SAND2020-11372R

Two new SciDAC institutes promote mathematical tools and software technology for high-performance computing

Gail Pieper, Karen Devine, Esmond Ng, Leonid Oliker, and Robert Ross

Bigger is often said to be better, and the newest extreme-scale computers certainly are bigger, with millions of processing units. Moreover, the breadth of science performed on the U.S. Department of Energy (DOE) computing facilities is expanding, with new technology such as artificial intelligence emerging. These advances are exciting, creating new opportunities for scientific discovery; however, they also raise new questions for scientists who want to exploit these advances for tackling more complex problems. Will my simulation code be able to utilize the accelerators in extreme-scale computing systems? Can I take advantage of the deepening memory hierarchy in heterogeneous processors? Is there a way around bottlenecks caused by the widening ratio of peak floating-point operations per second to I/O bandwidth? How can I manage my huge amounts of data effectively? Can I analyze data in situ, or must I transfer it to offline storage for later analysis?

To address such questions, DOE announced that it is providing $57.5 million over the next five years for two multidisciplinary teams – FASTMath and RAPIDS2 – to develop new tools and techniques to harness supercomputers for scientific discovery. The teams, called SciDAC Institutes, are part of the Scientific Discovery through Advanced Computing program.

A Brief Background of SciDAC

The SciDAC program was initiated in 2001 to accelerate scientific discovery using high-performance computing (HPC). SciDAC is a joint effort involving the six major program offices within DOE's Office of Science (SC) – Advanced Scientific Computing Research, Basic Energy Sciences, Biological and Environmental Research, Fusion Energy Sciences, High-Energy Physics and Nuclear Physics – as well as DOE's Office of Nuclear Energy. It addresses problems in disciplines including high energy and nuclear physics, condensed matter physics, materials science, chemistry, fusion energy sciences and Earth systems research.

The aim of the SciDAC program is to ensure that scientists from national laboratories, universities and other research organizations take full advantage of DOE's HPC resources. A key goal is to bridge the gap between mathematics and computer science research and domain science research – enabling the potential to significantly advance scientific discovery. Since its establishment, the SciDAC program has enjoyed tremendous success and produced significant achievements, ranging from new insights into the actions of supernovae to improving combustion to reduce pollution.

Now in its fourth five-year cycle, the SciDAC program is recognized worldwide as a leading force in accelerating the use of HPC to advance the state of knowledge in science.

The Two New SciDAC Institutes

With SciDAC-4 approaching an end, the DOE announced in March 2020 its plan to establish multidisciplinary teams to develop new tools and techniques in mathematics and computer science to harness state-of-the-art supercomputers for scientific discovery as part of SciDAC-5. These teams would take advantage of DOE supercomputing facilities at Argonne, Oak Ridge and Lawrence Berkeley National Laboratories. Following an open competition, DOE announced in August that it had selected two Institutes to be funded under the SciDAC-5 program.

- "Frameworks, Algorithms, and Scalable Technologies for Mathematics" (FASTMath) will focus on the development of new scalable mathematical algorithms and software tools that can exploit the power of extreme-scale computers.

- "RAPIDS2: A SciDAC Institute for Computer Science, Data, and Artificial Intelligence" will focus on helping application developers address challenges arising from the increasing deluge of data, new technologies and extreme-scale computing.

Both Institutes involve large research collaborations from academia and national laboratories. Members of the FASTMath and RAPIDS2 Institutes have a significant track record of successful collaboration within both the Institutes and the larger scientific community over the past 15 years; indeed, strong collaboration between FASTMath and RAPIDS was a highlight of SciDAC-4.

Impact on scientific applications remains the key focus of both Institutes. Toward that goal, FASTMath and RAPIDS2 researchers will engage with application developers and domain experts in science-focused SciDAC Partnerships and DOE SC projects to address some of the most complex computational problems of interest to the DOE. They will build on their suites of high-quality software with particular focus on lowering barriers to achieving high performance and high productivity on DOE computers.

Both Institutes are active in outreach to the broader scientific community. Through activities such as summer schools, tutorials and workshops, the Institute teams train the scientific computing community on leveraging their software and help educate the next generation of computational mathematicians and scientists.

The FASTMath Institute

The FASTMath Institute, led by Esmond Ng (Lawrence Berkeley National Laboratory) and Karen Devine (Sandia National Laboratories), is committed to providing robust mathematical techniques and expertise to enhance the performance and effectiveness of scientific simulations. FASTMath pursues three key goals:

- Deliver highly performant software with strong software engineering to run efficiently on current and next-generation advanced computer architectures at the DOE's major computing facilities;
- Work closely with domain scientists to share the FASTMath team's mathematical and machine learning expertise and deploy its software in large-scale modeling and simulation codes; and
- Build and support the broader computational mathematics and computational science communities across the DOE complex.

Several mathematical and computational challenges require the expertise of a team like FASTMath. The integration of graphics processing units (GPUs) in emerging exascale-scale computers, for example, means that computational scientists have to redesign their software to take full advantage of these accelerators. To tackle multiscale, multiphysics problems, researchers need to quantify the uncertainty in their computations and be assured of higher fidelity. Domain scientists also need to leverage new technologies such as machine learning in modeling and workflow simulations to quickly and accurately analyze the torrents of generated data. To address such challenges, FASTMath's efforts will span eight technical areas: structured mesh discretization, unstructured mesh discretization, time integration, linear and nonlinear equation solvers, eigensolvers, numerical optimization, uncertainty quantification and data analytics. Machine learning is a cross-cutting theme among these eight technical areas, with planned activities including numerical methods for machine learning and use of machine learning to optimize application usage of FASTMath software.

The FASTMath team comprises more than 50 mathematicians from five national laboratories (Argonne, Lawrence Berkeley, Lawrence Livermore, Oak Ridge and Sandia) and five universities (MIT, Rensselaer Polytechnic Institute, Southern Methodist University, University of Colorado Boulder and University of Southern California). Many FASTMath researchers also participate in SciDAC-4 Partnerships and DOE SC base math projects, enabling them to incorporate research developments into new tools for deployment in scientific applications.

The RAPIDS2 Institute

By providing high-performance computer science and data management tools, the RAPIDS2 Institute, led by Rob Ross (Argonne National Laboratory) and Lenny Oliker (Lawrence Berkeley National Laboratory), seeks to help DOE SC's application teams using leadership computing

resources to achieve scientific breakthroughs. To accomplish this objective, the Institute has identified the following goals:

- Solve computer science, data and artificial intelligence (AI) technical challenges for SciDAC and DOE science teams;
- Engage and work directly with SC scientists and facilities to identify needs and deploy new technologies; and
- Coordinate with other DOE computer science / applied mathematics activities and the DOE Exascale Computing Project to maximize impact on DOE science.

RAPIDS2 will build on the successes of the SciDAC-4 RAPIDS project and expand into several new areas that will have broad impact. Specifically, RAPIDS2 is addressing four technology thrusts: (1) data understanding, including ensemble analysis and feature detection; (2) HPC platform readiness, including heterogeneous programming and autotuning; (3) scientific data management, including workflow automation and storage systems and I/O; and (4) artificial intelligence, including representation learning and surrogate modeling. These thrusts provide a toolbox of advanced computation, information and data science technologies that address common challenges faced by scientific applications. Artificial intelligence is particularly exciting cross-cutting technology due to its potential to transform numerous scientific domains that utilize high-performance computing, such as materials science, high energy physics and chemistry.

RAPIDS2 has brought researchers from six universities (Northwestern University, Ohio State University, Rutgers University, University of Delaware, University of Florida and University of Oregon), five national laboratories (Argonne, Lawrence Berkeley, Lawrence Livermore, Los Alamos and Oak Ridge), and one software research and development company (Kitware), together providing a broad range of expertise and strong history of success in engagement with DOE scientists.

The Power of Two

By combining the knowledge and expertise of mathematicians and computer scientists working jointly with domain scientists, the FASTMath and RAPIDS2 Institutes complement each other and cover a wide spectrum of scientific computing needs. The teams' researchers are confident they have the expertise, algorithms, software and other computational tools required to provide end-to-end solutions to help application developers satisfy the SciDAC mission of advancing scientific discoveries through modeling and simulation on DOE's most advanced computers. Researchers interested in collaborating with the FASTMath and RAPIDS2 teams are invited to contact the Institute leaders or any Institute members.

For the press announcement from DOE, see https://www.energy.gov/articles/department-energy-provide-575-million-science-computing-teams. A list of lead and partner institutions for the two teams can be found on the homepage of the DOE Office of Science, Office of Advanced Scientific Computing Research, under the heading, "What's New."

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.