SIAM Task Force Report:

# The Future of COMPUTATIONAL SCIENCE

2024

# Table of Contents

# Executive Summary

Computational science uses math and computing to advance science and engineering. It is a core part of most scientific fields and underpins much of the modern world. Computational science enables the development of novel industrial products, the design of new drugs, weather and climate prediction, the protection of national security, forecasts of and responses to natural disasters, and much, much more. Computational science is on the cusp of making major advances and dramatic new impacts by leveraging recent progress in computing and applied mathematics. But the field also faces significant headwinds from hardware and workforce challenges.

Advances in artificial intelligence, powerful new computing platforms, and an increasingly complex landscape for future computing hardware represent both new opportunities and new challenges. These developments place computational science at an inflection point. At this time of simultaneous promise and challenge, the Society for Industrial and Applied Mathematics (SIAM) commissioned a task force on the future of computational science. SIAM is the primary professional society for computational science. It is a 14,000-member organization representing applied mathematicians, computational scientists, data scientists, and engineers. Members come from many different disciplines, but all share an interest in applying and developing state-of-the-art techniques of mathematics and computational science to solve real-world problems. The goal of this task force was to assess this complex landscape and to craft a strategic vision for the field for the next 15 years. The members of the task force were selected to represent a broad range of scientific backgrounds and perspectives, and included members from academia, national labs, and industry. This report is the result of their deliberations and assessments.

Future advances in computational science promise to pay dividends in sectors as diverse as healthcare, energy, science, resilience, and defense. Reaping these dividends will require a clear-eyed strategy and investment that enables progress across application areas while enabling new core capabilities such as digital twins, artificial intelligence-driven modeling, and real-time decision making. Changes in the computing landscape will add to the challenges as historical drivers for improvement in computer performance have run their course. Future advances in computer performance and power efficiency will require heterogeneous platforms with specialized hardware and exotic accelerators like quantum processors. Thus, as the Exascale Computing Project nears completion at the U.S. Department of Energy, this is no time to scale back investment. We have just begun to realize the power of exascale computing and much more research and development is needed to fully realize its benefits, let alone to prepare for post-exascale technologies. International competitors are investing strongly in computational science, challenging the United States to sustain and improve its lead in the underlying computing technologies and their uses across the application space. U.S. national security and economic competitiveness depend on continued leadership in computational science and that leadership can only be maintained through investments in applications, algorithms, and co-development with hardware vendors of the optimum combinations of hardware and computational science methods.

The SIAM Task Force on the Future of Computational Science articulated ***one*** overarching national priority.

> **Computational science is essential:** Computational science plays a crucial role in scientific discovery, the economy, and national security, but U.S. leadership is under threat. Investments that ensure continued U.S. leadership should be a high national priority.

Continued leadership will require overcoming

Back to TOC

significant challenges. The Task force identified ***two*** major impediments to further progress in the field.

**Future high performance computers will be challenging to design and difficult to program:** Since the 1960s, device engineers have made continuous progress on smaller and more energy efficient transistors. This has enabled exponential growth in computer performance, enabling many of the impacts of computational science. However, without significant hardware and software advances, future progress in computational science will be made at a much slower rate. Performance improvements in future computers will require new architectures that are significantly more energy-efficient than present supercomputers. Exotic technologies like quantum accelerators may be part of future, more heterogeneous supercomputers. These architectural changes will be highly disruptive to the applications, algorithms, and software used for computational science today. Significant new investments are needed to ensure that modeling, data science, simulation, and other activities can be adapted to the promising but challenging new computing environments just over the horizon.

**Existing approaches for attracting and preparing the future computational science workforce are insufficient:** The opportunities and challenges facing computational science can be addressed only through the efforts of a large, highly skilled workforce. Computational science is inherently interdisciplinary, and there is already a shortage of expertise. New approaches are needed to widen and deepen pathways into computational science fields and to ensure that the broad research community obtains the skills needed across all disciplines to advance increasingly complex computational efforts.

The Task Force identified ***three*** areas in which the scientific landscape is changing dramatically, and computational science has the opportunity to make transformative progress for the betterment of science and society.

**Exascale computing will enable unprecedented science:** With the arrival of the Frontier supercomputer at Oak Ridge National Laboratory, we have entered the exascale era ($10^{18}$ operations per second). Two more exascale machines will be standing up in the coming months. The Department of Energy's Exascale Computing Project (ECP) has created a large corpus of exascale-ready software ranging from tools and libraries to a diverse suite of applications. Never before has the community had a common core of high-quality software to build upon. ECP has built the tools, and now additional investments are needed to sustain, grow, and apply the software base to ensure the full promise of these machines in enabling new scientific discoveries and technological advancements.

**Science increasingly relies on large and complex data streams:** While modeling and simulation have been at the heart of computational science for many decades, in recent years, data science has become central to scientific progress. Scientific research is awash in data from high-throughput experiments, ubiquitous sensors, and simulations themselves. Investments to support research and development are needed to create a wide range of data management, data processing, and data analysis capabilities for scientific applications. In addition, much more research is needed to integrate data science with simulation and artificial intelligence.

**Artificial intelligence will create entirely new ways to do computational science:** In just the last few years, artificial intelligence (AI) and machine learning (ML) have begun to transform broad swaths of commerce and society. These technologies are beginning to have major benefits for science and engineering as well, but the field is still young. AI is being used to accelerate

Back to TOC

simulations, to combine experiments with simulations, to automate workflows, to propose new hypotheses, and much more. This rapidly developing area will be a major driver of scientific progress for the foreseeable future, but only if investments are made to ensure that existing or new AI technologies are appropriately reliable and trustworthy for scientific and engineering applications.

We offer this ***one-two-three*** framework as a constructive way to think about the most important decadal priorities in the field and the investments that must be made to address them. In meeting these challenges and seizing these opportunities, we will be creating the capabilities that enable critical future progress against problems that cannot be addressed today. Examples include cloud-resolving climate models, quantum-accurate electronic structure, materials-by-design, personalized healthcare, and numerous other applications that will create strategic and economic opportunities tomorrow. These applications require new mathematics, algorithms, and computer science, running on post-exascale computers. Without appropriate investments, the U.S. will lose leadership in this critical area and cede the technological advances of tomorrow to others.

Back to TOC

# 1. Introduction

Computational science is the discipline focused on the development and use of mathematics and computing to support and advance the frontiers of science and engineering. Driven by improvements in algorithms and vast increases in computer performance, computational science is now central to nearly all branches of science and engineering, and enables the development of novel industrial products, the design of new drugs, weather and climate prediction, enhancements to national security, forecasts of and responses to natural disasters, and much, much more. Computational science is poised to deliver further breakthroughs that will advance knowledge, create industrial advantage, and ensure national security. But the field is also experiencing unprecedented headwinds due to changes in computing technologies and a lack of workforce to meet future needs.

As the field enters an era of simultaneous promise and challenge, the Society for Industrial and Applied Mathematics (SIAM) commissioned a task force on the future of computational science. SIAM is the primary professional society for computational science. It is a 14,000-member organization representing applied mathematicians, computational scientists, data scientists, and engineers. Members come from many different disciplines, but all share an interest in applying and developing state-of-the-art techniques of mathematics and computational science to solve real-world problems. The goal of this task force was to assess this complex landscape and to craft a strategic vision for the field for the next 15 years. The members of the task force were selected to represent a broad range of scientific backgrounds and perspectives, to span from early career to senior researchers, and to reflect a breadth of professional affiliations including universities, national labs, and industry. The task force conducted its work through a series of emails and virtual meetings, and a two-day gathering facilitated by Lewis-Burke Associates. This report is the result of their deliberations and assessments.

As illustrated by the examples on p.5, computational science underpins much of our modern economy, technology, and national security. The Task Force had ***one*** overarching finding.

> **Computational science is essential:** Computational science plays a crucial role in scientific discovery, the economy, and national security, but U.S. leadership is under threat. Investments that ensure our continued leadership should be a high national priority. The latter part of **Section 1** discusses the competition for U.S. leadership in computational science.

Continued leadership will require overcoming significant challenges. The Task Force identified ***two*** major impediments to further progress in the field.

> **Future high performance computers will be challenging to design and difficult to program:** Since the 1960s, device engineers have made continuous progress on smaller and more energy efficient transistors. This has enabled exponential growth in computer performance, enabling many of the impacts of computational science. However, without significant hardware and software advances, future progress in computational science will be made at a much slower rate. Performance improvements in future computers will require new architectures that are significantly more energy-efficient than present supercomputers. Exotic technologies like quantum accelerators may be part of future, more heterogeneous supercomputers. These architectural changes will be highly disruptive to the applications, algorithms, and software used for computational science today. Significant new investments are needed to ensure that modeling, data science, simulation, and other activities can be adapted to the promising but challenging new computing environments just over the horizon.

Back to TOC

Although it is frequently invisible, computational science enables many of the technologies that drive modern economies and societies. For example:

## Medicine and health care

Advanced algorithms and computing were at the heart of the human genome project and are central to genomics and metagenomics. Drug design relies on high-fidelity, atomistic models of proposed therapeutics and their targets. Epidemiology utilizes social and behavioral models to optimize interventions against infectious diseases. Personalized medicine employs sophisticated data analysis tools to devise appropriate treatment plans for each patient.

## Earth and environmental science

Sophisticated models of the atmosphere enable weather predictions that inform everything from picnic plans to farming and disaster response. Whole earth models that combine the atmosphere, the oceans, and ice coverage shape our understanding of climate change and its consequences. Ubiquitous sensors combined with data management and analysis capabilities allow for the monitoring of biosystems and optimization of farm yields.

## Materials science

Computational modeling and machine learning accelerate the discovery of new materials for batteries, solar cells, and innumerable other applications. Quantum-accurate materials models provide deep, quantitative understanding of material properties, and support optimal design

## Manufacturing and industrial competitiveness

Model-based design allows for the design of safer cars, energy-efficient airplanes, improved industrial processes, and effective additive manufacturing technologies. Modeling coupled with data science enables "digital twins," which are used to optimize maintenance schedules.

## Energy

Modeling and real-time data assimilation help with the design and control of the electrical grid. Advanced simulations support the design of more efficient wind turbines and better layout of turbines in a wind farm. Simulations enable more efficient engine designs.

## National security

Modeling and machine learning are central to nuclear stockpile stewardship. State-of-the-art aircraft, submarines, and weapons systems are optimized via modeling and simulation. Data management and synthesis is essential to battlefield management.

## Fundamental science

Astrophysics relies on simulations of stars, galaxies, and entire universes. Particle physics employs simulations to understand the fundamental forces of the universe. Large scientific facilities (e.g. accelerators, telescopes, light sources) generate enormous amounts of data that must be managed, stored, and analyzed.

Back to TOC

**Section 3** discusses prospects for hardware advances, and the corresponding software developments, that pursue improvements to traditional digital microelectronics, while **Section 4** discusses more disruptive approaches such as quantum and neuromorphic computing.

**Existing approaches for attracting and preparing the future computational science workforce are insufficient:** The opportunities and challenges facing computational science can be addressed only through the efforts of a large, highly skilled workforce. Computational science is inherently interdisciplinary, and there is already a shortage of expertise. New approaches are needed to broaden and deepen pathways into computational science fields and to ensure that the broad research community obtains the skills needed across all disciplines to advance increasingly complex computational efforts. **Section 7** discusses approaches to create pathways into computational science for historically underrepresented communities and existing workers with skills in adjacent areas.

The Task Force identified ***three*** areas in which the scientific landscape is changing dramatically, and through which computational science will have the opportunity to make transformative progress to the betterment of science and society.

**Exascale computing will enable unprecedented science:** With the arrival of the Frontier supercomputer at Oak Ridge National Laboratory, we have entered the exascale era ($10^{18}$ operations per second). Two more exascale machines will be standing up in the coming months. The Department of Energy's Exascale Computing Project (ECP) has created a large corpus of exascale-ready software spanning from tools and libraries through a diverse suite of applications. Never before has the community had a common core of high-quality software to build upon. ECP has laid a foundation, but additional investments are needed to ensure the full promise of these technologies in enabling new scientific discoveries and technological advancements. **Section 2** discusses the importance of leveraging the advances of ECP, as well as the potential impact of emerging technologies like digital twins, and the combination of artificial intelligence and computer simulation.

**Science increasingly relies on large and complex data streams:** Modeling and simulation have been at the heart of computational science for many decades, but in recent years, data science has become central to scientific progress. The world is awash in data from high-throughput experiments, ubiquitous sensors, and simulations themselves. **Section 5** discusses the need for research and development to create a wide range of data management and processing capabilities for scientific applications and the need for much more research to integrate data science with simulation and artificial intelligence.

**Artificial intelligence will create entirely new ways to do computational science:** In just the last few years, artificial intelligence (AI) and machine learning (ML) have begun to transform broad swaths of commerce and society. These technologies are beginning to have major benefits for science and engineering as well, but the field is still very young. AI is being used to accelerate simulations, to combine experiments with simulations, to automate workflows, to propose new hypotheses, and much more. This rapidly developing area will be a major driver of scientific progress for the foreseeable future, but only if investments are made to ensure that existing or new AI technologies are appropriately reliable and trustworthy for scientific and engineering applications. **Section 6** discusses the opportunities for modern AI to transform the scientific landscape.

Back to TOC

# The competition for U.S. leadership in computational science

Until recently, the United States was the unquestioned leader in advanced computing and computational science. But in recent years that leadership has been challenged by strategic rivals. Europe and Japan have always had strong research programs in computational science, but the recent emergence of China has been particularly striking. The Top 500 List[1] tracks the fastest computers in the world by a widely embraced benchmark. The list has long been dominated by U.S. systems, but a Chinese computer took the top spot in November 2014, and it was surpassed by another Chinese system in June 2016. By November 2017, China had more computers in the list than the U.S. In recent years, China has stopped submitting results to the Top 500 List, so the current status of their efforts is hard to assess.

China has also made great strides in applications of computational science, as measured by the yearly Gordon Bell Prize. Teams of Chinese scientists garnered this award in 2016, 2017, and 2021.

The impacts of international competition are already evident. European weather-prediction models are generally viewed as superior to their U.S .counterparts. The European Union is implementing digital twins to better understand the European Earth System and make revolutionary discoveries. In national security, the U.S.'s adversaries have used their increased computational power to help develop new weapons, such as hypersonic missiles, which the U.S. is still working to master.

**Finding 1.0: Given the central role of computational science in scientific discovery, industrial competitiveness, and national security, the federal government must make the necessary investments to ensure continued U.S. leadership in the field.**

The United States federal government has long played a central role in the advancement of computational science and its application to a variety of scientific fields. Many federal agencies invest in the development and application of computational science, most critically the Department of Energy (DOE) and the National Science Foundation (NSF) as well as several others such as the Department of Defense (DOD), National Institutes of Health (NIH), and National Institute of Standards and Technology (NIST). Even more agencies depend on computational science including the National Oceanic and Atmospheric Administration (NOAA), National Aeronautics and Space Administration (NASA), and Environmental Protection Agency (EPA). DOE's unique role for many decades has been to push the frontiers of applied mathematics, computer science, and scientific software at the leading edge of high performance computing, both by providing computing platforms, and through support of basic research and development in computational science. Computational science enabled by high performance computing is central to many of DOE's missions including scientific discovery, alternative energy, environmental remediation, and stockpile stewardship. Because of DOE's leadership role in high performance computational science and the impact DOE-supported advances will have on computational science more broadly, the Task Force has chosen to focus its recommendations on actions that the DOE can take to ensure the nation maintains its leadership role in computational science.

The unique role of the DOE and its predecessors among the federal agencies in the development and promotion of computational science dates back to the early days of the modern supercomputer era. More recently, its role has led to the deployment of exascale computers and the advanced software technology of the DOE Exascale

1 Top 500 List: top500.org

Back to TOC

Computing Project (ECP), a partnership between DOE's Advanced Scientific Computing Research (ASCR) program and National Nuclear Security Administration (NNSA). In particular, ECP's software technology heavily leveraged decades of basic research in applied mathematics, computer science, and computational science supported by ASCR, whose mission is "to discover, develop, and deploy computational and networking capability to analyze, model, simulate, and predict complex phenomena important to the Department of Energy and the advancement of science." In addition, existing ASCR programs in computer science and applied mathematics have advanced capabilities and seeded new areas of research for revolutionary advancements. A key element of DOE's research programs over the past two and a half decades has been the support by multiple DOE Office of Science Program Offices of partnerships between researchers in the domain sciences and those in mathematics and computer science. In particular, the Scientific Discovery through Advanced Computing (SciDAC) program has supported collaborations bringing leading-edge mathematics and computer science advances to bear on scientific and engineering challenges of importance to the DOE. The recently established Energy Earthshots program[2] has funded similar partnerships aimed at advancing clean energy technologies in support of the nation's climate and clean energy goals.

DOE's leadership in computational science and engineering has been critical to U.S. leadership and innovation across the scientific spectrum, and sustained support for ASCR's ongoing programs is vital for this leadership to continue. ASCR continues to be uniquely positioned to play a pivotal and continuing role as one of the leading programs for the advancement of computational science to meet the problems facing our world.

However, with the completion of ECP, DOE stands at a crossroads on its next priorities and major efforts. DOE has provided significant support to deploy major computing platforms at the exascale, to develop an exascale computing software stack, and to build a suite of scientific applications that can take advantage of exascale capabilities. For progress into the future, a comprehensive plan and program is needed to leverage these advances for future computational science impact, and to begin the hard work of preparing for the next generation of computers past exascale. Given the advances of our competitors and the future U.S. needs for computational science, current federal investments are insufficient to ensure continued U.S. leadership. The SIAM Task Force looks to DOE to craft a clear plan and implement a program that will drive forward the future of computational science, including new capabilities, to ensure continued U.S. leadership.

**Recommendation 1.1: Investments should be made to support a comprehensive computational science program that leverages the results of the Exascale Computing Program and anticipates and exploits future high performance computing platforms. To fully take advantage of ECP technology and future hardware, this program will require further advances in applied mathematics and computer science and the establishment of partnerships between applied mathematicians, computer scientists, and application scientists to address critical national challenges.** Exascale hardware and software provide a foundation for discoveries through scientific computing, but further research and development is required to fully realize the potential of exascale. In addition, future advances in computer performance and efficiency will come from computer designs that will require new algorithms and software. Current support for mathematics and computer science research aimed at enabling the use of future computers is insufficient to ensure they will be usable for science in a timely way. It is essential that a comprehensive program be implemented to develop the mathematics and computer science and to support the partnerships needed to enable application scientists to exploit these computers to address critical future challenges.

2 Energy Earthshot Initiative: www.energy.gov/energy-earthshots-initiative

Back to TOC

# 2. Computational science is poised for transformative new impacts

## Summary

*The demand for additional computational resources is unrelenting. The capability of the world's fastest computers has grown by a mind-boggling nine orders of magnitude in the past 35 years. With each increase in performance, new scientific opportunities come within reach. With the large investments by the Department of Energy (DOE) in the Exascale Computing Project (ECP), scientists now have access to computers that can process $10^{18}$ numerical operations per second. This level of performance allows for new ways of doing science, including:*

- *more complex and accurate simulations;*
- *the conjoining of simulation with data science and artificial intelligence;*
- *the ability to simulate in real time and so to drive physical experiments or systems; and*
- *the capacity to understand uncertainties and fully explore design spaces.*

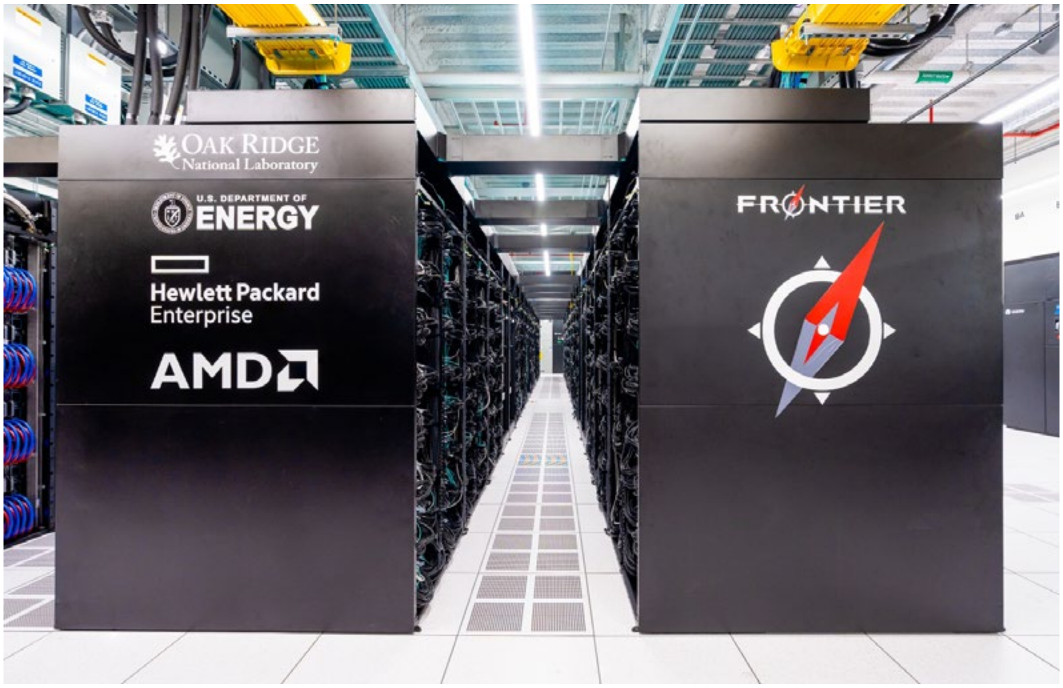
*Simultaneously, mathematical and algorithmic advances are opening new pathways to scientific discovery and engineering practice, such as applications of digital twins and the integration of machine learning with simulation. With the right investments, computational science is poised for a dramatic growth in applications and impact. As with past progress in computational capability, seizing these new opportunities will require advances in fundamental mathematics, algorithms, computer science, and application science.*

**Finding 2.0: DOE's recent investments have produced supercomputers, software tools, and applications with exascale capabilities. Similarly, recent mathematical and algorithmic advances have led to novel approaches for answering difficult scientific and engineering questions. A comprehensive investment strategy will be required to ensure that these capabilities can continue to advance and be leveraged to develop new pathways of scientific discovery and engineering practice.**

Computational science is used pervasively across science and industry to provide critical insights. Recent developments in computing and mathematics have created opportunities for a host of new, impactful applications. This section reviews three illustrative, emerging opportunities, each of which can be transformative.

### Leveraging exascale

With the success of the Department of Energy's (DOE) Exascale Computing Project (ECP), the research community now has access to transformational computing power and a highly sophisticated, community-endorsed software suite to enable their capabilities. The advent of the age of exascale computing has brought unprecedented opportunities to employ computational science to address scientific and societal grand challenges in astrophysics, biology, climate, energy, earth sciences, infrastructure, manufacturing, materials, medicine, and social sciences. The three-orders-of-magnitude increase in peak performance of current leading-edge systems over the past 15 years, accompanied by concomitant improvements in memory, networking, and storage,




Frontier, the world's first exascale computer, has about 8.7 million cores and performed $1.19 \times 10^{18}$ operations per second in a widely used supercomputing benchmark. Frontier is located at Oak Ridge National Laboratory.

have provided the computing power to make modeling and simulation tractable for many complex phenomena exhibiting multiscale, multirate, multiphysics, multidomain, and multimodel behavior. At the same time, the exascale era has provided new capabilities for analyzing and interpreting rapidly expanding volumes of observational and model-generated data. The DOE and other federal agencies need to provide the necessary support to seize the exascale opportunity and to sustain U.S. scientific leadership. Exascale applications will require methodological advances in multiphysics simulations, uncertainty quantification, design optimization, and the conjoining of artificial intelligence with simulation. Managing exascale simulations will require new approaches to system administration, workflow management, data analysis and visualization. All these needs can only be met with additional research investments from the federal government.

To manage the complexity of building exascale applications, ECP built a large collection of software tools and infrastructure including numerical libraries, mesh generation toolkits, performance profilers, programming models, and visualization libraries. All these tools were designed to run at exascale. This entire corpus of community software encompassing enabling technology and applications is a principal product of the nation's large investment in ECP. For the first time, the community has a very high quality and broadly embraced software stack that ranges from libraries and programming models through a variety of scientific applications. This software stack will underpin scientific discovery for many years to come. Further research investments in mathematics, algorithms, and software are needed to expand these capabilities and to customize and apply them to emerging applications.

**Recommendation 2.1: Further investments are required to support the research and development activities needed to fully leverage the promise of exascale computing.** These investments should include mathematics, computer science, application science, and system software. Exascale applications will require methodological advances in multiphysics simulations, uncertainty quantification, design optimization, and conjoining artificial intelligence with simulation. Managing exascale applications will require new approaches to system administration, workflow management, data analysis, and visualization. All of

Advanced manufacturing production processes, including 3D printing, are complex and produce parts that require careful inspection to ensure they are high quality and defect-free. Researchers at Lawrence Livermore National Laboratory are using digital twins to help optimize the process, catch existing errors, prevent new ones, and ensure the timely delivery of high-quality parts.

these needs can be met only with additional research investments from the federal government.

**Recommendation 2.2: An explicit program should be established to leverage and continue to advance the community software created by the Exascale Computing Project.** This body of software developed under ECP will underpin the community for many years to come, and will provide a cost-effective way to develop, support, port, and maintain a host of advanced applications. But this software will not be able to address continually evolving needs without further investment. DOE should oversee a research and development program to continue to extend this functionality, adapt it to new computers, and to apply it to new application areas.

## Digital twins

A digital twin[3] is a computer model of a natural, engineered, or social system that is dynamically updated with data from the real system and is used to inform decisions. Data streaming from the physical system is assimilated into the computational model to reduce uncertainties and improve predictions of the model. The updated model in turn is used as a basis for controlling the physical system, optimizing data acquisition, and/or providing decision support. The digital twin must execute rapidly enough to support decisions and controls in timescales relevant to the physical system and must manage and quantify uncertainties across its lifecycle.

The concept behind digital twins has been around for decades, but recent progress in mathematics, algorithms, and computing capability have opened broad, new vistas of possible applications. Digital twins are being developed in nearly every sector of industry—from transportation to the process industry, from defense to manufacturing, from energy to microelectronics, from civil infrastructure to aerospace. Moreover, digital twins are being deployed for many natural systems, such as severe weather, wildfires, tsunamis, earthquakes, and volcanoes, and earth and environmental systems including climate, biological, and ecological systems. Personalized medicine presents a future in which a predictive digital twin of an individual, continuously updated with sensor, lab test, and imaging data, will be used as a basis for diagnosis

3 Foundational Research Gaps and Future Directions for Digital Twins

Back to TOC

and treatment of diseases well before they can be caught using conventional techniques.

While digital twins present enormous opportunities, their development faces great mathematical, statistical, and computational challenges. This stems from the enormous complexity and scale of models describing complex physical systems targeted by digital twins, the numerous uncertainties that underlie these models, and the complexity of observing experimental systems and the indirect nature of the data they produce. Since digital twins subsume data assimilation and inverse problems, optimal control and model-based decision making, surrogates and model reduction, validation and uncertainty quantification, they inherit all of the individual challenges presented by those problems. Moreover, entirely new challenges arise when these constituent problems are integrated into digital twin frameworks, and research investments in mathematics and computer science must address the following needs:

- Faster solvers for real-time applications.
- Ensuring the stability of numerical methods for coupled systems.
- Uncertainty quantification with surrogate models (including machine- learned models).
- Feedback control methods for specific applications.

**Recommendation 2.3: Investment is needed to provide a strong scientific foundation for the development and use of digital twins.** This includes furthering the individual understanding of data assimilation and inverse problems, optimal control and model-based decision making, surrogates and model reduction, validation, and uncertainty quantification. In addition, understanding must be developed of the new challenges that arise when these constituent problems are integrated into digital twin frameworks.

## Combining simulation with artificial intelligence

The recent explosive progress in the capabilities of artificial intelligence (AI) and machine learning (ML) presents numerous opportunities and challenges for scientific computing. As discussed in Section 5, with the right investments, these technologies have transformative potential for scientific and engineering applications. The field is very young and new ideas are emerging rapidly, but areas of potential impact include:

- Fast approximate (surrogate) models for complex systems to enable quicker responses. Fast approximate models can enable real-time control of experiments or complex systems. They will also allow for more thorough explorations of parameter spaces to enable optimal designs, better understanding of uncertainties and margins, and higher confidence.
- New methods for merging experimental results with simulation outputs to allow for improved predictions of complex phenomena.
- Automated laboratories and user facilities to increase scientific throughput and discovery.
- Enhanced productivity for developers of scientific software, and fewer software bugs.

Much of the recent progress in AI and ML has been driven by industry through large investments in research, hardware, and software. The computational science community should leverage all of this to the extent possible and collaborate with industry where appropriate. These collaborations should examine scientific use cases and assess which existing capabilities are adaptable to scientific applications. They should also identify gaps where the scientific needs diverge from commercial applications and use these to prioritize new research directions. One possible example is how to exploit the low- and mixed-precision hardware being developed for AI,

Back to TOC

both mathematically and practically, for scientific software. New tools are also needed for managing heterogeneous hardware and complex workflows that integrate ML/AI languages like TensorFlow and PyTorch with traditional scientific computing languages like C++ and Fortran, allowing software developers to use the most appropriate hardware and language for each step of their workflows. These hybrid workflows will use computing resources differently than traditional simulations, so new approaches to resource management will also be needed. Further discussion of these opportunities and needs can be found in Section 6.

**Recommendation 2.4: Investments are needed in the research, software tools, and system management tools needed to enable complex workflows that combine simulation with machine learning.** The computational science community should leverage recent progress in AI and ML in industry and identify new AI and ML research directions that will enable new and faster scientific and engineering capabilities. Many existing applications involve the use of ML models as very fast "surrogates" (approximate replacements) for more expensive simulation codes.

Back to TOC

# 3. Rapid change is afoot in high performance computers

## Summary

*High performance computing (HPC) is a key element of computational science, but HPC itself is at an inflection point. Historical drivers for performance improvement have run their course, with transistors now nearly as small as they can get and supercomputer energy consumption at a practical maximum, making the energy-efficiency of future HPC platforms of paramount concern. The only viable way to continue improving performance will involve architectural specialization and heterogeneous supercomputers. A range of federal investments are needed to ensure these future machines will be applicable to computational science needs. These investments should include researcher access to early hardware and vendor incentives to serve the computational science market. In addition, the lack of clarity about future machines creates research challenges for algorithms and software engineering today, since software lives much longer than computers.*

**Finding 3.0: The nation needs a suite of investments to ensure that the development of high performance computing continues beyond exascale to meet the nation's continually evolving needs for advanced computational science.**

For more than 50 years, the density of transistors on a chip has doubled every 1.5 years. This exponential growth, combined with faster clock rates and larger machines, has led to the current era of exascale computing. While current supercomputing power promises revolutionary advances in many fields of science and engineering, there are important scientific and engineering questions that will remain out of reach. Examples include full understanding of climate change, increased accuracy in weather prediction, predictive design of materials, and understanding and predicting complex scientific and industrial processes. For reasons articulated below, the increased computing capability needed to make advances in these fields will no longer come from traditional approaches for scaling computer power, but will require innovations in digital microelectronics or markedly different approaches, such as quantum and neuromorphic computers. The current section is focused on the future of traditional digital microelectronics and its implications for computational science. The subsequent section focuses on the more disruptive alternatives including quantum and neuromorphic approaches.

Potentially radical changes in tomorrow's supercomputers will also require a broad set of new mathematical approaches, algorithms, and software tools. Changes will permeate the entire software stack from programming tools and numerical libraries through resource managers and system software. The computational science community can adapt to foreseeable changes in architectures, but the lack of clarity about future machines is an enormous challenge. It can take many years to develop new, optimized algorithms applicable to new computer designs. Modern computational science application codes take years to develop, and their lifespan is generally much longer than the lifespan of a single computer. So, the community must develop algorithms and software today that will be capable of running on future computers, even without clarity about what those computers will look like.

Back to TOC

# Hardware and architecture innovation for the post-exascale era

The exascale era has been enabled by a decrease in power requirements on the chip, increased heterogeneity of compute nodes with GPU accelerators, and the evolution of massive parallelism in compute nodes all incorporated in a monolithic system. However, the cost of designing, building, and operating future systems to provide even greater performance is becoming prohibitive. Transistor sizes have decreased to nearly the limits defined by physics, building a new semiconductor fabrication plant has become enormously expensive, and exascale-era supercomputers have the power requirements of a small city, dictating that any future hardware advances must include increased energy efficiency as a practical constraint.

New directions are possible and under development in architecture and hardware. Architectural design relates to how CMOS transistors are deployed on conventional chips. Hardware design refers to the development of the physical components that are employed in the computer. The GPU is a successful example of a specialized architecture, and it is likely that other accelerators will appear alongside evolved CPUs and GPUs.

The development of new hardware presents significant economic as well as technical challenges. The commercial viability of HPC-oriented hardware has depended on leveraging commercial hardware developed for other (larger) markets. Most recently, increases in supercomputer performance have depended on GPUs, which are commercially successful because of their applicability and use for the collectively large markets of HPC, AI, and crypto. A commercially viable future HPC market will most certainly depend either on leveraging investments in adjacent markets such as AI (risky and implying strong dependency on commercial markets) or discovering and leveraging less expensive approaches for developing new hardware. It is also important to note that computational science workloads are evolving to involve more data science and machine learning. Future HPC platforms will need to be able to address the increased complexity these developments imply for future applications.

Data movement increasingly dominates computation in both energy and time costs, motivating dramatic changes in architectures and software approaches. Potential hardware optimizations include specialized functional units and chiplets as discussed below. Software optimizations involve reducing the size of data, such as using reduced or mixed floating-point precision, lossy data compression, and selectively zeroing out weights in machine learning. These optimizations trade off accuracy for efficiency, which motivate new analytical and mathematical questions on error analysis and uncertainty quantification.

## Options for less expensive hardware

Three new hardware directions are present in the marketplace that have the potential to provide greater performance and increased energy efficiency with reduced cost: disaggregation, wafer-scale computing, and quantum computing (discussed in the next section).

### Disaggregation and chiplets

The cost, time, and expertise needed to bring a new chip to market has been a high bar to customization. An alternative lies in embedded systems on chip, which for many years have been customized for specific applications by combining often-reusable components (or “chiplets”) with new custom subsystems. To allow this trend to evolve and broaden, the Open Compute

Back to TOC

effort[4,5,6], a collaboration that cuts across companies and laboratories, is working toward standards of interoperability that would add available off-the-shelf components, reduce cost and risk, and generally lower the barriers to custom designs for application spaces including HPC. This may provide opportunities for the HPC field, the DOE, and perhaps eventually research teams to accelerate their computational science research through significantly cheaper and more powerful computers made possible by custom design.

### Wafer scale and near-memory processing

Recent development of wafer-scale devices[7] brings processing, fast SRAM memory, and fast, low latency on-wafer communication to a large (eight-inch square) wafer device having multiple petaflop compute capacity along with tens of gigabytes of memory. Most important, because of the locality of memory, the use of SRAM, and integration of the network with the processor, there is an attractive ratio of compute speed to memory and network bandwidth and single cycle memory and per hop network latency. Thus, wafer-scale integration achieves most of the long-promised benefits of near-memory computing, or taking the converse view, near-compute memory. The fast, low-latency communication that is possible on a wafer solves the data movement problem that occurs when processing is moved into DRAM devices and data is spread across these devices with low communication bandwidth between them. Although the main current use case is training neural networks for AI[8], recent exploratory investigations in combustion models[9], neutron transport[10], and seismic imaging[11] have shown hundredfold speedups over GPUs on modest-size problems with a similar reduction in energy cost. Much research remains to be done in applied mathematics and computer science to clarify the value of this approach for large-scale computational science.

## Architectures for complex workflows

Big data computing involves streaming massive data volumes in and out of compute systems with, frequently, insufficient computational intensity to tolerate the I/O rates of current systems. New tighter integration of optical communication with terabit per second per fiber bandwidth into compute systems, involving transducers and memory systems capable of sourcing or sinking these bandwidths, can open new and important use cases.

New applications will involve a mixture of simulation, data science, and machine learning—running on computers with specialized accelerators for some of these functions. Instead of having a single system composed of essentially the same components in each node, pools of CPUs, GPUs, memory, storage, and other accelerators will be available in a disaggregated form and composed for a particular application workflow. To achieve this, smart and fast networks will be required along with new algorithms and tools for resource and workflow management. Fundamental mathematical formulations and new algorithms will be needed to take maximum advantage of these emerging architectures.

---

4 Open Compute Project

5 Open Compute Project Pushes Fast Forward on an Open Chiplet Economy

6 RISC-V Moving Toward Open Server Specification

7 Cerebras

8 *GenSLMs: Genome-scale language models reveal SARS-CoV-2 evolutionary dynamics.* Maxim Zvyagin, et al. **The International Journal of High Performance Computing Applications**, Oct 2023

9 *Fast Stencil-Code Computation on a Wafer-Scale Processor*. Kamil Rocki et al. **Proceedings of SC 20**

10 *Efficient Algorithms for Monte Carlo Particle Transport on AI Accelerator Hardware.* John Tramm, et al. **Computer Physics Communications**, in press

11 *Massively Distributed Finite-Volume Flux Computation.* Ryuichi Sai, et al. **Proceedings of SC 23**

Back to TOC

## Longer-term technical directions

Some electronics researchers have responded to the impending end of Moore's Law by looking into alternatives to the established silicon orthodoxy. There is fascinating work being done in packaging, devices, and materials that provide possible avenues for progress beyond the end of CMOS scaling and perhaps a revival of exponential growth (in performance per unit power)[12]. These technologies include memristive memory and computing, analog representations, various "neuromorphic" technologies, carbon nanotube devices, multiple layers on a chip, integration of logic with memory, chip and wafer stacking, and others. Although each of these ideas has merit, many of them will be very challenging to integrate into existing manufacturing processes.

In summary, as Moore's Law comes to an end, further performance improvements will require new technologies and computer designs. There are many ideas that might become commercially important, but it is very difficult to discern which ideas will succeed from those which will not. It is clear that tomorrow's high performance computers will necessarily be different from today's, but the characteristics of those machines are not at all clear. For the computational science community, this requires a strategy of sustained engagement and continual (re-)evaluation, while working to minimize the disruptive impact of new architectures on existing software.

**Recommendation 3.1: Investments are needed in research and development collaborations between computational scientists and computer vendors to ensure development of future energy-efficient computers that meet the needs of the computational science community.** The high performance computing market is dominated by AI and Cloud, but there is still enough commonality that computational science can profit from developments in the larger market. Incentives are needed to encourage vendors to consider the needs of scientific applications beyond their own markets. In this emerging environment, DOE should continue to leverage its successful history of investments in holistic co-design collaborations aimed at advancing the scientific computing market.

**Recommendation 3.2: Investments are needed in research into methods to insulate applications from uncertain changes in future computing architectures.** These investments should include new methods and algorithms, modular software design, abstraction layers, and research in scientific software engineering.

**Recommendation 3.3: A facility or facilities should be established that will ensure that computational science researchers have access to emerging hardware, programming models, and heterogeneous systems to enable assessment of their utility for scientific applications.** With the plethora of emerging computing designs, it is difficult to know which are the most promising, how they can be best exploited, and which application areas are best suited to each. A program to provide broad access to new technologies, and to assist new users, will allow the community to identify the most promising approaches. A similar DOE program in the late 1980s and early 1990s supported (in part) under the federal High Performance Computing and Communications (HPCC) Program was essential for the successful transition from vector computers to parallel computers.

---

12 Aly, M.M.S., et al. Energy-Efficient Abundant-Data Computing: The N3XT 1,000X," *IEEE Computer,* Vol. 48 (2015), 24-33

Back to TOC

# 4. Potentially transformational computing technologies are on the horizon

## Summary

*Beyond traditional computer architectures, several disruptive alternatives are maturing and have the potential for impact in the next decade. This section reviews two such approaches, neuromorphic computing and quantum computing. Both technologies have compelling advantages over traditional approaches for a subset of important applications. But both also have considerable limitations. Investments in these areas are high-risk and high-reward and are an important part of a balanced portfolio in computational science.*

**Finding 4.0: Advances in traditional silicon-based computer architecture will likely be insufficient to address important scientific challenges of the future. Investments in alternatives to traditional computing, such as neuromorphic and quantum computing, will be necessary to achieve these goals.**

As was discussed in Section 3, advances in silicon CMOS microelectronics have slowed considerably due to fundamental engineering and manufacturing constraints. Future performance increases from traditional computers will be harder to come by and will likely require significant changes in computer architectures with disruptive consequences for existing software. Researchers have proposed several other technologies for digital computing with performance and power advantages, e.g. carbon nanotubes and cryogenic computing. Most of these technologies remain in their infancy and would require profound changes to manufacturing processes or computer machine rooms. But two technologies have credible paths to impact computational science over the coming decade: neuromorphic computing and quantum computing.

## Neuromorphic computing

As its name suggests, neuromorphic computing involves computing approaches that are inspired by biological brains. The extraordinary progress in deep learning in recent years is an exemplar of neuromorphic computing. The commercial promise of this technology has inspired firms big and small to design and market specialized accelerators for various aspects of machine learning (ML). The scientific computing community will benefit from ML, and specialized ML accelerators will be an important part of the high performance computing landscape for the foreseeable future. Beyond ML, these low-power, high-throughput accelerators may also be able to accelerate other elements of the computational science workflow. This is an important and active area of research.

The commercial market for ML computing is quite large, so limited opportunity exists for the computational science community to shape the landscape. But well-targeted investments could help ensure that these devices include computational science as one of their target customers. This will require close collaboration and codesign spanning hardware, software, and applications.

Current ML devices are built out of traditional microelectronics, but analog devices employing different physics (e.g. memristors) have the potential to dramatically reduce energy consumption and increase performance. The broader utility of these devices is unclear, but their potential makes them worthy of continued investigation.

Back to TOC

# Quantum computing

Richard Feynman's visionary proposition of quantum computers underscored the prospect of taming quantum many-body problems without succumbing to the curse of dimensionality. This proposition not only laid the groundwork for quantum computing, but also spotlighted the paradigm's potential to transcend today's limits of classical computing. Quantum computing is a fundamentally different approach from classical computing. It has the potential for significant speedups of some important algorithms for fields ranging from quantum chemistry and many-body physics to broader areas such as drug discovery, finance, and machine learning. Quantum devices are currently very fragile and error prone, but academia and industry are making steady progress in the design and fabrication of quantum computers. Combined with rapid advances in algorithms and applications, quantum computing has the potential to be a significant contributor to computational science within the next decade.

The rapid progress in quantum computing provides the scientific computing community with unprecedented opportunities. From an applied math perspective, quantum algorithms perform a sequence of matrix-vector multiplications using only unitary matrices. However, many scientific computing tasks are not formulated as multiplications of unitary matrices. Over the past few decades, with particularly exciting progress in recent years, ingenious methods have been devised to adapt non-unitary operations and express them in terms of unitary operations. Many of these advancements draw significantly from classical numerical analysis and numerical linear algebra, incorporating aspects like approximation theory and matrix decomposition techniques. The emphasis on unitary operations also provides fertile ground for novel numerical analysis and methodological developments.

Quantum algorithms promise to transform our problem-solving methods for a broad range of application areas, such as those mentioned above. Such transformations will require reformulating existing problems in terms of unitary matrices, such as Hamiltonian simulation problems. There may also be other innovative approaches that offer new directions and insights. Realizing these possibilities will demand substantial research in fundamental quantum-aware applied mathematics and computational science, as well as efforts to optimally map these new methods onto the various quantum systems being developed. Just as the transition from mainframes and vector computing to massively parallel computing systems required significant investment in numerical methods and algorithms research, so too will the quantum era require a similar level of dedication and innovation of research in applied mathematics and computational sciences.

Yet, realizing this potential is challenging. It demands thorough exploration of quantum speedups, the creation of new quantum techniques, and robust error correction and fault-tolerance strategies. To effectively utilize quantum capabilities, it is important to understand the true quantum cost. While it might be tempting to think that n qubits can represent an equivalent of 2n classical bits of information, suggesting exponential quantum speedups, all quantum algorithms will need to interface with classical processing. So, the quantum complexities should factor in the input-output models and specific needs of quantum algorithms. The focus should be on problems where results are obtained from a limited quantum measurement set, ensuring both practicality and optimized performance. Quantum computers might not bring about universal speedups but could act as specialized accelerators, similar to GPUs. They are meant to complement, not replace, classical computers.

Broadly, the quantum cost can be broken down into three main parts: input, running, and output costs.

1. *Input cost* refers to the cost of preparing the initial quantum state. Typically, a quantum algorithm starts with a standard state, and the input state for the quantum algorithm is prepared using a unitary

Back to TOC

matrix acting on the standard state. The input cost then is the gate complexity to achieve this input state. If a quantum algorithm

requires repeated access to the input state in a coherent manner, the input cost needs to be multiplied by the number of initial state preparations.

2. *Running cost* refers to the expense of executing the quantum algorithm a single time, not taking into account the cost for the input state. The running cost is typically the element most amenable to theoretical analysis.

3. *Output cost* refers to the number of times the quantum algorithm needs to be executed to perform quantum measurement on one or multiple qubits.

Considering the input, running, and output costs offers a comprehensive "end-to-end" analysis. Demonstrating a quantum advantage also requires comparing the quantum cost with that of classical solvers.

As quantum technology progresses, the spectrum of potential applications has been expanding notably. Many of these applications cater to inherently classical challenges, such as quantum solvers designed for classical differential equations, linear systems, optimization problems, and sampling tasks. The scope of scientific computing problems that could benefit from quantum algorithms is extensive, promising a wide range of opportunities. However, it is important to note that not every classical problem is well-suited for substantial quantum enhancements. For example, many low-dimensional differential equations have a solution cost that scales polynomially with system size, which restricts the possibility of exponential quantum speedups. Additionally, the input and output mechanisms for high-dimensional differential equations often present considerable hurdles. To accurately measure the advantage of quantum algorithms over classical counterparts, their performance should be benchmarked against the best classical methods, particularly those adept at handling high-dimensional data structures, such as tensor methods, Monte Carlo methods, and quantum-inspired techniques. It is also important to acknowledge that conducting a full end-to-end analysis can be challenging due to the nascent state of the field. Nevertheless, the concept of an end-to-end analysis should remain a guiding principle, with prioritized research providing a strategic roadmap toward comprehensive evaluation.

Addressing errors is another vital aspect. Over the past 50 years, classical computer gates have seen significant improvements, achieving error rates as low as $10^{-13}$ or even better. In contrast, quantum gates like single qubit rotations or two-qubit CNOT gates have error rates around $10^{-3}$ or higher. This presents a challenge for Noisy Intermediate Scale Quantum (NISQ) devices and highlights the importance of effective quantum error correction (QEC) mechanisms. The advancement of QEC codes can drastically change the landscape of quantum computation.

Strategies to address these challenges include:

1. Support fundamental quantum-aware computational science and mathematical research to enable the reformulation of challenging classical problems into quantum alternatives that are more amenable to be solved by quantum algorithms.

2. Improve quantum algorithms for scientific computing problems, while aiming to address the "end-to-end" complexity. Examples include, but are not limited to, Hamiltonian simulation, eigenvalue problems, open quantum systems, non-Hermitian quantum physics, linear and nonlinear classical differential equations, optimization, and sampling problems.

3. Address challenges posed by quantum algorithms for early fault-tolerant quantum computing (beyond NISQ), considering limited quantum resources and fault tolerance capabilities.

Back to TOC

4. Develop quantum-inspired classical algorithms, along with other advanced classical algorithms, to extend the boundaries of current classical computation capabilities.

**Recommendation 4.1: Investments are needed for the development of high-risk, high-reward alternatives to traditional computing hardware, such as quantum and neuromorphic computing, along with research and development of the algorithms and software that will be required to use them for scientific and engineering applications.**

**Recommendation 4.2: Investments are needed to advance quantum computing through the improvement of mathematical understanding and algorithm development.** These investments should be focused on the development and refinement of quantum algorithms for a broad range of unsolved scientific computing challenges, aiming at addressing end-to-end complexity in quantum computing applications.

**Recommendation 4.3: A facility or facilities should be established to ensure that researchers have access to experimental non-traditional hardware that will inform a holistic codesign cycle between manufacturers and users of transformative computing technologies.** As the non-traditional computing hardware industry evolves and the roles of different enterprises change, the government should remain open to establishing new types of collaborations and relationships with both existing and emerging industrial partners. These new devices will likely be initially employed as accelerators, so their ability to integrate with existing HPC platforms must be part of the design process.

A quantum computer at Lawrence Berkeley National Laboratory is exploring the potential for quantum phenomena to enable groundbreaking computational power.

Back to TOC

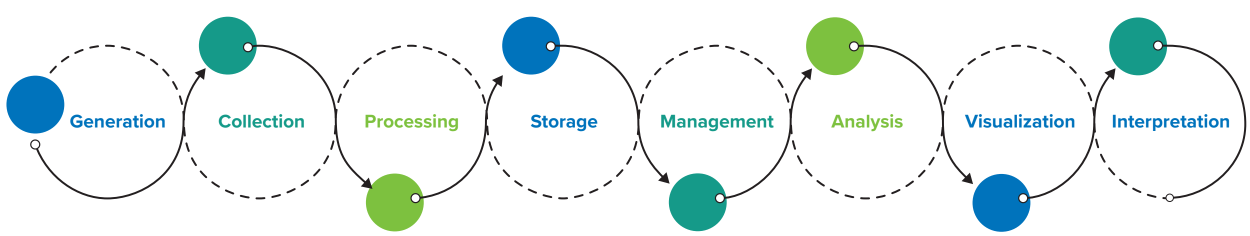
# 5. Scientific progress increasingly relies on vast amounts of data

## Summary

*A new generation of scientific facilities is providing unprecedented volumes and quality of data. In the broader world, ubiquitous sensors are providing novel data about earth systems and social systems. And ever larger and more complex simulations generate vast amounts of data. All these very different kinds of data can provide important insight into science and decision making. Consequently, managing and leveraging data is now a central element of computational science. Scientific data science requires capabilities that are distinct from those required for traditional simulation. As detailed in this section, these capabilities include a broad tool set for collecting, processing, storing, and analyzing data. Significant research investments are needed in mathematics and computer science to create, build, and apply these capabilities and to gain maximum value from existing and emerging data streams.*

**Finding 5.0: Scientific progress increasingly relies on vast amounts of data. Broad investments will be required to ensure that scientific advances will continue to be fueled by the increasingly large and complex data streams produced by scientific instruments and facilities.** These investments must both leverage current exascale and edge computing technologies and be able to take advantage of future computer architectures as they are developed.

Our world is awash with sensors that can be used for societal good like weather forecasting, traffic routing, and environmental monitoring. New scientific facilities are generating unprecedented amounts of data. Simulations themselves generate enormous outputs. These data, separately and combined, can be used to answer important questions in applied and fundamental science ranging from better battery materials to the health of our planet to the origins of the universe. Modern data science requires sophisticated mathematical and computational tools. Data needs to be managed, processed, organized, stored, analyzed, and interpreted to produce insights for science and for society. Thus, new approaches are needed across the data lifecycle (See figure below) for sensor development, efficient data storage, safe and secure management, and novel techniques for analysis and visualization. Open research questions remain around analyzing highly distributed data sources, enabling data discovery and integration, tracking data provenance, coping with sampling biases and heterogeneity, ensuring data integrity, privacy, security, and sharing, and visualizing massive datasets. Integrated research and development (R&D) are needed throughout the full pipeline of the data lifecycle, informed and shaped by the specific needs of the relevant scientific communities.




Lifecycle of data. New types and quantities of data are reshaping science.

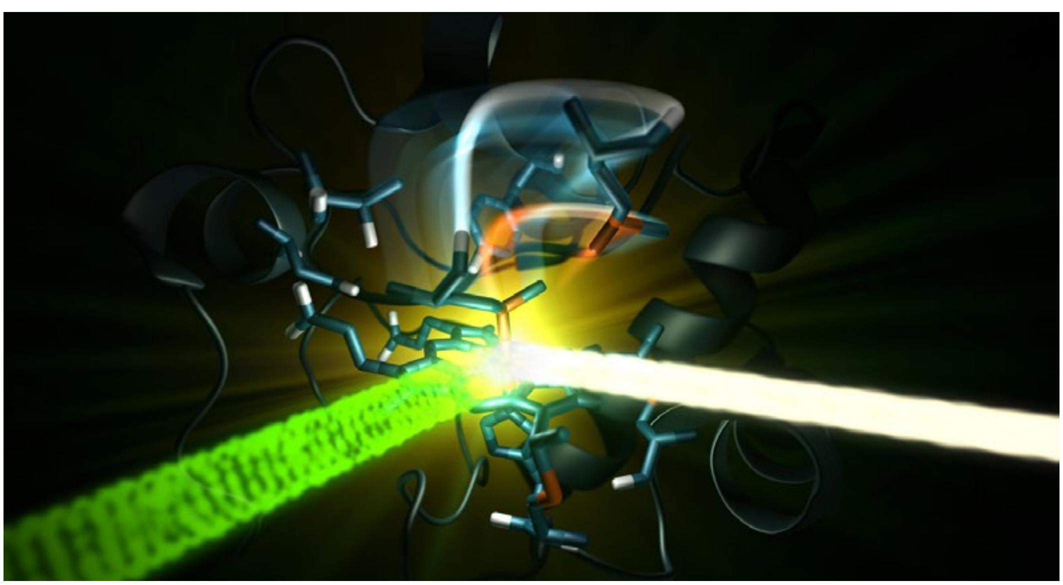
Back to TOC

DOE's national laboratories provide researchers with access to the largest and most diverse suite of scientific experimental facilities in the world—from x-ray synchrotrons and neutron sources to integrative genomics and atmospheric radiation facilities—as well as to the world's most capable high performance computing facilities. Upgrades to these user facilities and the advent of new facilities coming online now and over the next decade will dramatically increase the amount and complexity of new data produced. Given the specialization of many of these facilities and the scientific questions they will help answer, it will not be possible to rely on general advances in data science to fully support the needs associated with growth in data. The DOE will need to build its own R&D programs that focus on science-driven questions and the required computing capabilities to answer them.

An important example is the opportunity to build on DOE's co-design experience by enabling application scientists, software infrastructure developers, data science researchers, and hardware specialists to define and develop a common software stack for data science computing resources at the different user facilities. The software infrastructure should support some generic services but also allow the creation of specialized applications specific to the facility.

Data science can also enable breakthroughs in operating complex infrastructure. To realize this potential, new paradigms must be developed and tested to integrate real-time information from sensors with nearby computing resources (known as "edge computing"), to enable real-time predictive analytics, control, and optimization. Ideally these models will support particle accelerators, light sources, and complex instruments, many of which involve interconnected subsystems of magnets, mechanical, vacuum, and cooling equipment, power supplies, and other components. Such instruments have many control points, and require high levels of stability, making their operation a complex optimization problem. It is a challenging, ongoing problem to develop models for reliable and safe control.

The application of data fusion in technical systems requires mathematical and heuristic techniques from fields such as statistics, AI, operations research, digital signal processing, pattern recognition, cognitive psychology, information theory, and decision theory. In scientific scenarios, data fusion sensors can be used to observe electromagnetic radiation, acoustic and thermal energy, nuclear particles, infrared radiation, noise, and other signals.



Using intense X-ray pulses from SLAC National Accelerator Laboratory's Linac Coherent Light Source, scientists can determine the structures of proteins from tiny nanocrystals, including proteins important in disease and its treatment. The analyses of these data sets require advanced data science and supercomputing capabilities.

Back to TOC

Multisensor data fusion is a relatively new engineering discipline used to combine data from multiple and diverse sensors and sources to make inferences about events, activities, and situations. These systems are often compared to the human cognitive process where the brain fuses sensory information from the various sensory organs, evaluates situations, makes decisions, and directs action. Current sensor fusion solutions perform object-level fusion wherein each sensor with its inherent limitations identifies and classifies objects individually. This results in poor performance and is not optimal because no single sensor is capable by itself of detecting all objects under all conditions. Furthermore, when sensor data is not fused, operators may get contradicting inputs from sensors and be unable to determine with a degree of certainty on the next action. New, application-informed approaches to scientific data fusion are needed.

One of the main challenges is dealing with data that is heterogeneous in nature, such as data that is in different formats, units, or scales. This requires developing methods for data harmonization and normalization, which can be complex and time-consuming. Another challenge is ensuring the quality and accuracy of the data, especially when dealing with large and complex datasets. This requires developing methods for data cleaning, validation, and error correction. Data fusion requires integrating data from different sources and modalities into a unified format that can be analyzed and interpreted, as well as advanced analytics and machine learning techniques to extract meaningful insights and patterns.

In all of these settings and many more, advances in data science and machine learning will accelerate and transform the practice of science and engineering. But scientific applications are different from commercial applications of these technologies, and considerable effort will be required to modify existing approaches, to develop entirely new approaches, and to build the requisite web of complex software tools.

**Recommendation 5.1: Investments are needed in research to develop the mathematics and computer science technologies supporting multisensor data fusion.** Development of these capabilities will require research in fields such as mathematics, statistics, AI, operations research, digital signal processing, pattern recognition, cognitive psychology, information theory, and decision theory.

**Recommendation 5.2: Investment is needed to define and develop a common software stack for data science edge computing resources at scientific user facilities**. This will include developing and testing new paradigms for integrating real-time information from sensors with edge computation, enabling real-time predictive analytics, control, and optimization in support of operation of complex infrastructure at DOE scientific user facilities. The software infrastructure should support all aspects of scientific user facility operations, from generic operation and analysis services to specialized software applications specific to each facility.

**Recommendation 5.3: Investment is needed in research and development to create an integrated suite of data lifecycle methods and tools, informed by the specific needs of DOE scientific communities.** Techniques that lead to data harmonization and normalization, such as data cleaning, validation, and error correction must be developed to overcome the natural heterogeneity of data sources which is in inherent conflict with the need for data fusion. Open research questions remain around analyzing highly distributed data sources, enabling data discovery and integration, tracking data provenance, coping with sampling biases and heterogeneity, ensuring data integrity, privacy, security, and sharing, and visualizing massive datasets.

Back to TOC

# 6. Modern artificial intelligence is transforming the scientific landscape

## Summary

*Extraordinary recent advances in artificial intelligence (AI) and machine learning (ML) have showcased the potential for these technologies to disrupt many aspects of society. The field is being driven by industry, with a focus on commercial applications. But it is clear that AI and ML also have enormous potential to accelerate scientific discovery. The field of scientific ML is quite young, and the underlying technologies are changing rapidly, so it is difficult to see the future with great clarity. A successful research program will need to remain agile and adaptive. To complement and leverage the large industrial investment in these areas, DOE should focus on the unique needs associated with scientific applications and with the needs of DOE's broader missions.*

**Finding 6.0: Artificial intelligence and machine learning have enormous potential to impact the processes of scientific research, but broad investments in mathematics and computing will be required to realize these opportunities.** In some cases, AI and ML capabilities developed by industry can be applied to science, but in many cases, significant research will be required to make these capabilities usable in the scientific domain. In particular, DOE should focus on the unique needs associated with scientific applications and with the needs of DOE's broader missions. These investments must both leverage current exascale and edge computing technologies and be able to take advantage of future computer architectures as they are developed.

In just the past decade, explosive progress in AI and ML have created capabilities that are changing numerous areas of commerce and everyday life. These capabilities are also showing the potential to dramatically accelerate progress in science, and there is much yet to be discovered. Many scientists are exploring ideas and developing new insights, but the field of scientific machine learning is very young. Many existing applications involve the use of ML models as very fast "surrogates" (approximate replacements) for more expensive simulation codes. The speed of these ML surrogates allows for more instances to be run, and for a larger search space to be explored which can lead to more optimized designs or better characterizations of uncertainties. Alphafold[13], Google's breakthrough technology for predicting protein conformations, showcased the potential for ML to outperform traditional scientific approaches. Other potential roles for ML in science include automatically monitoring and running experiments or series of experiments, improved methods for combining simulation results and experimental data, and the generation of new hypotheses from data. Quite likely, the most important ideas have not yet been thought of. AI may also dramatically change the software development process, which might make it much easier to build new computational science tools and applications.

Modern methods in AI, defined broadly to include ML, optimization, statistical inference, and supporting systems, are yielding unprecedented results in many application areas. Recent AI successes can be attributed to huge datasets, enormous computing power, and innovations in the underlying mathematics, statistics, and algorithms. Scientific applications of AI can have quite different characteristics than commercial applications,

13 AlphaFold Protein Structure Database

Back to TOC

so existing methods may need to be modified and new methods developed for scientific needs. Scientific datasets may be limited in size and are seldom labeled with the information required for many AI/ML algorithms. The best use of algorithms for scientific ML may involve new methods that take into account physical principles and constraints. Many scientific and engineering applications are high-consequence and require high accuracy and a careful assessment of uncertainties. The scientific community needs new algorithms and approaches to meet these challenges, and progress will have profound impact on many areas of science.

Some of the foundational areas to explore include:

- **Effective ML training for small or sparse datasets.** Many scientific areas, ranging from materials science to astronomy to biology, have very few datasets with labeled training data. Many current AI algorithms, for example, deep neural networks, cannot perform well when trained on these limited samples. New developments in foundational AI, such as few-shot learning, self-supervised learning, mixed-scale dense networks, and meta-learning techniques, need to be developed to enable important scientific AI applications.

- **Incorporation of physical models into AI structure**. The current theory and practice of AI struggles to incorporate physical models and constraints. Formulations of new AI algorithms structured to include physical models, rather than having to "learn" the physics, will allow the designs of new classes of AI applications optimized exactly for the scientific tasks that require them. To do this will require advances in such formalisms as projection operators that enforce physical principles, data structures that encode symmetries and constraints, and the construction of physical priors and their injection into mathematical models.

- **Verification, validation, and uncertainty quantification**. DOE has been a leader in verification, validation, and uncertainty quantification (VVUQ) for computational models, where verification determines the accuracy and correctness of a code's output, and validation determines the degree to which a model represents the real world. Augmenting computational models with data, AI, and ML algorithms opens new VVUQ challenges that will require a comprehensive framework for assessing the uncertainty associated with AI predictions and leveraging this knowledge to develop better predictors.

- **Real-time decision making**. Many facilities in the DOE portfolio and beyond generate vast quantities of heterogeneous data that must be analyzed in real-time. Further work is required to achieve sub-microsecond decision times for applications such as particle physics, real-time particle accelerator or fusion reactor control, or the kinds of processing required for applications such as radio astronomy. Real-time decision making can also be used to optimize the use of scientific experiments and facilities and to manage key systems like the power grid.

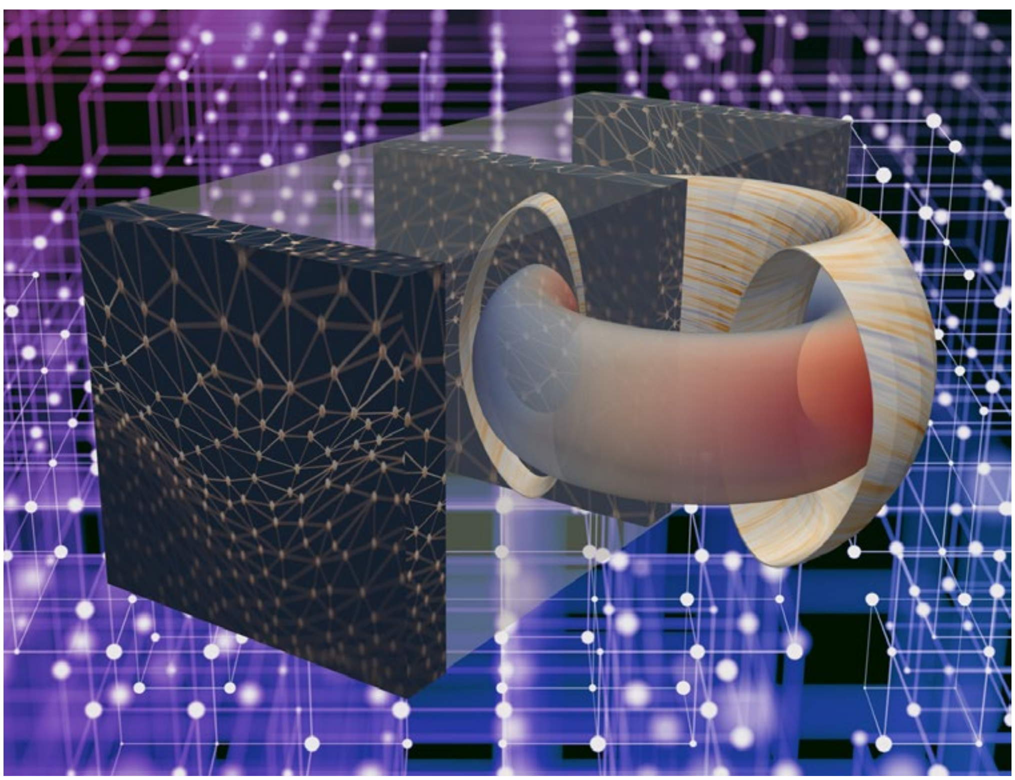
- **AI methods for management of computational resources and workflows**. Computer systems employed in high-end scientific applications have become extremely complex. Computer architectures have deep hierarchies with multicore CPU and GPU components, are often heterogeneous, and are increasingly integrated into distributed systems that incorporate edge computing. AI methods are being developed to schedule and coordinate execution of workflows and to choreograph applications that include capture, reduction, assimilation, and analysis of streaming data.

- **Interpretable models and algorithms**. Many AI models focus on prediction accuracy without a focus on explainability. For scientific applications, prediction without insight is seldom enough. A key area of exploration is the development of interpretable models and algorithms that provide insight on why and how a model is producing a particular output.

**Recommendation 6.1: Investments should be made in fundamental research to develop AI technologies that can advance DOE science.** This could include efforts through existing programs and new focused opportunities that enable research collaborations among applied mathematicians, computer scientists, and computational scientists. Possible examples include the following:

- effective ML training for small or sparse datasets;
- incorporation of physical models into AI structure;
- verification, validation, and uncertainty quantification;
- real-time decision making;
- AI methods for management of computational resources and workflows;
- AI methods for planning, monitoring and running experiments;
- interpretable models and algorithms; and
- hypothesis generation from data.

**Recommendation 6.2: Investments should be made to develop partnerships between computational and applications scientists to customize and apply AI capabilities to new science areas.** As with all areas of science, active collaboration between science domain experts and computational experts will ensure that the best possible computational capabilities can be developed and brought to bear on challenging scientific problems. The SciDAC program provides an attractive model for this kind of investment.



An AI/Deep Learning/Machine Learning Project at the Princeton Plasma Physics Laboratory has enabled the FRNN code to predict the onset of powerful instabilities known as "disruptions" in magnetic confinement fusion devices ("tokamaks"). This is an important step towards a system to control plasmas for fusion power generation.

Back to TOC

# 7. Workforce

## Summary

*In industry, academia, and the national labs, there is high and growing demand for workers with expertise in computational science. The current workforce is insufficient to meet these needs, and demands will continue to grow as employers embrace the transformational opportunities discussed above. Current pipelines are tapped out, so new initiatives are needed to create pathways for historically underrepresented communities. Retraining programs are also needed, e.g. in artificial intelligence and machine learning, for workers with skills in adjacent areas. As an inter- or even multidisciplinary field, training in computational science requires specialized educational programs that cut across traditional academic departments and continue in the workplace.*

**Finding 7.0: The current workforce is insufficient in both size and diversity to meet the national need for computational science expertise in industry, academia, and the national labs. Active steps must be taken to create pathways into computational science for historically underrepresented communities and existing workers with skills in adjacent areas.**

High performance computing (HPC) plays a "vital role in driving private-sector competitiveness"[14] driven by advances in computing technology and computational science and engineering (CSE). The rapidly changing HPC environment is creating new challenges and exciting opportunities for CSE professionals. Thus, there will be an intensifying demand for a well-trained, increasingly diverse CSE workforce that can continually adapt and grow to meet these challenges, and that can take advantage of the opportunities for new and exciting advances in science and engineering that will result.

In a workforce environment where industry rather than the government sector increasingly draws the brightest minds in computer science and mathematics, it will be important to attract talented young researchers into the CSE workforce at government laboratories and in academia. The U.S. Department of Energy (DOE) has played a critical role in the development and use of CSE and HPC, maintaining an international leadership position. However, it has been noted by the DOE hallenges in workforce development and recruitment. This is exacerbated due to the inter- and multidisciplinary nature of the work and the reliance on an understanding of advanced and high-performance computing[15]. Thus, it will be increasingly important to draw from an expanded base of talent, which makes recruiting from historically underrepresented communities even more important.

Creating new workforce pipelines that expose students to scientific computing earlier in their careers is vital to expand the pool of practitioners. DOE's Workforce Development for Teachers and Scientists (WDTS) has been successful at collaborating with the national laboratories to engage with educators and undergraduate, graduate, and postdoctoral students. However, with the changing landscape in scientific computing, more investment is needed to address the growing workforce issues at all levels. Exposure to DOE has proven to be an effective way to inspire students at all levels to pursue careers in the field and ensure

---

14 Council on Competitiveness, "Advance: Benchmarking Industrial Use of High Performance Computing for Innovation," 2008. [Online]. Available: https://compete.org/2008/03/16/advance

15 Roscoe Giles, et al. 2020. Transforming ASCR after ECP. https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202004/Transition_Report_202004- ASCAC.pdf

Back to TOC

Department of Energy (DOE) Computational Science Graduate Fellowship (CSGF) scholars and alumni at the 2023 DOE CSGF Annual Program Review in Washington, D.C. The CSGF scholars are part of an innovative group learning to use high performance computing to solve scientific and engineering problems.

they gain the requisite skills. The national labs provide these opportunities but are a finite resource; additional programs at educational institutions to attract and train the next generation of the CSE workforce will help to alleviate the current shortages.

**Recommendation 7.1: DOE should invest in programs at diverse educational institutions to increase the reach of workforce development programs.** One model DOE could follow is the National Science Foundation Research Experiences for Undergraduates, which funds students to work on research projects at the host institution. DOE should also consider additional K–12 engagement to ensure students are aware of and interested in CSE fields as they make college decisions. To support the workforce needs of the future, students need to be exposed early and often to DOE, and partnering with other institutions will multiply the impact of this effort.

Over the past decades, workforce development programs have been created that aim to build a skilled CSE workforce. A principal example is the DOE Computational Science Graduate Fellowship[16], which has enabled students to pursue a multidisciplinary program of education in CSE coupled with practical experiences at the DOE laboratories. Graduates of this program have pursued professional careers in industry, academia, and at federal laboratories. More recently, the ECP Broadening Participation Initiative [17], including the Sustainable Research Pathways program[18, 19,] has successfully brought hundreds of faculty and students from underrepresented communities to the DOE national laboratories for summer research experiences and HPC training. Continuation and expansion of these programs, and the development of additional programs modeled after them, will be essential for building the diverse CSE workforce needed to address the challenges and take advantage of the opportunities discussed in this report.

---

16 Brown, D, Hack, J, Voigt, R. The Early Years and Evolution of the DOE Computational Science Graduate Fellowship Program. Computing in Science & Engineering. 2021 November 1; 23(6):9-15. Available: https://ieeexplore.ieee.org/document/9580661/ DOI: 10.1109/MCSE.2021.312068)

17 The ECP Broadening Participation Initiative: https://www.exascaleproject.org/hpc-workforce

18 Brown, DL, Crivelli, S, Leung, MA. Sustainable Research Pathways: Building Connections across Communities to Diversify the National Laboratory Workforce. CoNECD 2019 - Collaborative Network for Engineering and Computing Diversity; 2019; c2019. source-work-id: 3228546

19 Sustainable Horizons Institute: https://shinstitute.org

Back to TOC

Computational science has always been an interdisciplinary endeavor, drawing on mathematics, computer science, and application domains. With the growth in the importance of data science and AI discussed elsewhere in this report, the field is becoming even broader. Interdisciplinary education has long been challenging due to the departmental structure of universities. But recent years have seen a steady growth in cross-campus programs that facilitate interdisciplinary education, collaboration, and research. Many of the most pressing scientific and societal challenges can only be solved through the efforts of multidisciplinary teams. The federal government should create incentives for interdisciplinary education and collaboration at both the graduate and undergraduate levels.

**Recommendation 7.2: The federal government should expand and broaden investments in workforce development in computational science.** Areas of focus should include pathways for underrepresented communities, retraining opportunities for existing workers, and academic programs with strong interdisciplinary elements.

**Recommendation 7.3: DOE should continue to support, and look for opportunities to expand, the Computational Science Graduate Fellowship.** The CSGF program has been integral in ensuring early career researchers have the skills and experiences they need to effectively contribute to DOE. As the future of scientific computing continues to evolve, the program should be supported in expanding to new areas, such as quantum computing and artificial intelligence for science.

Back to TOC

# 8. Findings and recommendations

## SECTION 1 | Introduction

**Finding 1.0: Given the central role of computational science in scientific discovery, industrial competitiveness and national security, the federal government must make the necessary investments to ensure continued U.S. leadership in the field.**

**Recommendation 1.1: Investments should be made to support a comprehensive computational science program that leverages the results of the Exascale Computing Program (ECP) and anticipates and exploits future high performance computing platforms. To fully take advantage of ECP technology and future hardware, this program will require further advances in applied mathematics and computer science and the establishment of partnerships between applied mathematicians, computer scientists, and application scientists to address critical national challenges.** Exascale hardware and software provide a foundation for discoveries through scientific computing, but further research and development is required to fully realize the potential of exascale. In addition, further advances in computer performance and efficiency will come from computer designs that will require new algorithms and software. Current support for mathematics and computer science research aimed at enabling the use of future computers is insufficient to ensure they will be usable for science in a timely way. It is essential that a comprehensive program be implemented to develop the mathematics and computer science and to support the partnerships needed to enable application scientists to exploit these computers to address critical future challenges.

## SECTION 2 | Computational science opportunities

**Finding 2.0: DOE's recent investments have produced supercomputers, software tools, and applications with exascale capabilities. Similarly, recent mathematical and algorithmic advances have led to novel approaches for answering difficult scientific and engineering questions. A comprehensive investment strategy will be required to ensure that these capabilities can continue to advance and be leveraged to develop new pathways of scientific discovery and engineering practice.**

**Recommendation 2.1: Further investments are needed to support the research and development activities needed to fully leverage the promise of exascale computing.** These investments should include mathematics, computer science, application science, and system software. Exascale applications will require methodological advances in multiphysics simulations, uncertainty quantification, design optimization, and conjoining artificial intelligence with simulation. Managing exascale simulations will require new approaches to system administration, workflow management, data analysis, and visualization. All of these needs can only be met with additional research investments from the federal government.

**Recommendation 2.2: An explicit program should be established to leverage and continue to advance the community software created by the Exascale Computing Project.** This body of software developed under ECP will underpin the community for many years to come and will provide a cost-effective way to develop, support, port, and maintain a host of advanced applications. But this software will not be able to address continually evolving needs without further investment. DOE should oversee a research and development program to continue to extend this functionality, adapt it to new computers, and apply it to new application areas.

Back to TOC

**Recommendation 2.3: Investment is needed to provide a strong scientific foundation for the development and use of digital twins.** This includes furthering the individual understanding of data assimilation and inverse problems, optimal control and model-based decision making, surrogates and model reduction, validation, and uncertainty quantification. In addition, understanding must be developed of the new challenges that arise when these constituent problems are integrated into digital twin frameworks.

**Recommendation 2.4: Investments are needed in the research, software tools, and system management tools needed to enable complex workflows that combine simulation with machine learning.** The computational science community should leverage recent progress in AI and ML in industry and identify AI and ML research directions that will enable new and faster scientific and engineering capabilities. Many existing applications involve the use of ML models as very fast "surrogates" (approximate replacements) for more expensive simulation codes.

## SECTION 3 | Future architectures, traditional microelectronics

**Finding 3.0: The nation needs a suite of investments to ensure that the development of high performance computing continues beyond exascale to meet the nation's continually evolving needs for advanced computational science.**

**Recommendation 3.1: Investments are needed in research and development collaborations between computational scientists and computer vendors to ensure development of future energy-efficient compute platforms that meet the needs of the computational science community.** The high performance computing market is dominated by AI and Cloud, but there is still enough commonality that computational science can profit from developments in the larger market. Incentives are needed to encourage vendors to consider the needs of scientific application beyond their own markets. In this emerging environment, DOE should continue to leverage its successful history of investments in holistic co-design collaborations aimed at advancing the scientific computing market.

**Recommendation 3.2: Investments are needed in research into methods to insulate applications from uncertain changes in future computing architectures.** These investments should include new methods and algorithms, modular software design, abstraction layers, and research in scientific software engineering.

**Recommendation 3.3: A facility or facilities should be established that will ensure that computational science researchers have access to emerging hardware, programming models, and heterogeneous systems to enable assessment of their utility for scientific applications.** With the plethora of emerging computing designs, it is difficult to know which are the most promising, how they can be best exploited, and which application areas are best suited to each. A program to provide broad access to new technologies, and to assist new users, would allow the community to identify the most promising approaches. A similar program in the late 1980s and early 1990s was essential for the successful transition from vector computers to parallel computers.

## SECTION 4 | Future architectures, transformational

**Finding 4.0: Advances in traditional silicon-based computer architecture will likely be insufficient to address important scientific challenges of the future. Investments in alternatives to traditional computing, such as neuromorphic and quantum computing, will be necessary to achieve these goals.**

Back to TOC

**Recommendation 4.1: Investments are needed for the development of high-risk, high-reward alternatives to traditional computing hardware, such as quantum and neuromorphic computing, along with research and development of the algorithms and software that will be required to use them for scientific and engineering applications.**

**Recommendation 4.2: Investments are needed to advance quantum computing through the improvement of mathematical understanding and algorithm development.** These investments should be focused on the development and refinement of quantum algorithms for a broad range of unsolved scientific computing challenges, aiming at addressing end-to-end complexity in quantum computing applications.

**Recommendation 4.3: A facility or facilities should be established to ensure that researchers have access to experimental non-traditional hardware that will inform a holistic codesign cycle between manufacturers and users of transformative computing technologies.** As the non-traditional computing hardware industry evolves and the roles of different enterprises change, the government should remain open to establishing new types of collaborations and relationships with both existing and emerging industrial partners. These new devices will likely be initially employed as accelerators, so their ability to integrate with existing HPC platforms must be part of the design process.

## SECTION 5 | Data

**Finding 5.0: Scientific progress increasingly relies on vast amounts of data. Broad investments will be required to ensure that scientific advances will continue to be fueled by the increasingly large and complex data streams produced by scientific instruments and facilities.** These investments must both leverage current exascale and edge computing technologies and be able to take advantage of future computer architectures as they are developed.

**Recommendation 5.1: Investments are needed in research to develop the mathematics and computer science technologies supporting multisensor data fusion.** Development of these capabilities will require research in fields such as mathematics, statistics, AI, operations research, digital signal processing, pattern recognition, cognitive psychology, information theory, and decision theory.

**Recommendation 5.2: Investment is needed to define and develop a common software stack for data science edge computing resources at scientific user facilities**. This will include development and testing of new paradigms for integrating real-time information from sensors with edge computation, enabling real-time predictive analytics, control, and optimization in support of operation of complex infrastructure at DOE scientific user facilities. The software infrastructure should support all aspects of scientific user facility operations, from generic operation and analysis services to specialized software applications specific to each facility.

**Recommendation 5.3: Investment is needed in research and development to create an integrated suite of data lifecycle methods and tools, informed by the specific needs of DOE scientific communities.** Techniques that lead to data harmonization and normalization, such as data cleaning, validation, and error correction must be developed to overcome the natural heterogeneity of data sources which is in inherent conflict with the need for data fusion. Open research questions remain around analyzing highly distributed data sources, enabling data discovery and integration, tracking data provenance, coping with sampling biases and heterogeneity, ensuring data integrity, privacy, security, and sharing, and visualizing massive datasets.

Back to TOC

## SECTION 6 | Artificial intelligence

**Finding 6.0: Artificial intelligence and machine learning have enormous potential to impact the processes of scientific research, but broad investments in mathematics and computing will be required to realize these opportunities.** In some cases, AI and ML capabilities developed by industry can be applied to science, but in many cases, significant research will be required to make these capabilities usable in the scientific domain. In particular, DOE should focus on the unique needs associated with scientific applications and with the needs of DOE's broader missions. These investments must both leverage current exascale and edge computing technologies and be able to take advantage of future computer architectures as they are developed.

**Recommendation 6.1: Investments should be made in fundamental research to develop AI technologies that can advance DOE science.** This could include efforts through existing programs and new focused opportunities that enable research collaborations among applied mathematicians, computer scientists, and computational scientists. Possible examples include:

- Effective ML training for small or sparse datasets;
- Incorporation of physical models into AI structure;
- Verification, validation, and uncertainty quantification;
- Real-time decision making;
- AI methods for management of computational resources and workflows;
- AI methods for planning, monitoring and running experiments;
- Interpretable models and algorithms;
- Hypothesis generation from data.

**Recommendation 6.2: Investments should be made to develop partnerships between computational and applications scientists to customize and apply AI capabilities to new science areas.** As with all areas of science, active collaboration between science domain experts and computational experts will ensure that the best possible computational capabilities can be developed and brought to bear on challenging scientific problems. The SciDAC program provides an attractive model for this kind of investment.

## SECTION 7 | Workforce

**Finding 7.0: The current workforce is insufficient in both size and diversity to meet the national need for computational science expertise in industry, academia, and the national labs. Active steps must be taken to create pathways into computational science for historically underrepresented communities and existing workers with skills in adjacent areas.**

**Recommendation 7.1: DOE should invest in programs at diverse educational institutions to increase the reach of workforce development programs.** One model DOE could follow is the National Science Foundation Research Experiences for Undergraduates, which funds students to work on research projects at the host institution. DOE should also consider additional K–12 engagement to ensure students are aware of and interested in CSE fields as they make college decisions. To support the workforce needs of the future, students need to be exposed early and often to DOE, and partnering with other institutions will multiply the impact of this effort.

**Recommendation 7.2: The federal government should expand and broaden investments in workforce development in computational science.** Areas of focus should include pathways for underrepresented

Back to TOC

communities, retraining opportunities for existing workers, and academic programs with strong interdisciplinary elements.

**Recommendation 7.3: DOE should continue to support, and look for opportunities to expand, the Computational Science Graduate Fellowship.** The CSGF program has been integral in ensuring early career researchers have the skills and experiences they need to effectively contribute to DOE. As the future of scientific computing continues to evolve, the program should be supported in expanding to new areas, such as quantum computing and artificial intelligence for science.

# Appendix I: Acronym glossary

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **ASCR** | U.S. Department of Energy Advanced Scientific Computing Research Program |
| **CMOS** | Complementary Metal-Oxide Semiconductor |
| **CNOT** | controlled NOT (quantum logic) gate |
| **CPU** | Central Processing Unit |
| **CSE** | Computational Science and Engineering |
| **CSGF** | U.S. DOE Computational Science Graduate Fellowship program |
| **DOD** | U.S. Department of Defense |
| **DOE** | United States Department of Energy |
| **DRAM** | Dynamic Random-Access Memory |
| **ECP** | DOE Exascale Computing Project |
| **EPA** | U.S. Environmental Protection Agency |
| **E.U.** | European Union |
| **GPU** | Graphics Processing Unit |
| **HPC** | High Performance Computing |
| **HPCC** | (Federal) High Performance Computing and Communications Program |
| **IO** | input-output |
| **ML** | Machine Learning |
| **NASA** | The National Aeronautics and Space Administration |
| **NIH** | U.S. National Institutes of Health |
| **NISQ** | Noisy Intermediate Scale Quantum (device) |
| **NIST** | U.S. National Institute of Standards and Technology |
| **NOAA** | U.S. National Oceanic and Atmospheric Administration |
| **NSF** | National Science Foundation |
| **QEC** | Quantum Error Correction |
| **SIAM** | Society for Industrial and Applied Mathematics |
| **SciDAC** | DOE Scientific Discovery through Advanced Computing program |
| **SRAM** | Static Random-Access Memory |

Back to TOC

# Appendix II: Task Force members

**Bruce Hendrickson**
Lawrence Livermore National Laboratory (Chair)

**Alejandro Aceves**
Southern Methodist University

**Elie Alhajjar**
RAND Corporation

**Mario Bañuelos**
California State University, Fresno

**David Brown**
Lawrence Berkeley National Laboratory (emeritus)

**Karen Devine**
Sandia National Laboratories (emeritus)

**Qiang Du**
Columbia University

**Omar Ghattas**
University of Texas, Austin

**Roscoe Giles**
Boston University

**Mary Hall**
University of Utah

**Tanzima Islam**
Texas State University, San Marcos

**Kirk Jordan**
IBM

**Lin Lin**
University of California, Berkeley

**Alex Pothen**
Purdue University

**Padma Raghavan**
Vanderbilt University

**Robert Schreiber**
Cerebras

**Craig Thalhauser**
Genmab

**Alyson Wilson**
North Carolina State University

Staff: Support for this report was provided by Lewis-Burke Associates LLC, Miriam Quintal, and Griffin Reinecke

Back to TOC

Back to TOC

Back to TOC