whether distributed or singly held), is that no organization has the clear accountability to address the need to provide stable, fulfilling career options in this space. The subcommittee recommends that software-focused career paths sit with ASCR research rather than facilities and be shared with the software stewardship program depending on its structure. One motivation is that continuous delivery will counteract a perception created by ECP's duration and focus, which have led early career researchers (both PD and staff) to perceive ASCR as oriented towards projects and facilities. Furthermore, due to the continuing evolution of application needs and also the diversity of HPC platforms at different facilities, much of the necessary software talent should be resourced under the research teams and possibly to a lesser extent under the shared-software stewardship program. In contrast, within a facility, the natural focus is on the platforms within that facility. For early career researchers, becoming extremely focused on one piece of software, or one specific HPC facility, may lead to difficulty seeking academic positions. Industry approaches like those mentioned above provide an alternative with a more attractive balance. Being intentional about avoiding overspecialization, by situating software-focused individuals with research, helps ensure that developers continue to feel connected to their distributed teams and their larger communities with which they identify. Increasing connectedness increases satisfaction, which supports both productivity and retention. It has been noted that the "postdoc at facilities" model yielded about half staying, but half moving on to research roles. Hiring processes should be transparent to candidates specifically about a position's focus. Both recruitment and retention may benefit from increased transparency about the career paths, especially if there are software career paths that are clearly software development within the research community.

One workshop participant, a PI at a Lab, expressed that from their perspective, the creation of this kind of stable career path for software development talent is a bigger problem than the end of ECP. The PI noted that research grants do start and end, and that for PIs who lead groups/teams, the end of ECP is to some extent not that much out of the ordinary pattern, in that there's a known end date, and that the PIs are responsible for finding additional funding afterwards. This perspective also furnishes an additional perspective for funding these career paths under ASCR research.

Recommendation C.5: Support diversity, equity & inclusion (DEI)

A diverse workforce assures fresh perspectives and aspirations to achieve innovations targeted toward the breadth of national needs.

Everyone with whom the subcommittee spoke agreed on the importance of increasing diversity, equity, and inclusion in the DOE/SC/ASCR workforce. Researchers contributed numerous suggestions of opportunities to improve DEI. On a broader scale, the subcommittee's work coincides with a large-scale effort by SC to review DEI efforts across the SC labs, and the subcommittee recommends – supporting ambitious goals based on their findings. The ECP transition period, as ASCR re-affirms its commitment to national and international leadership in the wide range of disciplines and communities, offers a timely opportunity to simultaneously reaffirm its commitment to DEI. Results from the prior Workforce letter, and studies across the research disciplines stewarded by ASCR, establish that. Programs to engage diverse students in high school and younger in STEM careers are essential, and should be supported, along with broader efforts in recruitment and retention.

In recent years, industry has adapted its hiring practices significantly in pursuit of recruiting and retaining top talent, and the subcommittee recommends that SC/ASCR support comparable adaptations wherever feasible and appropriate. Researchers who spoke to the subcommittee noted particularly: support for employees who work remotely, flexibility in work hours for personal reasons, childcare options (on-site as well as support for finding it), and parental and family leave. Visible and consequential support from ASCR leadership on these aspects are especially important for DEI because members of under-represented groups are perceived differently with respect to majority groups when during hiring discussions[18,19,20]. Satellite offices, where possible, have also been helpful in attracting new diverse talent, and could be expanded.

ASCR should promote more family friendly work policies, mentoring, to specifically address women in the workforce and forms of diversity such as (but not limited to) ethnic diversity, LGBTQ, and people with varying abilities**.**

The subcommittee was encouraged to hear that at the Labs, many of these adaptations have been made already in isolated cases. Typically, a Lab staff member recognized the importance of recruiting a uniquely capable individual for the given position, and "went the extra mile" to find ways to make the Lab position work for them. AD support to establish local guidance will greatly reduce the activation barriers and support DEI significantly; members of under-represented groups and candidates with different needs. One striking example is that ECP ST has been highly successful, and ST lead Mike Heroux has worked remotely for Sandia for over 20 years.

---

[18] Counteracting Racial and Gender Bias in Job Negotiations. Harvard Law School Daily Blog, Program on Negotation. December 31, 2018. https://www.pon.harvard.edu/daily/leadership-skills-daily/counteracting-racial-and-gender-bias-in-job-negotiations-nb

[19] Bohnet I. What Works: Gender Equality by Design. Belknap Press, 2016

[20] Hernandez M., Avery D.R. Getting the Short End of the Stick: Racial Bias in Salary Negotiations. MIT Sloan Management Review. June 15, 2016. https://sloanreview.mit.edu/article/getting-the-short-end-of-the-stick-racial-bias-in-salary-negotiations/

D: National and International Leadership

As has happened throughout its history, DOE is pioneering the vision, design, creation, and application of the next incarnation of advanced scientific computing.

DOE's investments and vision for scientific computing continue to lead national and international efforts to advance the scale and impact of scientific computing. Finding D.1 highlights that ECP represents the latest chapter of DOE leadership in this space. At the same time, the rise of a large private-sector market for large-scale computing (as well as the proliferation of applications and the diversification of an international supply chain) necessitate a new strategy for maintaining leadership. For context, in 2019 the market for cloud computing alone was approximately $227 billion, and is projected to grow at approximately 17% per year for the next several years[21]. Consequently, even with large-scale research efforts such as ECI/ECP, vendors make rational decisions to focus their R&D spend on industry-driven needs and forecasts.

The subcommittee notes that the imbalance in R&D spending is not unique to DOE ASCR, but extends broadly across the US: in the mid-1960s, federal funding for R&D was approximately twice that of private-sector R&D, but by the mid-1990s the imbalance went the other way, and in recent years commercial R&D has exceeded federal funding by almost a factor of three. Furthermore, the coming decade promises to advance a variety of disruptive, transformational computing technologies from theory to practice: examples include neuromorphic computing, quantum computing, deep co-design, and artificial intelligence. Finding D.2 addresses the opportunities and challenges of these multiple competing innovations.

The importance of maintaining leadership, the context motivating a new strategy for doing so, and the subcommittee's perspective on opportunities in this increasingly complex space, form the basis for Recommendation D.1. Numerous other US government agencies and entities share interests and capabilities in emerging areas of computing technologies. Recommendation D.2 covers the subcommittee's suggestions for engaging these organizations to ensure timely and cost-effective delivery of the DOE/ASCR mission through the ECP transition and beyond.

Finding D.1: ECI/ECP is the leading national and international exascale computing effort

When envisioned, the goal of developing exascale computing was advanced in SC/ASCR and NNSA/ASC and supported to meet the needs of DOE. As plans to design, invest in, and create exascale were formed, the consideration expanded to a national effort involving multiple agencies and stakeholders. It also helped spawn exascale projects around the world.

[21] Gartner Forecasts Worldwide Public Cloud Revenue to Grow 17% in 2020. Gartner. November 13, 2019. https://www.gartner.com/en/newsroom/press-releases/2019-11-13-gartner-forecasts-worldwide-public-cloud-revenue-to-grow-17-percent-in-2020

Finding D.2: New horizons in computing will impact DOE's mission

These new horizons challenge the traditional operating structure of ASCR. The subcommittee heard repeatedly that ASCR research should be 5 – 10 years ahead of the facilities, and anticipating the needs of facilities. For over a decade, and perhaps since the rise of vector processing, ASCR has successfully leveraged this shared view to harness explosive growth in computing. We find that the excitement around these emerging technologies arises in part because of their disruptive potential, that is, there is a breakdown in the consensus of what the facilities will look like in 5 – 10 years. The consensus has rather stepped back to an understanding that many aspects are likely to change: as one researcher put it, when it comes to "[maintaining an] international edge in the post-exascale era, Moore's law won't save us."

It is important to view this emerging gap in context with ASCR's traditional strengths in basic research across a wide range of disciplines, and the workforce's enthusiasm for exploring new directions in basic research in the Exascale era. It is more important than ever that ASCR be aware and sensitive to the fact that some key technologies and breakthroughs will come from outside of DOE/ASCR's domains.

Recommendation D.1: DOE/ASCR should maintain national and international leadership in advanced computing

The recommendations we have made for the transition lay the foundation for ASCR research and ASCR technologies to impact future computing and DOE's mission. To maintain world leadership in scientific computing, DOE and ASCR need to be able to connect to stakeholders across US universities, industry, laboratories, and agencies.

An ecosystem-engagement strategy

A general challenge ASCR faces is that the rise of computing means that both industry and academia are making substantially greater investments in ASCR disciplines – yielding a large, rapidly growing number of research activities and subfields relevant to ASCR's mission. As noted in Finding A.3 (Ecosystem), ECP has successfully created its own ecosystem for Exascale computing, and begun to integrate it with the larger ecosystem. The ecosystem view is a new paradigm for ASCR research and delivery of HPC-enabled science mission goals. The present section focuses holistically on this framing for the subcommittee's recommendations for maintaining national and international leadership in advanced computing.

Noteworthy ASCR activities that extend impact and leadership.

ECP is one of multiple ASCR efforts in the last several years that provide noteworthy examples of the kinds of activities that the subcommittee encourages for the transition and beyond.

- For example: recognizing the need for distributed, multi-institutional teams to access data and computing resources seamlessly for complex workflows, ASCR together with the National Laboratory Research Computing Group supported studies of building a DOE/SC-crossing Distributed Computing and Data Ecosystem (DCDE), to gather input on needs from across the DOE/SC and assess gaps and capabilities [reference April2019 report].

- In addition, ASCR's recognition of the growing importance of innovation in microelectronics, including deep co-design, drove a BRN workshop that included participants across DOE, government, academia, and industry [ref microelectronics report].
- Within ECP, the PathForward partnership with hardware vendors has yielded significant benefits that accrue to its longevity, and as described in Recommendation A.4.
- ECP has also supported the software ecosystem through efforts to increase standardization, interoperability, and software best practices, especially across Labs, to reduce barriers to access.
- ECP's contributions to Jupyter project for security [ref] have not only helped enabled DOE researchers to use this popular and powerful tool, but also advanced security for the data science community around the world.
- As an example of leadership in a large ecosystem, the ASCR-BER collaboration IDEAS led to the adoption and dissemination of software policies [ref: IDEAS Report].

In addition to direct benefits to users (enabling, for instance, one application to experiment with different multigrid implementations, a test which would have been essentially impossible before ECP), organizations outside DOE are expressing interest, including NSF and in Japan. As we describe below, ECP support has also enabled contributions to standards bodies for LLVM compiler infrastructure and the C++ language.

Adopting an ecosystem view brings implications for both internal and external-facing activities. More can be accomplished than with a purely transactional approach, as SC has long recognized through cross-disciplinary partnerships such as SciDAC (in fact, to some degree many of the following comments amplify what we have heard as the benefits of SciDAC). Relationships, at both individual and organizational levels, are more important, as are indirect paths between nodes in the network. Building these relationships requires stable, durable funding, as in SciDAC, many ECP project teams, and as described below between ASCR-funded researchers and the LLVM community.

With the shift from compartmentalized communities, to integrated, empowered teams and ecosystems, more can be done with less resources. However achieving this in a sustained way requires foresight in developing the relationships before they are needed. Business literature supports an ecosystem-engagement approach as a "shaping strategy"[22] or one of "fostering generative relationships."[23]

In summary, ECP has created a successful ecosystem of tools and practices, capable of supporting increased productivity as well as facilitating integration with the much larger computing ecosystem outside DOE and the research community. Maintaining national and international leadership can be achieved by sustaining and building on these aspects of ECP.

---

[22] Hagel J III, Seely Brown J., Davison L. Shaping Strategy in a World of Constant Disruption. Harvard Business Review, October 2008 http://johnseelybrown.com/shapingstrategy.pdf

[23] Lane D., Maxfield R. Strategy under Complexity: Fostering Generative Relationships. Long Range Planning, 29(2):215-231. 1996. https://doi.org/10.1016/0024-6301(96)00011-8

Recommendation D.2: ASCR should engage and collaborate with national stakeholders in other agencies

As illustrated by ECP's Flang activities (see text box), engaging and contributing to large open-source project communities provides an effective and productive way to demonstrate leadership while raising awareness about ASCR's work. Similarly, standards bodies are often important in organizing multi-institutional community priorities, and DOE's ability to support standards that align with its roadmap and mission are valuable, as can be seen through ECP's contributions to the LLVM and C++ communities. As ASCR continues to move towards the ecosystem model of engagement, maintaining and growing relationships (and expectations) requires increased consideration, for instance in increasing the regularity and predictability of funding opportunity announcements (Recommendation C.3, strengthen ties to universities).

ECP and LLVM: an exemplar of delivering DOE mission by engaging the ecosystem as a partner

ECP's Flang project represents an exemplar of the ecosystem-engagement strategy that the subcommittee has described as a cost-effective means to maintain and grow national and international leadership. ECP Flang builds on a significant investment by NNSA's ASC (National Nuclear Security Administration, Advanced Simulation and Computing) to establish the Fortran front-end ("Flang") within the open-source LLVM compiler infrastructure. The ECP effort will ensure that the LLVM infrastructure supports critical features for Fortran performance on the accelerator-driven platforms up to and including Exascale machines. The LLVM community is large and multi-institutional, primarily composed of industry and academic partners, who together contribute a diverse range of expertise, needs, and requirements. Flang's integration with LLVM yields multiple high-impact outcomes for DOE, including for example making it cost-effective for vendors to continue providing Fortran compilers for their own hardware (by leveraging Flang). Such an impact would be prohibitively expensive for DOE/ASCR to achieve on is own (as it would require supporting the entire toolchain, not only the Fortran front-end).

Connecting to stakeholders and maximizing situational awareness

For ASCR to maximize the impact of its research budget, it should leverage the investments of the larger ecosystem wherever doing so makes sense. This requires both situational awareness and mutual understanding of how the associated communities' long-term priorities and commitments align with those of DOE/SC/ASCR. Without situational awareness, it will be difficult to avoid expensive and time-consuming reinventions of the wheel; without understanding how other communities may respond to the changing technological landscape, ASCR risks the long-term stability and predictability that are essential for many mission needs. Therefore, situational awareness and understanding partner communities are prerequisites for ASCR to deliver the fundamental computing sciences and technologies that DOE needs but cannot acquire – or is not likely to acquire – elsewhere. AI and ML represent an important case study: industry has made massive investments in AI/ML libraries, and it does not seem feasible or necessarily valuable for DOE to duplicate such efforts. However, for high-consequence applications DOE can and will remain a central sponsor. [24,25]

[24] Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence. https://doi.org/10.2172/1478744

[25] AI Charge Letter, October 2019.

Situational awareness is always an important dimension of stewarding a research program such as ASCR's. The challenge is that its importance and complexity have grown significantly faster than budgets, as have the costs of shortcomings in awareness. Because these trends are likely to continue, it is especially important that ASCR-supported activities to map landscapes and research needs, such as workshops, including (wherever possible) significant representation from the associated communities. Where participation is for whatever reasons constrained, it is increasingly essential that ASCR field alternate mechanisms for receiving the input from the groups unable to participate. For program managers and leadership, tracking rapidly evolving fields is challenging even for technical experts; achieving satisfactory awareness requires support for travel and meaningful engagement with both practitioners and stakeholders. The need for support is especially pronounced for program managers whose portfolios differ from their technical background. Our understanding is that such support has not increased, and instead has decreased, over the past decade. There are also other opportunities, including partial sponsorship of meetings and workshop series whose participants span the breadth of communities that contribute to research areas stewarded by ASCR.

The subcommittee heard a strong consensus, from early career researchers to senior leadership, that continued investments are needed to ensure continued delivery of DOE science in spaces where Fortran applications remain the dominant tools. A senior researcher highlighted the tension between ECP's impact and the need for substantially more investment by saying, "ECP has allowed us to accelerate development by an order of magnitude; we are replacing code that took 20 years to develop and we are going to do it in 5 years. But I do not [know] how we are going to keep that going." Another characterized the work done on Fortran applications as "surgical," that is, limited and strategic, rather than comprehensive as will be needed in the future.

Importantly, ASCR's ability to work closely with LLVM arises not from the ECP funding itself, but rather from the sustained, steady support of key Lab staff *prior* to ECP, that involved significant contributions to the LLVM community. This highlights that engaging the computing ecosystem can be a cost-effective approach to delivering essential capabilities, and that doing so requires supporting the valuable work of building trusted relationships, identifying opportunities where interests align, and pursuing those opportunities in ways that further improve the relationship.

Conclusions

Over its history, ASCR research has laid foundations for the world's most powerful and impactful scientific computers and DOE facilities have led in making them available for science and in support of DOE's mission. With ECP, DOE through ASCR and ASC built mightily on this tradition to develop and deploy transformational exascale systems.

In this report, we have considered what has been learned from the ECP experience and its impact on ASCR and DOE and investigated how this experience should inform ASCR as it moves forward in the post-ECP environment. The systems created by ECP open new avenues for ASCR activity, including a systematic approach to software sharing and making some research software reliable and broadly available. Some of the project management practices and development paradigms that were used in ECP can be applicable to ASCR. We must re-invigorate ASCR's research portfolio and relationships to rapid development of computing technologies and capabilities in the nation and around the world. We must value and retain the workforce which has been engaged in ECP and in the Labs during exascale development. Finally, ASCR must move forward with full awareness of the changing landscape and new opportunities for advanced computing around the nation and the world.

We hope this report will be useful in helping the Office of Science and ASCR develop the programs that will come during the transition. We have already seen ASCR take positive steps in preparation of the transition. We anticipate a bright future for ASCR in continuing to lead in scientific computing.

Acknowledgments

The work of the Transition Subcommittee would not have been possible without generous engagement and time commitments from the ECP leadership team and many members of the ECP community. We also thank the ASCR researchers who met with us as well as the many CSGF Fellows who shared their views of the future at the CSGF program review in 2019.

We especially thank the Office of Science AD's for sharing their perspectives on ECP and ASCR and the ASCR office for its coordination and support of the subcommittee's work.

Special thanks to Christine Chalk and Tiffani Conner for their support of the subcommittee.

Summary

For convenience, here is a summary of our findings and recommendations.

Finding A.1: ECP has been successful overall

ECP is a successful project in multiple dimensions including its primary objective of organizing large complex resources to design and develop exascale computing systems that will be deployed to satisfy DOE mission needs. At its conclusion, ECP will have created artifacts (evaluation systems, software libraries, demonstration applications) and adopted practices (in software engineering, project management, co-design, stakeholder collaboration) that should be expanded and built upon to realize the full impact of exascale computing.

Finding A.2: ECP successfully managed a large distributed collaboration

ECP's management practice contributed to the successes of the project. Adaptation of project management for research/development activities is challenging; ECP's lessons learned include ways to balance lightweight PM with supporting delivery by adding the right process at the right time.

Finding A.3: ECP created a software ecosystem focused on DOE science mission needs

ECP has created a well-designed software ecosystem for development, curation, and distribution of exascale systems and application software. This ecosystem integrates the fruits of years of basic research in: mathematics, computer science, applications, and systems software. In particular, the ecosystem greatly reduces barriers for ASCR fundamental research maturation and impactful delivery at the facilities and with users. ECP has also created significant amounts of educational and training materials, which have been important to the project's successes.

Finding A.4: ECP supported collaboration with facilities and industry

ASCR and ECP/ECI have effectively collaborated with industry and the facilities to develop exascale computing technology and industry applications.

Recommendation A.1: Create a shared-software stewardship program within ASCR

ASCR should create a comprehensive program that leverages the ECP ecosystem to support and curate shared software. This should incorporate ASCR program office oversight while delegating operational control to a software engineering team of laboratory and academic experts.

Recommendation A.2: Engage current, and anticipate future, software needs

Important software and algorithms can originate outside of DOE and ASCR. ASCR should continue to monitor and anticipate external developments in critical areas and incorporate this information in planning the evolution and modernization of software.

Recommendation A.3: Collaboratively support applications

ASCR should collaborate with other DOE offices and select outside entities to support development of key applications, especially those which continue to defy attempts to address them at the exascale level of computing performance and problems involving edge computing. Also there should be a pathway for deployment of "hardened" shared scientific software.

Recommendation A.4: Broaden industry and academic engagement

We recommend that the ECP collaboration models be extended as appropriate to hardware and independent software vendors to engage them early and substantively in new directions and that similar collaboration with university groups should be explored.

Recommendation A.5: Adopt modern project management practices

We recommend that ASCR adopt and incorporate modern project management practices and tools into its programs to facilitate collaborative work between labs and programs.

Finding B.1: Applied mathematics and computer science research is essential for future progress in advanced scientific computing

There are still many open research issues in traditional Applied Mathematics and Computer Science for HPC that are relevant for current and emerging computational science efforts.

Finding B.2: ASCR's base research program has been constrained during the ECP era

The funding for math and computer science research has suffered during the exascale project. Even in FY20, it is still substantially below the levels of FY14-15.

Recommendation B.1: Substantially reinvest in ASCR research

We recommend a significant expansion of the ASCR research investments in computer science and applied math in a series of areas important for the future.

Recommendation B.2: Renew a stable environment for basic research

In order to create a stable environment that nurtures long term research, we recommend restoring the research budget of ASCR, including funding for basic research in high performance computing. Collaborative Applied Mathematics and Architecture research and interdisciplinary research should be encouraged. Long term research should be encouraged.

Recommendation B.3: Distribute research software

Building on some of the ECP experience with "hardening" research software for wide reliable use, we recommend that ASCR create pathways to wider distribution and uptake of research results that make it to the threshold of distribution. This should be an ongoing continuous effort within the research programs.

Finding C.1: ASCR has a skilled and motivated workforce

ECP built ties and trust between the DOE/SC/ASCR communities. This was enabled in large part through collaborative delivery of high-quality results. There have also been extraordinary efforts at the Facilities, Centers, and Labs.

Finding C.2: Retention, diversity, and opportunities are challenged beyond exascale

Addressing the workforce issues of retention, diversity, and opportunities for innovation will be critical during the transition to the post-ECP environment. As ECP winds down, the career paths for both researchers and developers supported by ASCR and the labs are uncertain.

Finding C.3: Diversity, equity, and inclusion are valued

Diversity, equity, and inclusion questions were raised at every discussion between subcommittee members and external groups. Researchers uniformly expressed strong support for DEI and DEI initiatives.

Recommendation C.1: Support researchers' re-engagement in "blue sky research"

The ASCR workforce must be supported to deliver fundamental, "blue sky" research through the transition and beyond. The subcommittee recommends that ASCR capitalize on ECP's successes by crafting programs that will develop the diverse, multi-generational workforce as researchers shift focus back to basic/fundamental problems.

Recommendation C.2: Retain the current workforce

The committee heard loud and clear from leadership and researchers that there are concerns about what funding and programs will be available after the ECP ends, and the need for ASCR to put forward the plan for engaging laboratory talent after ECP ends.

Recommendation C.3: Strengthen ties to universities and the ecosystem

Realizing the full potential of ASCR's contributions to advanced scientific computing over the coming decades will require a strong, diverse pipeline of talented and well-trained professionals.

Recommendation C.4: Create career paths for scientific software professionals

The subcommittee heard from many communities that ECP's increased focus on development was a positive shift for their community. Sustaining and growing the benefits of this focus will require increasing the ability to recruit and retain top talent in software development.

Recommendation C.5: Support diversity, equity & inclusion (DEI)

A diverse workforce assures fresh perspectives and aspirations to achieve innovations targeted toward the breadth of national needs. We recommend that SC/ASCR support adaptations to attract and retain such a workforce. Visible and consequential support from ASCR leadership on these aspects are especially important for DEI because members of under-represented groups are perceived differently with respect to majority groups when during hiring discussions

Finding D.1: ECI/ECP is the leading national and international exascale computing effort

When envisioned, the goal of developing exascale computing was advanced in SC/ASCR and NNSA/ASC and supported to meet the needs of DOE. As plans to design, invest in, and create exascale were formed, the consideration expanded to a national effort involving multiple agencies and stakeholders. It also helped spawn exascale projects around the world.

Finding D.2: New horizons in computing will impact DOE's mission

The emergence of new disruptive computing paradigms and technologies presents a real challenge to ASCR's ability to anticipate and pioneer research in areas critical to the future needs of DOE facilities. It is more important than ever that ASCR be aware and sensitive to the fact that some key technologies and breakthroughs will come from outside of DOE/ASCR's domains.

Recommendation D.1: DOE/ASCR should maintain national and international leadership in advanced computing

The recommendations we have made for the transition lay the foundation for ASCR research and ASCR technologies to impact future computing and DOE's mission. To maintain world leadership in scientific computing, DOE and ASCR need to be able to connect to stakeholders across US universities, industry, laboratories, and agencies.

Recommendation D.2: ASCR should engage and collaborate with national stakeholders in other agencies

For ASCR to maximize the impact of its research budget, it should leverage the investments of the larger ecosystem wherever doing so makes sense. This requires both situational awareness and mutual understanding of how the associated communities' long-term priorities and commitments align with those of DOE/SC/ASCR.

Appendices

Transition Charge Letter to ASCAC

Department of Energy
Office of Science
Washington, DC 20585

September 6, 2018

Professor Daniel Reed
Chair, Advanced Scientific Computing Research
Advisory Committee
University of Utah
Salt Lake City, Utah 84112

Dear Professor Reed:

Thank you for your continued service to the Office of Science (SC) as the Chair of the Advanced Scientific Computing Research Advisory Committee (ASCAC). Your reports and recommendations continue to help us improve the management of the Advanced Scientific Computing Research (ASCR) program. ASCAC's studies and reports on the Exascale Computing Initiative's Conceptual Design (2015), Ten Technical Approaches to Address the Challenges of Exascale Computing (2013), DOE Data-intensive Science and Exascale (2012) and the Opportunities and Challenges of Exascale Computing (2007) were instrumental in developing the Department's Exascale Computing Initiative and identified critical research opportunities for applied mathematics, computer science, computational partnerships and advanced networking during the exascale era. The upcoming ASCAC report on the opportunities and challenges for future high performance computing capabilities should further identify areas for ASCR investments in the "Beyond Moore's Law" era.

I am writing to ask ASCAC to assemble a subcommittee to identify the key elements of the Exascale Computing Project (ECP) that need to be transitioned into ASCR's research program or other new SC/ASCR initiatives after the end of the project to address the opportunities and challenges for future high performance computing capabilities. The subcommittee should consider ECP lessons learned for managing large collaborations, ASCR's historic fundamental research investments in applied mathematics, computer science and computational partnerships at the National Labs, the Administration's new Research and Development priorities in artificial intelligence, quantum information systems and strategic computing.

As history has shown, basic research advances have been the bedrock of American innovation and prosperity. These advances often gave rise to new lines of scientific inquiry and led to inventions of new technologies and industries that transformed our society. Breakthrough discoveries emerging from Federal investment can have broader impacts beyond the original field of scope and have made Federal programs, such as ASCR, an essential part of the Nation's science and technology strategy.

By examining ECP, I expect this report to provide guiding strategies and approaches that will be key to ensuring future U.S. leadership, and more generally, U.S. leadership in the full range of disciplines stewarded by ASCR. With these high-level objectives in mind, the report should provide recommendations for capturing the lessons learned from ECP, supporting the software and hardware technologies and application development from ECP activities and informing ASCR's future investment strategy for its basic research programs.

I would appreciate receiving a written report by September 30, 2019.

Sincerely,

J. Stephen Binkley
Deputy Director for Science Programs

cc: Barbara Helland, SC-21
Christine Chalk, SC-22

Meetings of Transition Subcommittee

2019

Committee Calls

January 31
March 21
April 16
May 28
June 11
July 16
July 20
July 23
August 6
August 20
August 27
September 3
September 10
September 17
October 1
October 8
November 5
November 12
November 26
December 3
December 10
December 17
December 26
December 30

Meetings with other entities

March 26 – ASCAC
June 3 – ADs
June 25 – Math Discussion
July 2 – ST Discussion
July 9 – CR
July 11 – BESAC
July 15-16 – CSGF
July 29 – CRLC
July 30 – Gamblin
August 13 – Barb Helland
September 23 – ASCAC
October 15 – TH
October 17 – IC
November 15 – IC
November 19 – CASC
November 21 – SC19 – BoF
December 16 – CLRC

2020

Committee Calls

January 6
January 9
January 21

Meetings with other entities

January 13 – ASCAC
February 6 – ECP – AHM