Can the US Maintain its Leadership in High Performance Computing?

A report by the ASCAC Subcommittee on International Competitiveness

U.S. DEPARTMENT OF ENERGY | Office of Science

Outline

- Subcommittee Charge
- Team
- Executive summary
- Key Findings and Recommendations
- Discussion

Charge Letter

Professor Daniel A. Reed, Chair of the ASCAC
Senior Vice President for Academic Affairs
Professor of Computer Science and Electrical & Computer Engineering
The University of Utah
201 Presidents Circle, Room 205
Salt Lake City, Utah 84112-9007

Dear Professor Reed:

Thank you for your work as Committee Chair on the Advanced Scientific Computing Advisory Committee (ASCAC's) and for the ongoing review of the collaboration with the National Cancer Institute. The ASCAC recommendations will help us to improve the management of this important program.

As you know, the Administration and Congress have been keenly interested in the recent issues with the supply chain and U.S. competitiveness and innovation. Looking to the future, we want to ensure that the U.S. continues to be a leader in advanced computing, high end computational science and engineering, advanced scientific networks, and the fields and workforce that underpin these efforts.

To that end, we must develop and maintain world-leading capabilities in key technologies, especially microelectronics, high performance computer architectures and software, computer science, applied mathematics Artificial Intelligence, Quantum Information Science; and also provide compelling, inclusive, and equitable opportunities for all those who want to work in this fast paced and ever-changing area of research.

Therefore, I request that ASCAC develop a report to address the following questions:

- How can the Department maintain critical international cooperation in an increasingly competitive environment for both talent and resources? In areas where the U.S. is leading, how can we sustain our roles and attract the best industry and international partners? In other areas, how can the Department build and maintain its reputation as a "partner of choice"? In general, are there barriers that can hinder our ability to form effective and enduring international and industry partnerships?

- Identify key areas where the U.S. currently has, or could aspire to, leadership roles in advanced computing and high-end computational science and engineering, including unique or world-leading capabilities (i.e., advanced scientific facilities, testbeds and networks) or leading scientific and technical resources, such as highly trained personnel and supporting infrastructure. This may include emerging areas or opportunities that offer significant promise for leadership.

- To preserve and foster U.S. leadership roles within reasonable resource constraints, are there particular technical areas or capabilities that could be emphasized? Are there other technical resources and capabilities that could be leveraged in to achieve these goals, possibly through collaborations within and beyond the ASCR community?

- How can programs and facilities be structured to attract and retain talented people? What are the barriers to successfully advancing careers of scientific and technical personnel in advanced computing, computational science and engineering, and related fields and how can the Department address those barriers? A complete answer to these questions should address how we can ensure that we are recruiting, training, mentoring, and retaining the best talent from all over the world, including among traditionally underrepresented groups within the U.S.

We would appreciate receiving a written report by the Spring meeting in 2023.

If you or the subcommittee chair have any questions, please contact Christine Chalk, Designated Federal Official for ASCAC at 301-903-5152 or by e-mail at christine.chalk@science.doe.gov.

I appreciate ASCAC's willingness to undertake this important activity.

Sincerely,

JOHN BINKLEY
Digitally signed by JOHN BINKLEY
Date: 2022.03.22 19:30:31 -04'00'

J. Stephen Binkley
Acting Director
Office of Science

DOE Office of Science Charge to the Subcommittee

1. **How can the Department maintain critical international cooperation in an increasingly competitive environment for both talent and resources?**
   – In areas where the U.S. is leading, how can we sustain our roles and attract the best industry and international partners?
   – In other areas, how can the Department build and maintain its reputation as a "partner of choice"?
   – In general, are there barriers that can hinder our ability to form effective and enduring international and industry partnerships?
2. **Identify key areas where the U.S. currently has, or could aspire to, leadership roles in advanced computing and high-end computational science and engineering, including unique or world-leading capabilities (i.e., advanced scientific facilities, testbeds and networks) or leading scientific and technical resources, such as highly trained personnel and supporting infrastructure.**

DOE Office of Science Charge to the Subcommittee

3. **To preserve and foster U.S. leadership roles within reasonable resource constraints, are there particular technical areas or capabilities that could be emphasized?**
   – Are there other technical resources and capabilities that could be leveraged in to achieve these goals, possibly through collaborations within and beyond the ASCR community?

4. **How can programs and facilities be structured to attract and retain talented people?**
   – What are the barriers to successfully advancing careers of scientific and technical personnel in advanced computing, computational science and engineering, and related fields and how can the Department address those barriers?
   – How can we ensure that we are recruiting, training, mentoring, and retaining the best talent from all over the world, including among traditionally underrepresented groups within the U.S.

ASCAC subcommittee on American Competitiveness

Committee Members

ASCAC Members:

- Jack Dongarra (UTK/ORNL) (chair)
- Tony Hey (STFC)
- Satoshi Matsuoka (RIKEN)
- Vivek Sarkar (GATech)

ASCR Office: Christine Chalk
ASCAC Chair: Dan Reed

External Members:

- Greg Bell (Corelight)
- Ewa Deelman (USC/ISI) (vice chair)
- Ian Foster (Argonne/Chicago)
- David Keyes (KAUST)
- Dieter Kranzlmueller (LRZ)
- Bob Lucas (ANSYS)
- Lynne Parker (UTK) (excused herself)
- John Shalf (LBNL)
- Dan Stanzione (TACC)
- Rick Stevens (ANL)
- Kathy Yelick (UC-Berkeley)

Support for the Subcommittee

- **People who provided support**
  - **Logistics and Technical Support**
    - Ann B. Gonzalez, ORISE
    - Deneise Terry, ORISE
    - Gail Pieper, ANL

Process

- **July 2022: Subcommittee Formed**
- **August 26, 2022: Zoom call to organize into subgroups**
  - Critical Scientific Areas for Leadership in ASCR
    - Tony(lead), Dan, Greg, Rick
  - Advanced Research Tools
    - David(lead), Ian, Kathy, John
  - Building and Maintaining Strategic Industry and International Partnerships
    - Ewa(lead), Dieter, Jack
  - Strategies for Success, Recruitment, and Retention
    - Vivek(lead), Satoshi, Bob
- **October 7, 2022: Zoom call to review and discuss sections**
- **November 16, 2022: SC informal meeting of subcommittee to discuss progress**
- **December 11, 2022: Zoom call to discuss progress and plan for January meeting**
- **January 11-12, 2023: Face to face meeting to discuss report**
- **February 24, 2023: Zoom call to discuss what's needed**
- **March-May: Multiple rounds of report editing**

- Reviewing relevant reports and publications.
- Discussions with colleagues in DOE and academia.

Executive Summary

- The US has long been the international leader in areas of Advanced Scientific Computing Research critical to DOE's Office of Science
  - This includes the development of large-scale networking and computational facilities, and research in applied math and computer science.

- The ASCR program has provided the necessary HPC and networking capabilities and expertise to support DOE's mission to advance the national, economic, and energy security of the United States. The products of this program are in use world-wide.
  - ASCR can potentially maintain scientific leadership in the critical areas while strengthening our research infrastructure and train a large, diverse cohort of scientists.

- Leadership requires significantly increased investments, as well as innovative policies and programs.
  - Competing priorities must be balanced, but it should not limit our imagination or silence our advocacy.

- The DOE's ASCR program is a key part of the US research infrastructure and has a responsibility to pursue its mission, including advanced scientific computing and the advanced research facilities it requires, with determination and enthusiasm.
  - However, US scientific leadership in HPC is at risk, and with it all of the research and technological tools and processes that it undergirds in commerce, defense, environment, and society.

- ASCR's program must not only develop and publish a clear vision with an associated list of goals, priorities, and recommendations but also enable scientific leadership by consistently securing long-term funding.
  - This will allow the program to build on its achievements to date, to realize its ambitious vision, and to make lasting contributions to the field.

Key Findings and Recommendations

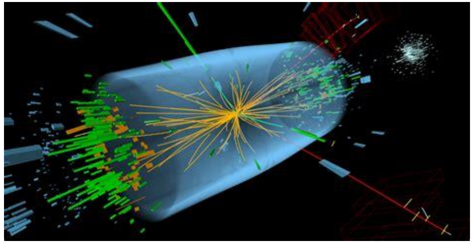
Key Findings (Strengths)

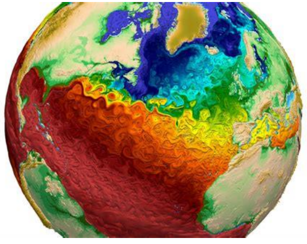
1. **Science and engineering applications of national importance will continue to require increasingly more capable advanced computing systems.**

    - Complex phenomena, process, analyze, and manage vast amounts of data, and support cutting-edge experiments.

    - Will require major and sustained advances in computing, networking, mathematics, computer science, and AI technologies.

    - The national labs and their university partners are uniquely qualified to produce those advances, but only if supported appropriately in terms of leadership, vision, and predictable and sustained funding.

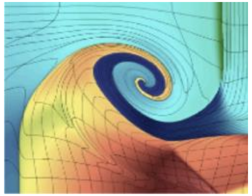


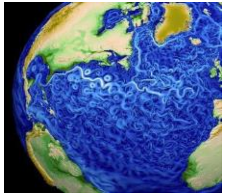
Key Findings (Strengths)

2. Led in large part by DOE, the US has been an international leader in applied mathematics and computational science research.

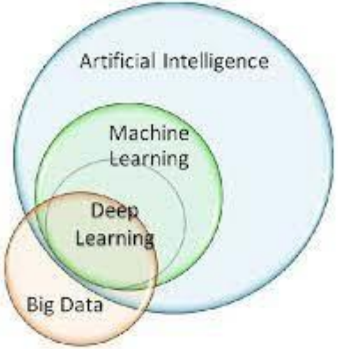
- The United States has also been a leader in computer science, with DOE's role focused on those areas related to HPC (e.g., programming, parallel algorithms, and performance optimization techniques) as well as networking and data science (methods and tools for scientific discovery).

- In part because of: Sustained Funding, Collaborative Research Centers, Interdisciplinary Approach, HPC facilities, Labs, Global Collaboration





Key Findings (Strengths)



3. **HPC enables Big Data, simulation, and AI/ML, which is critical to scientific discovery and are synergistic.**
    - Experimental facilities across DOE’s Office of Science are increasing the demands for leading-edge computing and networking facilities, methods, and services.
        - These demands include the ability to move, analyze, share, and manage exponentially growing datasets from observational sensors and increasingly powerful scientific instruments and to use AI technologies to integrate that data with physics-based and data-driven models, which may themselves produce enormous datasets and require massive computing for model training and inference.
    - HPC catalyzes scientific discovery by enabling efficient processing and analysis of Big Data, supporting high-fidelity simulations, and empowering AI/ML applications.

Key Findings (Strengths)

4. The Exascale Computing effort is an exemplar of US leadership in high performance computing.

- Incorporates the latest mathematical and computational innovations into scientific applications, creating a comprehensive exascale software stack, and advancing the capabilities at the leadership-class computing facilities to enable future scientific breakthroughs.

- Cutting-edge hardware and software development, has an impact on scientific discovery, economic competitiveness and national security, global collaboration and influence.

Key Findings (Strengths)

5. The DOE has a history of partnering closely with industry collaborators.

- Developing, deploying, and applying advanced technology, particularly in the context of leadership-class computer systems and cutting-edge network services.

- DOE laboratories also work closely with end users from industry and have achieved numerous high-impact results extending participating companies’ capabilities. (See Cristina Thomas' talk from yesterday)

    - Technology Transfer, Shared Facilities and Expertise, Workforce Development and Training, and Small Business Collaboration

ECP Industry Council Members

Recommendation 1: Critical Areas

Building on its existing strengths in computational science, advanced computing, and unique user facilities, ASCR should be largely focused on four key areas:

- **(a)** **High-end modeling and simulation for science and engineering** (e.g., applied math, software, advanced applications)
- **(b)** **Artificial intelligence for science and engineering** (e.g., AI methods, software, data sets, advanced applications)
- **(c)** **Leading-edge computing architectures and systems on the path toward Post-Exascale** (e.g., hardware architecture, software, deployed infrastructure)
- **(d)** **Advanced networks and future internet architectures for an integrated research infrastructure** (e.g., architecture, software, deployed infrastructure).

Each of these four areas has long-term research challenges that should be pursued through a combination of base program funding (promotes career development), and opportunistic calls (provides flexibility). Each of these areas also demands the development and deployment of infrastructure (e.g., codes, libraries, models, HPC, AI, data and edge hardware facilities, national facilities) that supports the broader research enterprise.

Key Findings (Strengths & Challenge)

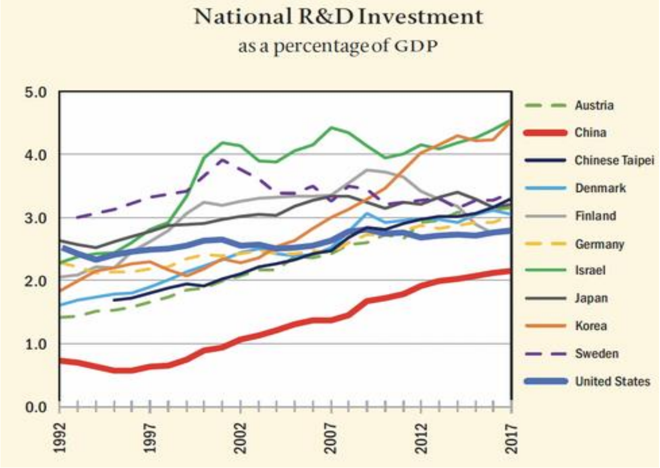
6. **The end of the Exascale Computing Project (ECP) is both a success and a huge risk.**

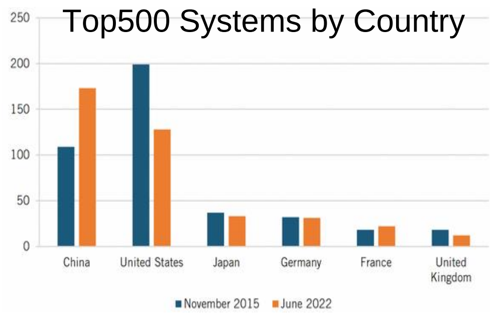
- The project delivered great capabilities, both human and technical.

- However, **DOE is highly vulnerable to losing the knowledge and skills of trained staff** as future funding is unclear.
  - This has become a major source of uncertainty and anxiety, since no clear message has been communicated as to what the post-ECP plans.
  - There needs to be a sustained program to look beyond exascale.

Key Findings (Challenges)

7. **The US, DOE, and ASCR leadership in critical areas is under threat.**
   - Global Competitiveness: China, Japan, and several European nations, have been investing heavily in their scientific research and development capabilities.
   - Funding and Budgetary Constraints: There needs to be a clear strategy to grow funding for sustaining leadership in critical areas. However, the US government's budgetary constraints and shifting priorities can pose challenges to sustained funding for scientific research and computing initiatives.
   - Brain Drain and Talent Retention: The retention of top scientific talent is essential for sustaining leadership in critical areas.





U.S. DEPARTMENT OF ENERGY | Office of Science

Key Findings (Challenges)

8. The technology landscape has fundamentally changed.

- DOE and ASCR must embrace agility, innovation, and strategic foresight.
  - Dennard scaling ending, Moore’s law waning
  - Hyperscalers
  - Rise of custom silicon
  - Other disruptive technologies, i.e. quantum

- By adapting to emerging technologies, fostering interdisciplinary collaboration, forging strong partnerships, engaging with global counterparts, and considering ethical and societal implications, the DOE can effectively navigate this transformed landscape and continue to fulfill its mission of advancing scientific discovery, addressing energy challenges, and ensuring national security.

Key Findings (Challenges)

9. **Unlike in the past, today's scientific research landscape and HPC supply chain is horizontal and international.**

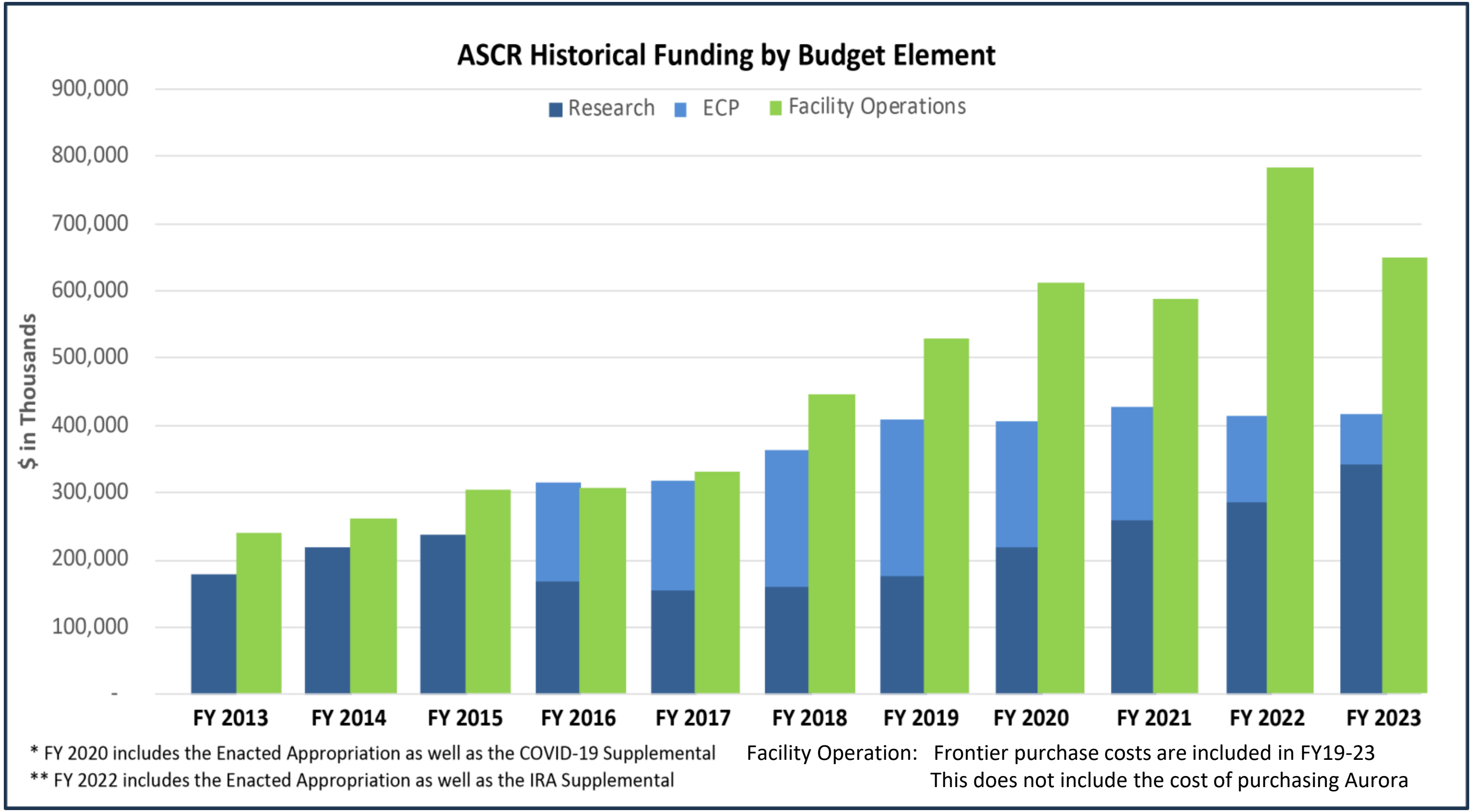
- Leadership in HPC requires proactive, long-term, and sustained engagement with this broad international HPC ecosystem.

- Industry partnerships are essential and merit attention and improvement.
    - Particularly in co-design, there are lessons to be learned from ECP and other international efforts so that the process can be improved in the future.

Standardization and interoperability
Recruiting from a worldwide talent pool
Global collaboration and knowledge exchange
Open science and open source
Collaborative funding models
Global HPC supply chain

Key Findings (Challenges)

10. ASCR funding levels for research are level and declining in real terms.



Key Findings (Challenges)

11. The attractiveness and prestige of careers in national labs have been on the decline.

- Increasing competition from other sectors.
  - Lack of long-term program vision and stable funding from ASCR and DOE.
- The funding uncertainty at the end of ECP has creating a "cliff" regarding confidence and the ability to retain and recruit talent.
  - Given the shakeup in the tech industry, this is a huge missed opportunity.
- Funding challenges have impacted the attractiveness of careers in national labs.
  - In an era of budget constraints, national labs have faced reduced funding and tighter research budgets.
- The bureaucratic nature of national labs may prevent joint appointments with academia and industry.

Recommendation 2: Vision and Planning

- ASCR leadership should work with the DOE labs to develop a **decadal plus post-exascale vision and strategy that builds on its strengths in the mathematics and computing research program working together with its world-class facilities.**

The focus should be to provide sustained investments to preserve and extend ASCR's current leadership in computational science research and multidisciplinary team science while also establishing new application areas in emerging topics such as:

- AI for science, energy, and security,
- Digital twins, and
- Post-Moore technologies.

Key Findings (Outlook)

12. Big science and advanced scientific computing increasingly require international collaborations.

- International collaborations facilitate the exchange of knowledge, skills, and best practices.
- Enhance the global impact and relevance of scientific research.
- International collaborations foster data sharing and access to software and large datasets.
- Significant opportunities for international collaboration exist in areas such as standard interfaces and libraries and many more.

Recommendation 3: Vision and Planning

ASCR needs to articulate a **vision, associated goals, and milestones for international collaboration focused on post-Exascale computing**.

ASCR, in partnership with NNSA/ASC and other SC and energy offices should strive to become international leaders in:

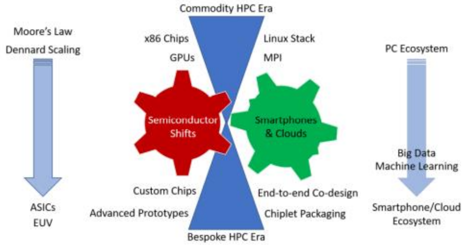
- AI for science

  Research in AI for science, energy, and security problems

  Software stack that uses the best of AI, simulation, and the combination.

ASCR should work to establish trust relationships with strategic partners, evangelize and socialize these efforts, define agreement structures (perhaps beyond the traditional MOU), and provide resources to develop flexible multiparty collaborations.

Key Findings (Outlook)



13. ASCR facilities will continue to require an interdisciplinary approach requiring co-design, rather than a reliance solely on the vendor marketplace.

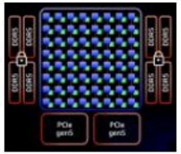
This approach fosters innovation, deep integration of scientific knowledge, and a culture of collaboration, ultimately driving meaningful advancements and improving the effectiveness of research outcomes.

Current approaches to designing HPC systems must change in deep and fundamental ways,

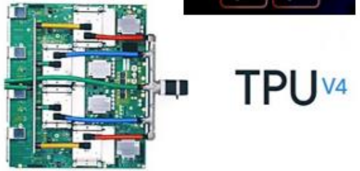
- Embracing end-to-end co-design;
- Custom hardware configurations and packaging;
- Large-scale prototyping; and
- Collaboration between the dominant computing companies, smartphone and cloud computing vendors, and traditional computing vendors.

- Alibaba
  - CIPU, 128 core ARM based
  - Alibaba's Elastic Compute Service

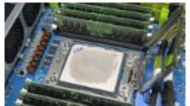
- AWS Graviton3
  - 64 ARM Neoverse V1 cores, chiplet design
  - 55 billion transistors, DDR5 memory



- Google TPU4
  - 2X TPU3 performance
  - 4096 units per "pod"
  - Reconfigurable optical interconnect



- Microsoft Azure
  - Ampere Alta ARM processors
  - Project Catapult/Brainwave



Recommendation 4: Strategy

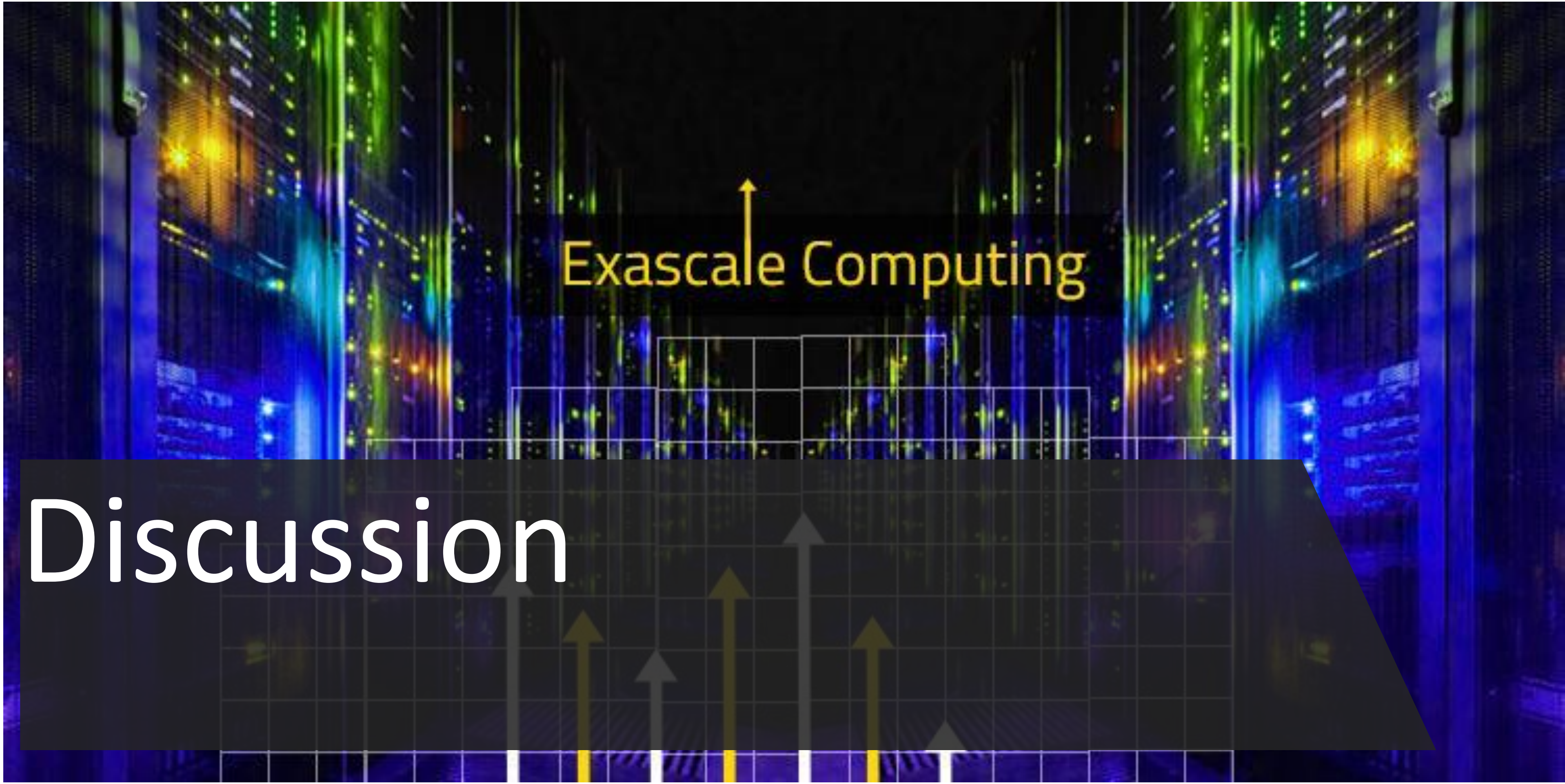
ASCR needs to invest in long-term forward-looking co-design research in advanced computer architecture, hardware, and system concepts to identify potential solutions for sustaining continued scientific productivity increases for future scientific computing systems.

Such a co-design effort will require substantially increased government investment in basic research and development. In addition, DOE should fund the building of real hardware and software prototypes at some scale to test new ideas using custom silicon and associated software.

Business-as-usual will not be adequate.

Conclusions – DOE, ASCR, and the Labs Face an Existential Crisis

- The US is losing its historical leadership position in advanced scientific computing research.
  - The trend will likely lead to the United States producing a smaller overall share of technological innovations.
  - A major factor contributing to this decline is the significant investment in fundamental research by other countries.
  - Reduced investment in intellectual underpinnings raises concerns about the future success of DOE’s scientific endeavors.
- The United States was long considered a highly desirable destination for the career development of scientists, and its national laboratories were respected worldwide as hosting the most talented researchers. This is no longer the case.
  - Funding uncertainties and a move away from sustained funding intended to produce foundational innovations to short-term research contracts focused on near-term goals have made DOE laboratories a far less attractive location for both junior and established researchers.
- ASCR should revive stable funding, maintain its stewardship of state-of-the-art facilities, and develop a long-term visionary research program for advanced scientific computing.



Discussion