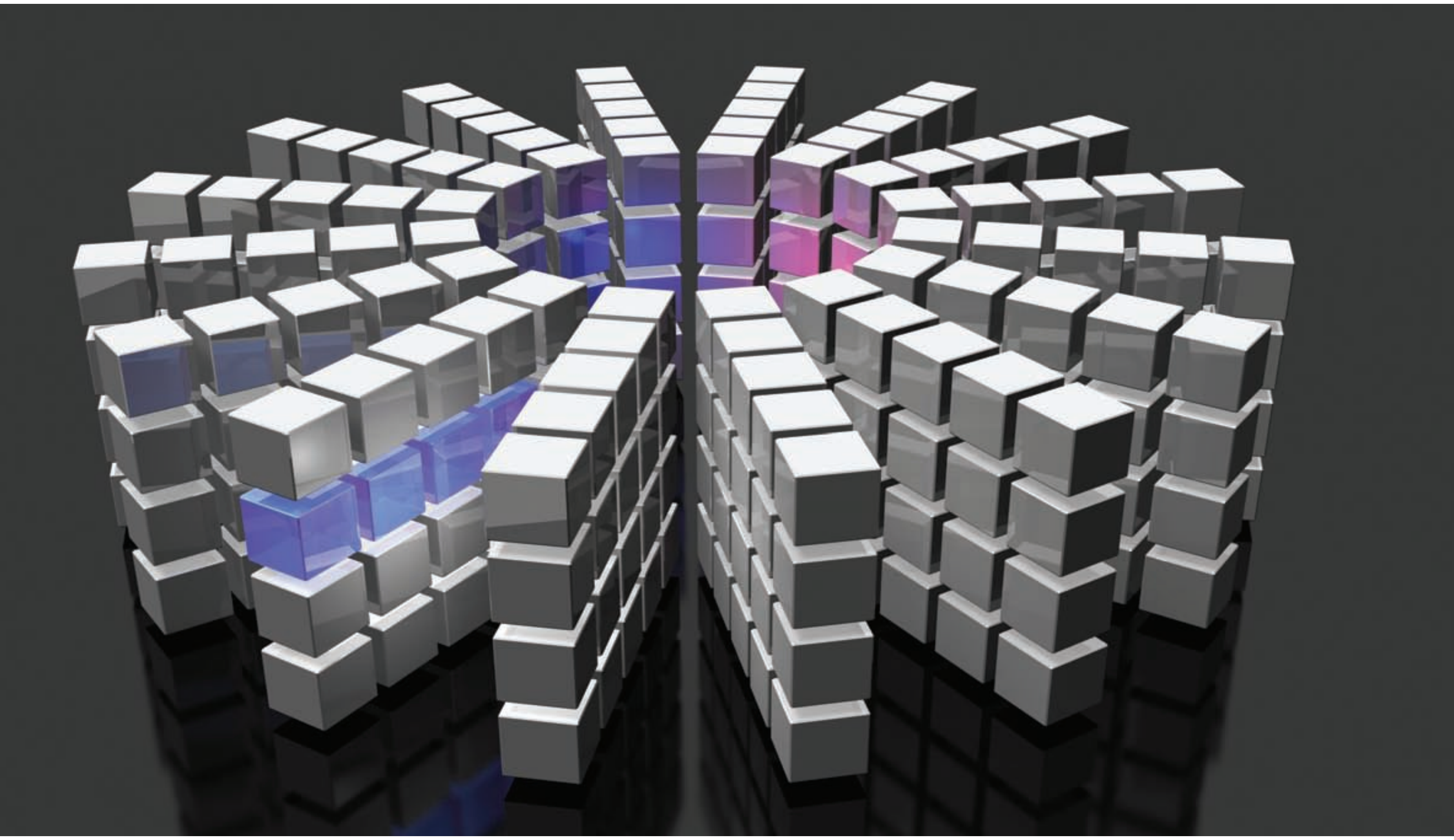
Editors: James J. Hack, jhack@ornl.gov | Michael E. Papka, papka@anl.gov



The Exascale Computing Project

Paul Messina | Argonne National Laboratory

Exascale computing—computing capability that can achieve at least a billion billion operations per second, 50 to 100 times more powerful than the nation's fastest supercomputers in use today—is the next waypoint in high-performance computing (HPC). Achieving such speeds isn't an easy task. To be useful to a wide spectrum of applications, in addition to peak speed, supercomputers need to have large memories and the ability to store and read vast quantities of data at high speed. In addition, the supercomputer must have a software environment that facilitates its efficient and productive use.

The value of exascale computing is primarily in the applications that it will enable, which include national security, fundamental science, industrial design, and health.

The Exascale Computing Project

The US Department of Energy (DOE) launched the Exascale Computing Project (ECP; https://exascaleproject.org/exascale-computing-project) in 2016 as part of a coordinated effort to achieve the next generation of HPC and to accelerate scientific discovery and economic competitiveness. In addition to enabling traditional supercomputing applications at the exascale, the ECP seeks to contribute to the emerging convergence of supercomputing, big data analytics, and machine learning across a wide variety of science and engineering domains and disciplines to address problems of national significance and to help meet our national security needs.

This seven-year project aims to advance all aspects of the computing ecosystem: applications that require exascale computing to tackle previously intractable problems, support software that provides the bridge between applications and hardware, and computer architectures and technologies that will result in usable, reliable, and affordable exascale computers. The long-term benefit of the ECP's approach will be a thriving ecosystem of capable, US-based, exascale computing products and scientists and engineers with expertise in using them.

The ECP is a collaboration among the nation's six premier computing labs (Argonne National Laboratory, Los Alamos National Laboratory, Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Oak Ridge National Laboratory, Sandia National Laboratories) with funding from the DOE Office of Science (SC) and National Nuclear Security Administration (NNSA). The project is led by a collaborative national laboratory staff: the current director of the ECP is from Argonne, the project office is at Oak Ridge, and the technical leadership team is populated by staff from the six core labs.

Background

Starting in 2007, the DOE sponsored numerous workshops on exascale, some in collaboration with other government agencies. The first dozen or so focused on identifying the potential benefits of exascale computing (https://science.energy.gov/ascr/community-resources/workshops-and-conferences/grand-challenges). These were followed by workshops and studies on the technical challenges—both hardware and software—in achieving exascale capability (http://exascaleresearch.labworks.org/ascrMar2011/ascr2011). A series of international workshops developed a software road map for exascale (www.exascale.org/mediawiki/images/2/20/IESP-roadmap.pdf) and studies carried out by the Advanced Scientific Computing Advisory Committee (https://science.energy.gov/ascr/ascac) identified opportunities and challenges of exascale computing (https://science.energy.gov/~/media/ascr/ascac/pdf/reports/Exascale_subcommittee_report.pdf). Advancing from petascale to exascale computing presents challenges that are much greater than those that were surmounted in the transitions from gigascale to petascale. For 40 years, each generation of computer circuits got smaller and faster than the previous one. Around 2004, chips stopped getting faster when thermal limitations put an end to frequency scaling (also called Dennard scaling). The only way to continue to increase performance was to increase the number of processors, thus requiring programming models and algorithms that can exploit massive parallelism: hundreds of millions of operations. Memory and storage—getting data in and out of such systems—will require new approaches. The large number of processors leads to reliability issues. Power requirements for exascale systems using current technologies would be unacceptably high, hundreds of megawatts, hence the decision by the DOE to create a project to achieve the exascale goal; business as usual would not yield affordable and usable exascale systems anytime soon.

The technological advances necessary to achieve affordable and productive exascale computing will also yield important benefits to the HPC ecosystem. Other nations grasp the value of HPC for national and economic security and have begun investing heavily in HPC systems and in exascale computing. China has had the number one system on the semiannual TOP500 list since June 2013; it occupies the top two spots on the current list. China is targeting deployment of at least one exascale system by 2020; it has three potential exascale architectures under development and will deploy their prototypes in 2017. Japan and the European Commission have also launched major programs to develop exascale computing capabilities.

Here in the US, the ECP has organized its technical activities into four focus areas: application development, software technology, hardware technology, and exascale systems.

Application Development

The exascale application development effort targets specific R&D activities and outcomes that address critical DOE missions as well as applications with broad societal and economic impact. For selected critical problems in DOE science, energy, and national security mission space and other federal agency mission space, this focus area funds and manages application software teams to create or enhance the predictive capability of relevant applications through targeted development of requirements-based models, algorithms, and methods, as well as integration of appropriate software and hardware co-design methodologies.

Currently, 25 applications projects have been launched, covering topics ranging from cosmology to combustion to wind energy to precision medicine for cancer treatment, all of which require exascale computing capability. Each application has identified a "challenge problem" that the team will tackle once the first exascale systems are deployed. These application teams are complemented by five co-design centers that are developing versions of frequently used methods such as particle-in-cell, adaptive mesh refinement, and graph algorithms. Related activities include training on new computing methodologies and software engineering.

Software Technology

To achieve the full potential of exascale computing, an expanded and vertically integrated software stack must be developed that includes advanced math libraries, a programming environment, tools, and extreme-scale system management. The scope of the software technology effort spans low-level operational software to high-level applications software development environments, including the software infrastructure to support large data management and data science for the DOE SC and NNSA computational science and national security activities at the exascale level. Software technology will also help develop required software for novel architecture HPC ecosystems as well as address portability issues across all HPC ecosystems, focusing on the 2021 and 2023 DOE exascale system acquisitions. An ambitious goal of this activity is to partner with vendors and software developers to create an HPC software stack that supports applications on systems from midscale HPC to exascale.

Hardware Technology

The hardware technology focus area supports vendor-based node and system design R&D activities, called PathForward, required to develop exascale systems with at least two diverse architectures in support of DOE exascale system acquisitions in 2021 and 2023, and addressing the challenges of massive parallelism, memory and storage, reliability, and energy consumption. This focus area also applies laboratory architectural analysis capabilities and abstract machine models to PathForward designs to support ECP co-design interactions.

Exascale Systems

The exascale systems focus area supports advanced system engineering development by vendors needed to produce capable exascale systems. The focus area also includes acquisition and support of testbeds for the application, software, and hardware activities, and will provide exascale requirements for planned procurements by the DOE supercomputing facilities. Computer procurements will be executed by the laboratories hosting the systems, including any site preparations needed.

A Holistic Solution

A fundamental approach of the ECP is to emphasize co-design and integration of activities to ensure that the result is a robust exascale ecosystem. Accordingly, the ECP combines application, software, and hardware development into a holistic solution that uses a co-design process across all areas, an approach that's necessary for these activities to result in a coherent exascale ecosystem. Co-design refers to a system-level design process in which scientific and mission problem requirements influence architecture design and hardware technology, and architectural characteristics inform the formulation and design of applications, algorithms, and system software. The co-design methodology requires the combined expertise of vendors, hardware architects, software stack developers, domain scientists, computer scientists, applied mathematicians, and systems staff working together to make informed decisions about the design of applications, algorithms, and the underlying system software and hardware.

Because both hardware and ECP application software require many years between concept, design, and production, co-design must be executed early in the ECP in partnership with users, vendors, and developers of applications, system software, tools, and other technology researchers. The software and hardware requirements identified by the application development teams have a particularly important role in the co-design process of the ECP.

Why We Need Exascale Computing

Every time computing power increases by large factors, new benefits open before us. The benefits of exascale computing—which range from creating novel, more efficient combustion engines and new energy solutions to advances in healthcare, biology, and storm prediction—could potentially impact every person reading this article.

National Security Needs

HPC-based modeling and simulation tools already play a role in the NNSA's annual assessment process and help provide stewardship certification and assessment of stockpile safety, security, and reliability. The importance and role of HPC in stockpile stewardship will intensify in the coming decade as enhanced computing capabilities provide more confident results.

Advances in Materials Science

To create certain new technologies and inventions, we need to discover new materials with specific properties. One crucial way to make these discoveries is by complex calculations that simulate how materials behave in nature by using massive databases of known compounds to identify combinations that have the desired properties. For example, much more efficient, less expensive, and longer-lived batteries will be developed through discovery of new materials. Batteries are very important, not only for portable consumer electronics and electric automobiles, but to store energy from variable energy sources, such as wind and solar, to use when they're needed. Deep learning techniques and classical simulation are in use today in this field, and exascale computing will enable faster and more complex designs.

New Energy Solutions

Solar energy will also be made more cost-effective through the discovery of materials that convert the sun's rays to electricity more efficiently. Wind turbines can be made more efficient and quieter through computer-based design of their blades, and simulations can optimize the locations of individual wind turbines in wind farms to yield substantially more energy from the same number of turbines.

Another challenge in the use of alternative energy sources is managing the electric power grid, such as when the wind dies down in one area of a wind farm, for example. The electric power grid is complex, and adjustments such as firing up a fossil fuel generator have to be made quickly to avert problems; it isn't possible to predict accurately such events far ahead of time. Computationally intensive optimization methods can provide the needed guidance, but to get the results in time to be useful requires fast computers.

More than 85 percent of the world's energy is generated by burning fossil fuels, but we can minimize pollution and optimize efficiency by understanding and controlling the chemical process of combustion. Through exascale computing, we expect it will be possible to increase the efficiency of combustion systems in engines and gas turbines for transportation and power generation by potentially 25 to 50 percent, and to thereby lower emissions.

Advances in Healthcare

Exascale computing will accelerate cancer research by helping scientists understand the molecular basis of key protein interactions and by automating the analysis of information from millions of cancer patient records to determine optimal treatment strategies. It will also enable doctors to predict the right treatments for the patient by modeling drug responses. These tasks are daunting: for example, drug combination response prediction might require the search of 1 trillion drug combinations.

Exascale computing applications in biology might enable the prediction of feasible parameter values for dynamic models of metabolism that would enable scientists to design organisms that would perform a variety of tasks. These models might also contribute to the development of treatments for emerging types of infections.

Predicting Severe Weather

Weather prediction models will be able to predict more accurately and quickly the timing and path of severe weather events such as hurricanes by using much higher spatial resolution, incorporating more physics, and assimilating more observational data. Timely delivery of the predictions is essential to provide sufficient time for evacuations and protective measures.

Improving Quality of Life

The use of exascale computing in urban science promises to mitigate health hazards, reduce crime, and improve the quality of life in cities by optimizing infrastructure (such as transportation, energy, housing, and healthcare) access and usage choices. Such optimization requires gathering and analysis of numerous types of data from databases, sensors, and simulation results and conducting thousands of potential scenarios. The resulting analyses will be useful for planning new cities or neighborhoods as well as restructuring the infrastructure in existing large urban areas.

Fundamental Science

Cosmology, high-energy physics, astrophysics, and theoretical chemistry are among the sciences that seek to understand the evolution of the universe, the laws that govern the creation and behavior of matter. All of them rely heavily or even completely on computer modeling and simulation to conduct their research. Although such studies aren't explicitly aimed at results that immediately benefit society, historically they have a track record of making discoveries that have major impacts in everyday life. An obvious example is the World Wide Web, which was invented to support certain aspects of particle physics research.

Improved Product Design and Manufacturing

Computer modeling has long been used in the design of products ranging from airplanes to diapers. With exascale computing capabilities, it will be possible to optimize design of such products even further as well as to exploit new technologies such as additive manufacturing with metals.

The ECP has a broad scope with a correspondingly broad spectrum of goals:

- develop applications that will tackle a broad spectrum of mission-critical problems of unprecedented complexity with unprecedented performance;
- contribute to the economic competitiveness of the nation;
- support national security;
- develop a software stack, in collaboration with vendors, that is exascale-capable and is usable on smaller systems by industry and academia;
- partner with vendors to develop computer architectures that support exascale applications; and
- train a large cadre of computational scientists, engineers, and computer scientists who will be an asset to the nation long after the end of the ECP.

There are additional important applications that exascale computing will advance but will require even more computing power to accomplish, such as reverse engineering the human brain to understand complex neural systems; new methods to integrate large-scale data; and information and simulation dynamics.

Exascale computing will potentially benefit society in myriad ways, only some of which are mentioned in this article. Novel discoveries, new materials, and solutions to problems that impact everyday life will be possible through new capabilities of exascale computing systems. In addition to benefits to society, key technological advances also hold the promise of maintaining economic security. ■

Acknowledgments

This research was supported by the Exascale Computing Project (ECP), project number: 17-SC-20-SC, a collaborative effort of two DOE organizations—the Office of Science and the National Nuclear Security Administration—which are responsible for planning and preparing a capable exascale ecosystem, including software, applications, hardware, advanced system engineering, and early testbed platforms, to support the US exascale computing imperative.

Paul Messina is an Argonne Distinguished Fellow at Argonne National Laboratory and director of the Exascale Computing Project for the US Department of Energy. Contact him at messina@anl.gov.