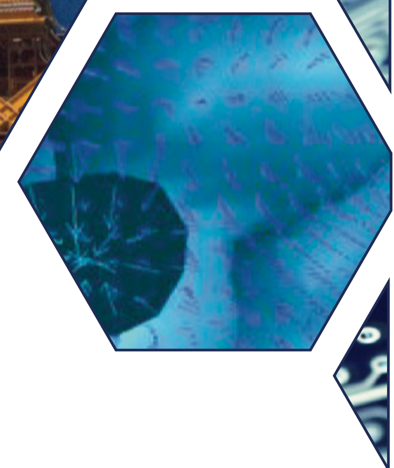
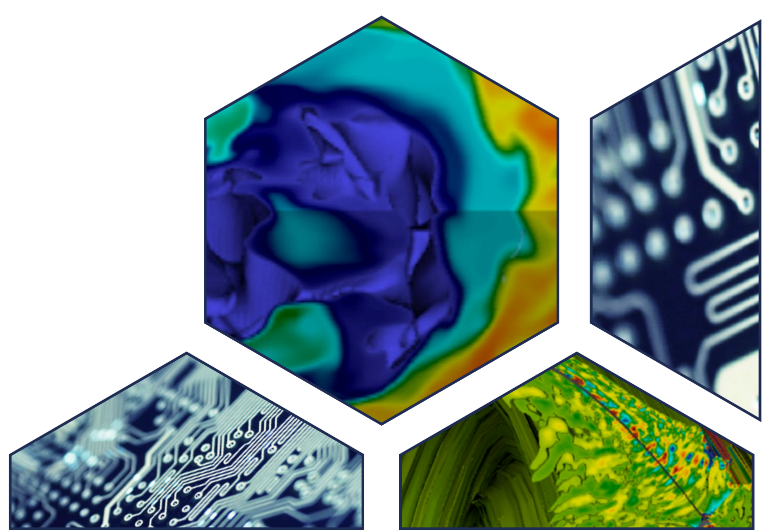


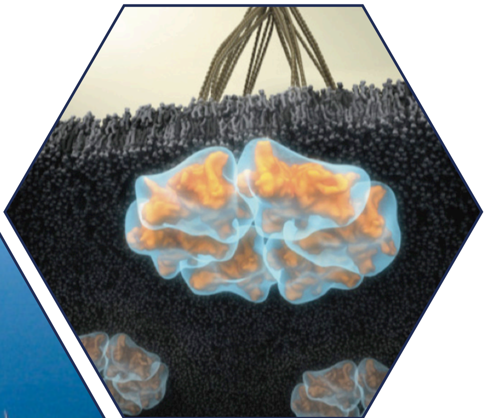
BASIC RESEARCH NEEDS FOR

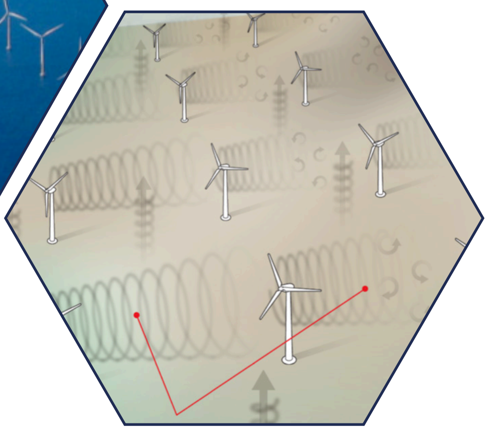
Productive Computational Science in the Era of Extreme Heterogeneity

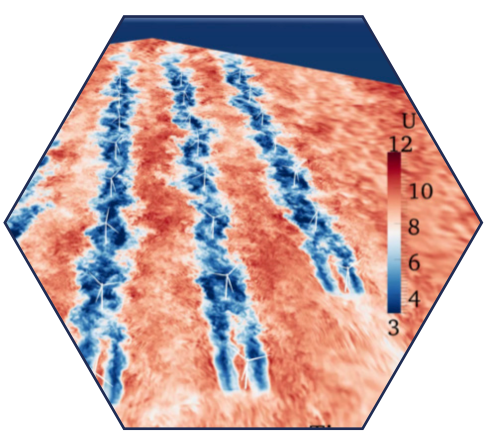
Reducing the time to scientific discovery in a transformational era of diverse scientific applications and revolutionary computer architectures



Revolutionizing how we utilize leadership class computing facilities for scientific innovation...







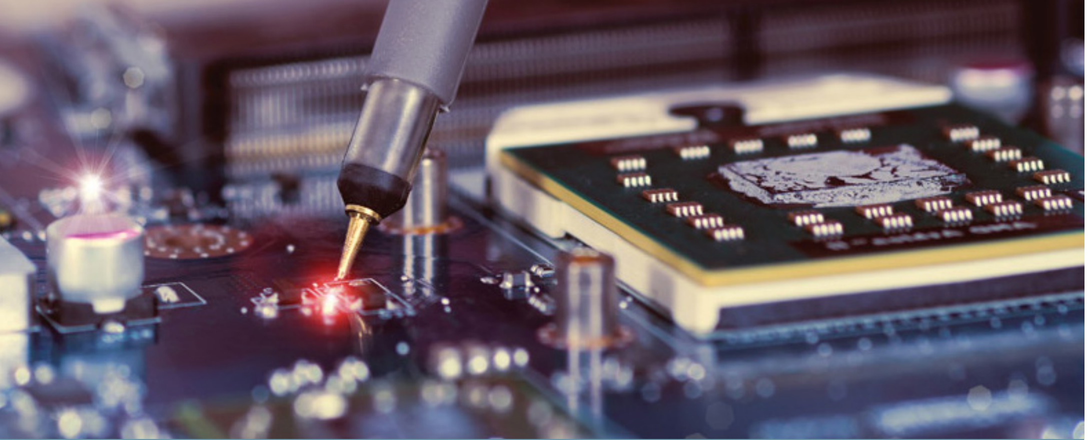
Providing a software environment that can overcome complexities in the design of future supercomputers will play a key role in improving the Nation's rate of scientific discovery, innovation and competitiveness.

During the last three decades, advances in computer technology have allowed the performance and functionality of processors to double every two years. This trend, known as Moore's Law, enabled both computational and experimental science to leverage the seemingly unending growth of the broad computing industry with very little change to the supporting software environment. As computer chip manufacturing techniques reach the limits of the atomic scale, this era of predictable performance improvements is ending and directly impacting the stability of today's software environment.

To address this challenge, computer vendors are already pursuing systems built from combinations of different kinds of processors in order to improve capabilities, boost performance, and meet energy efficiency goals. Many of DOE's supercomputers already rely on computational accelerators to meet the growing demands of increasingly complex computational workloads and the number of such accelerators will increase sharply in the future. This change will have a significant impact on both the design of high-performance computers and the software infrastructure required to effectively utilize the Nation's Leadership Computing Facilities.

According to the U.S. Department of Energy (DOE) Office of Science Advanced Scientific Computing Research program studies, several types of special-purpose accelerated processing units that are currently under development will play a major role in allowing future supercomputers to meet ever- increasing performance goals within acceptable power budgets. These processors will be augmented with new and diverse types of memory and data storage capabilities. The majority of these new developments are driven by rapid growth in the data-centric, machine learning and artificial intelligence marketplaces that far exceed the vendor revenues represented by high-performance computing for computational and experimental science.
In the 2025-2030 timeframe, these external economic drivers and resulting hardware design diversity will result in leadership-class supercomputers built from custom aggregations of diverse processors, accelerators, and memory types, resulting in a dramatic increase in the complexity of developing scientific software. This fundamental change in how supercomputers are designed has been termed as *extreme heterogeneity*.

A report from a 2018 Basic Research Needs workshop identified five priority research directions that are imperative for addressing the challenges posed by extreme heterogeneity. They are summarized below. The full report is available via https://doi.org/10.2172/1473756.



Priority Research Directions

Maintaining and improving programmer productivity

Key question: *In an era of an increasingly diverse and complex computing environments, what advances in programming models, environments, and tools are required to improve the productivity of a broad range of scientific software developers?*

The very high levels of effort and expertise required to develop verifiable, high-performance scientific software limit the ability to achieve key scientific milestones and transformational discoveries. Diversity and heterogeneity in the design of supercomputer architectures will increase the difficulty of these efforts and directly impact the overall time-to-solution. New programming environments, methodologies, and tools, combined with new approaches for effectively utilizing rapidly emerging capabilities such as machine learning and artificial intelligence, will be essential for reducing complexity and boosting productivity. This will be critical to maintaining and establishing competitive advantages across any number of mission critical areas of study.

Managing resources intelligently

Key question: *Can artificial intelligence and machine learning be effectively incorporated into system software to coordinate and control a large and diverse set of computing resources?*

Manually attempting to coordinate, reason about and schedule the placement of data; which types of processors to select for certain calculations; and when computations will occur will become intractable as the complexity, diversity and scale of high-performance computing systems grow alongside increasingly challenging scientific missions. Optimized resource management decisions must be made at a pace, scale and level of complexity that exceeds human ability. Furthermore, today's system software is not designed to manage rapid changes in resource scheduling and workloads that cutting-edge science and extremely heterogeneous systems will demand. Infusing artificial intelligence, especially machine learning, into system software capabilities provides an opportunity for improving and automating system use, increasing overall productivity, and accelerating scientific discovery and innovation.

Modeling & predicting performance

Key question: *Can advanced modeling and simulation predict the performance characteristics of applications running on emerging hardware technologies and provide insight into the design of future systems?*

With a growing number of options for processors, accelerators, networks and memories, optimally configuring a high-performance supercomputer for a wide range of scientific domains becomes overwhelmingly complex. New, intelligent modeling and simulation capabilities that facilitate evaluation of application behavior on novel hardware components would provide important guidance to software developers and intelligent system software, as well as enabling potential system designs to be evaluated for their suitability for science and mission needs. This would also allow DOE to consider customized systems that can be tailored to mission critical needs, thus improving productivity and maximizing the return on investment.

Enabling reproducible science despite diverse processors and non-determinism

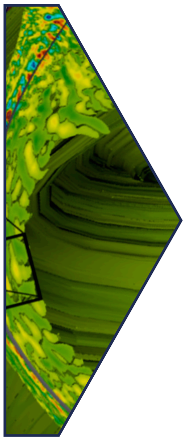
Key question: *What novel methods and techniques are needed to support productive, reliable and verifiable scientific findings in the face of architectural diversity and variability in future systems?*

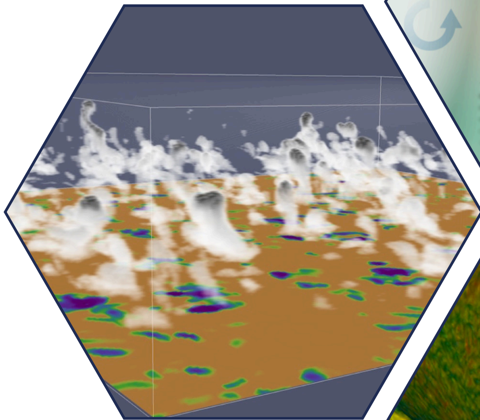
The ability to validate scientific outcomes is essential, but it becomes increasingly difficult as the number and heterogeneity of hardware components grows and computations become increasingly non- determinate and asynchronous. Effectively leveraging the increasing diversity of accelerated processors will require different approaches to computation that have the potential to impact the precision and thus exact reproducibility of numerical calculations. New methods, algorithms, and supporting software infrastructure are needed to enable developers to better reason about, evaluate and limit the impact of these uncertainties inherent in the variability among diverse hardware components and workloads.

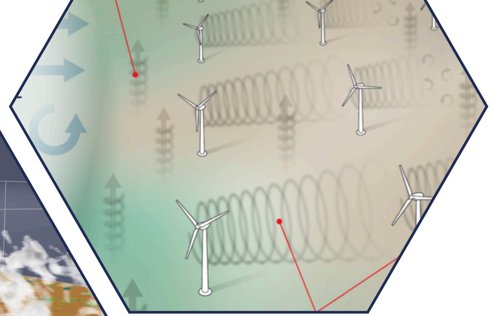
Facilitating data management, analytics, and workflows

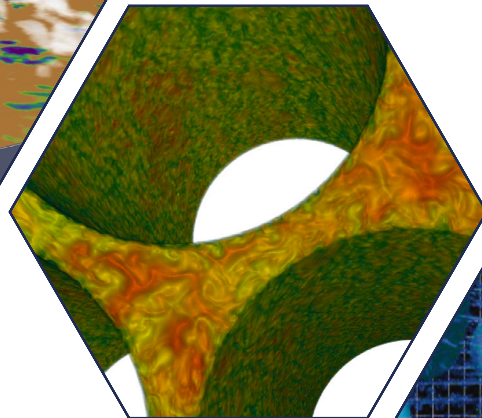
Key question: *What software infrastructure and tools will be necessary to achieve usable and productive scientific workflow across multiple, different and increasingly complex computing environments?*

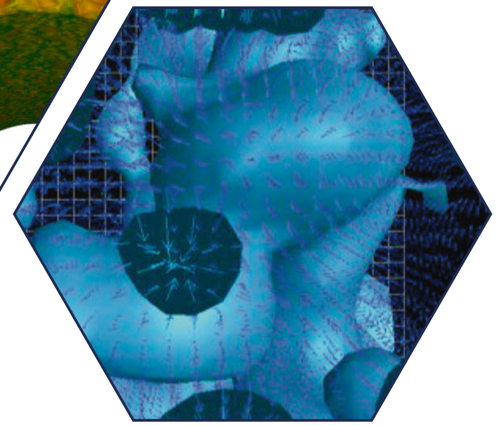
A scientific campaign relies on multiple simulations and/or experiments, which must be coordinated across an increasingly complex array of scientific instruments, distributed data resources and large geographically distributed teams. The system software environment must improve to facilitate this process across a range and combination of different computing and experimental facilities, from finding and scheduling the available resources to composing and executing the complete workflow for a broad set of scientific domains. New and innovative tools will be needed for tracking the scientific process from formulation of a hypothesis to final discovery, spanning both the dynamic and static selection of appropriate computing and data storage resources, analyzing the resulting data, and cataloging experimental results and findings.











U.S. DEPARTMENT OF
ENERGY

Office of
Science