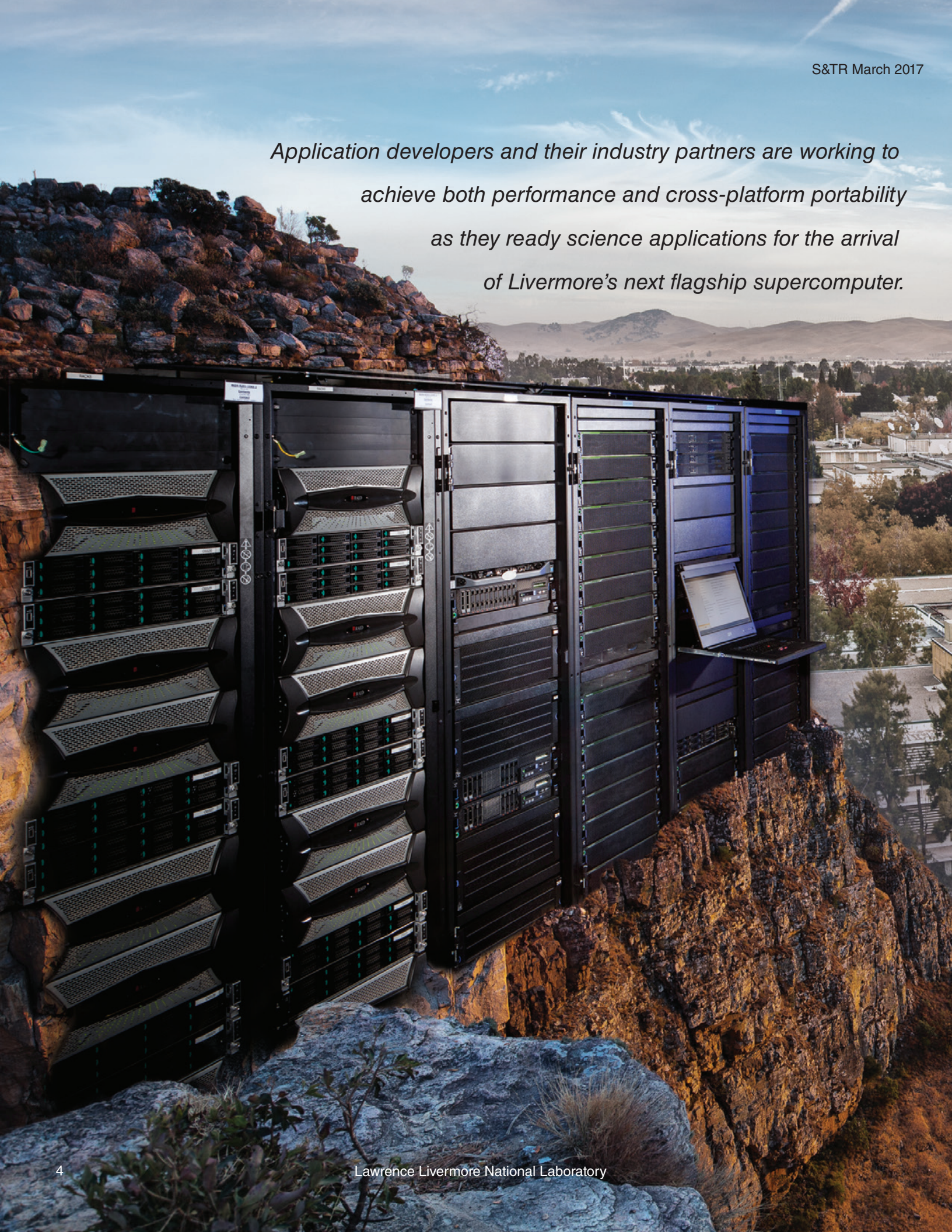
Application developers and their industry partners are working to achieve both performance and cross-platform portability as they ready science applications for the arrival of Livermore's next flagship supercomputer.



A Center of Excellence Prepares for SIERRA

In November 2014, then Secretary of Energy Ernest Moniz announced a partnership involving IBM, NVIDIA, and Mellanox to design and deliver high-performance computing (HPC) systems for Lawrence Livermore and Oak Ridge national laboratories. (See *S&TR*, March 2015, pp. 11–15.) The Livermore system, Sierra, will be the latest in a series of leading-edge Advanced Simulation and Computing (ASC) Program supercomputers, whose predecessors include Sequoia and BlueGene/L, Purple, White, and Blue Pacific. As such, Sierra will be expected to help solve the most demanding computational challenges faced by the National Nuclear Security Administration's (NNSA's) ASC Program in furthering its stockpile stewardship mission.

At peak speeds of up to 150 petaflops (a petaflop is 10^{15} floating-point operations per second), Sierra is projected to provide at least four to six times the performance of Sequoia, Livermore's current flagship supercomputer. Consuming "only" 11 megawatts, Sierra will also be about five times more power efficient. The system will achieve this combination of power and efficiency through a heterogeneous architecture that pairs two types of processors—IBM Power9 central processing units (CPUs) and NVIDIA Volta graphics processing units (GPUs)—so that programs can shift from one processor type to the other based on computational requirements. (CPUs are the traditional number-crunchers of computing. GPUs, originally developed for graphics-intensive applications such as computer games, are now being incorporated into supercomputers to improve speed and reduce energy usage.) Powerful hybrid computing units known as nodes will each contain multiple CPUs and GPUs connected by an NVLink network that transfers data between components at high speeds. In addition to CPU and GPU memory, the complex node architecture incorporates a generous amount of nonvolatile random-access memory, providing memory capacity for many operations historically relegated to far slower disk-based storage.

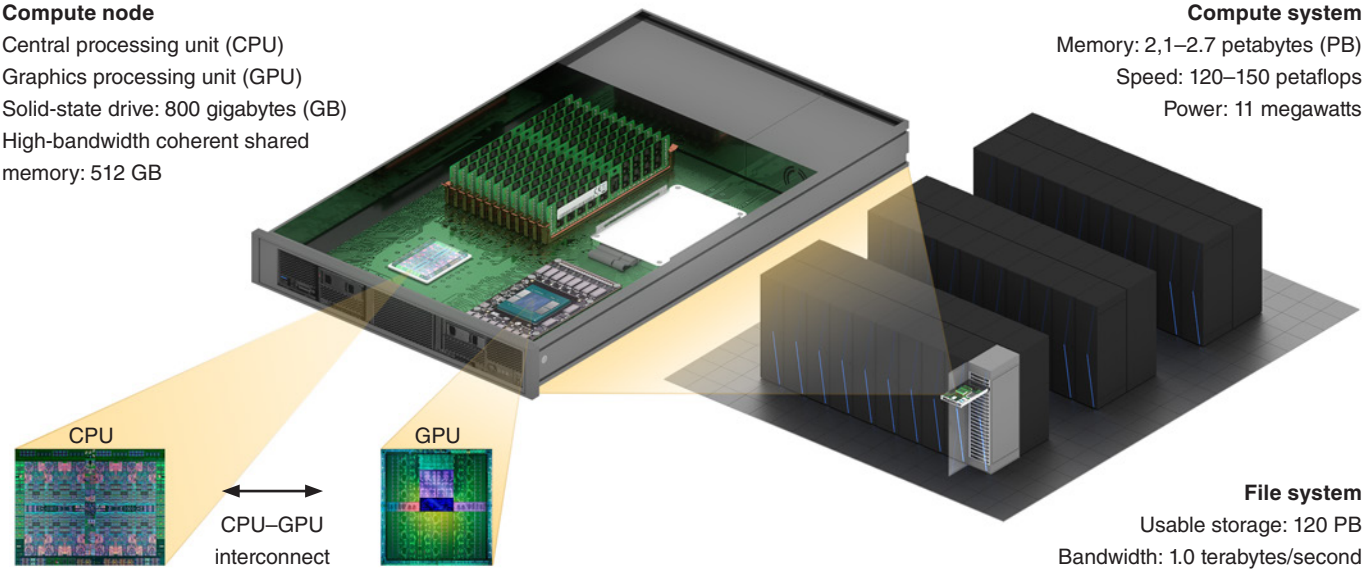
The Department of Energy has contracted with an IBM-led partnership to develop and deliver advanced supercomputing systems to Lawrence Livermore and Oak Ridge national laboratories beginning this year. These powerful systems are designed to maximize speed and minimize energy consumption to provide cost-effective modeling, simulation, and big data analytics. The primary mission for the Livermore machine, Sierra, will be to run computationally demanding calculations to assess the state of the nation's nuclear stockpile.

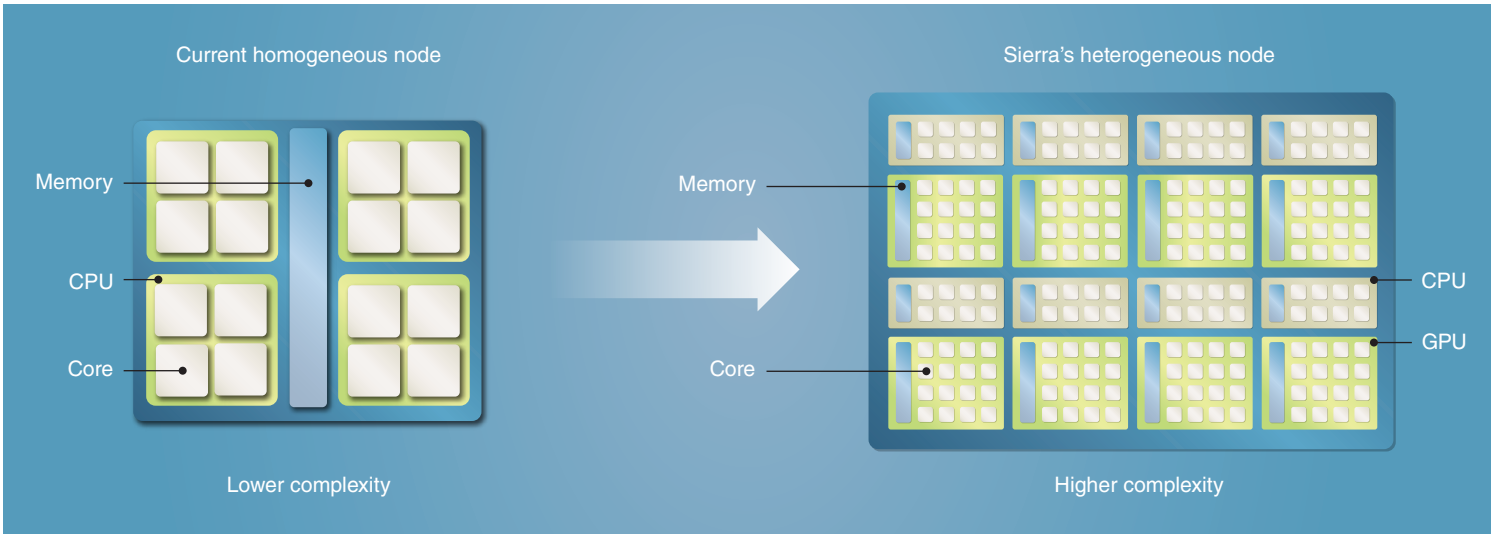
These hardware features—heterogeneous processing elements, fast networking, and use of different memory types—anticipate trends in HPC that are expected to continue in subsequent generations of systems. However, these innovations also represent a seismic architectural shift that poses a significant challenge for both scientific application developers and the researchers whose work depends on those applications running smoothly. "The introduction of GPUs as accelerators into the production ASC environment at Livermore, starting with the delivery of Sierra, will be disruptive to our applications," admits Rob Neely. "However, Livermore chose the GPU accelerator path only after concluding, first, that performance-portable solutions would be available in that timeframe and, second, that the use of GPU accelerators would likely be prominent in future systems."

To run efficiently on Sequoia, applications must be modified to achieve a level of task division and coordination well beyond what previous systems demanded. Building on the experience gained through Sequoia, and integrating Sierra's ability to effectively use GPUs, Livermore HPC experts are hoping that application developers—and their applications—will be better positioned to adapt to whatever hardware features future generations of supercomputers have to offer.

A Platform for Engagement

IBM will begin delivery of Sierra in late 2017, and the machine will assume its





Compared to today's relatively simpler nodes, cutting-edge Sierra will feature nodes combining several types of processing units, such as central processing units (CPUs) and graphics-processing units (GPUs). This advancement offers greater parallelism—completing tasks in parallel rather than serially—for faster results and energy savings. Preparations are underway to enable Livermore's highly sophisticated computing codes to run efficiently on Sierra and take full advantage of its leaps in performance.

full ASC role by late 2018. Recognizing the complexity of the task before them, Livermore physicists, computer scientists, and their industry partners began preparing applications for Sierra shortly after the contract was awarded to the IBM partnership. The vehicle for these preparations is a nonrecurring engineering (NRE) contract, a companion to the master contract for building Sierra that is structured to accelerate the new system's development and enhance its utility. "Livermore contracts for new machines always devote a separate sum to enhance development and make the system even better," explains Neely. "Because we buy machines so far in advance, we can work with the vendors to enhance some features and capabilities." For example, the Sierra NRE calls for investigations into GPU reliability and advanced networking capabilities. The contract also calls for the creation of a Center of Excellence (COE)—a first for a Livermore NRE—to foster more intensive and sustained engagement than usual among domain scientists, application developers, and vendor hardware and software experts.

The COE provides Livermore application teams with direct access to vendor expertise and troubleshooting as codes are optimized for the new architecture. (See the box on p. 9.) Such engagement will help ensure that scientists can start running their applications as soon as possible on Sierra—so that the transition sparks discovery rather than being a hindrance. In turn, the interactions give IBM and NVIDIA deeper insight into how real-world scientific applications run on their new hardware. Neely observes that the new COE also dovetails nicely with Livermore's philosophy of codesign, that is, working closely with vendors to create first-of-their-kind computing systems, a practice stretching back to the Laboratory's founding in 1952.

The COE features two major thrusts—one for ASC applications and another for non-ASC applications. The ASC component, which is funded by the multilaboratory ASC Program and led by Neely, mirrors what other national laboratories awaiting delivery of new HPC architectures over the coming years are doing. The non-ASC component, however, was pioneered by Lawrence Livermore, specifically the Multiprogrammatic and Institutional Computing (M&IC) Program, which helps make powerful HPC resources available to all areas of the Laboratory. One way the M&IC Program achieves this goal is by investing in smaller-scale versions of new ASC supercomputers, such as Vulcan, a quarter-sized version of Sequoia. Sierra will also have a scaled-down counterpart slated for delivery in 2018. To prepare codes to run on this new system in the same manner as the COE is doing with ASC codes, institutional funding was allocated in 2015 to establish a new component of the COE, named the Institutional COE, or iCOE.

To qualify for COE or iCOE support, applications were evaluated according to mission needs, structural diversity, and, for iCOE codes, topical breadth. Bert Still, iCOE's lead, says, "We chose to fund preparation of a strategically important subset of codes that touches on every discipline at the Laboratory." Applications selected by iCOE include those that help scientists understand earthquakes, refine laser experiments, or test machine learning methods. For instance, the machine learning

application chosen (see *S&TR*, June 2016, pp. 16–19) uses data analytics techniques considered particularly likely to benefit from some of Sierra's architectural features because of their memory-intensive nature. Applications under COE purview make up only a small fraction of the several hundred Livermore-developed codes presently in use. The others will be readied for Sierra later, in a phased fashion, so that lessons learned by the COE teams will likely save significant time and effort. Still says, "At the end of the effort, we will have some codes ready to run, but just as importantly, we will have captured and spread knowledge about how to prepare our codes." Although some applications do not necessarily need to be run at the highest available computational level to fulfill mission needs, their developers will still need to familiarize themselves with GPU technology, because current trends suggest this approach will soon trickle down to the workhorse Linux cluster machines relied on by the majority of users at the Laboratory.

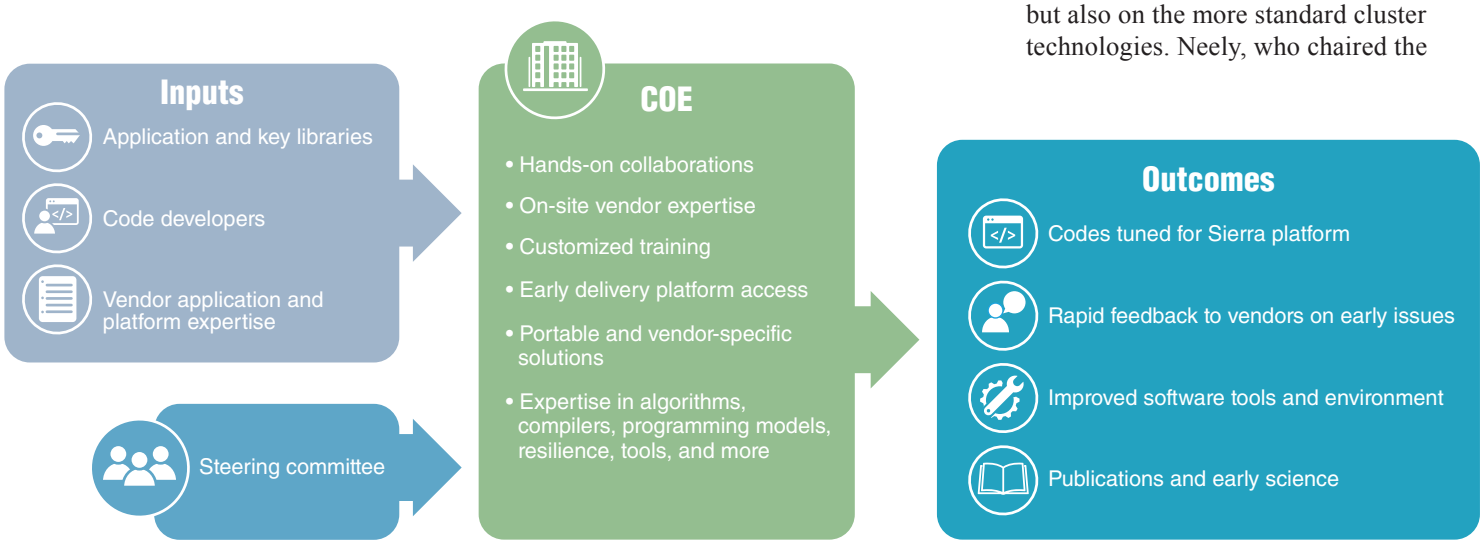
Performance and Portability

Nearly every major domain of physics in which the massive ASC codes are used has received some COE attention, and nine of ten iCOE applications are already in various stages of preparation. A range of onsite and offsite activities has been organized to bolster these efforts, such as training classes, talks, workshops focused on hardware and application topics, and hackathons. At multiday hackathons held at the IBM T. J. Watson Research Center in New York, Livermore and IBM personnel have focused on applications, leveraging experience gained with early versions of Sierra compilers. (Compilers are specialized pieces of software that convert programmer-created source code into the machine-friendly binary code that a computer understands.) With all the relevant experts gathered together in one place, hackathon groups were able to quickly identify and resolve a number of compatibility issues between compilers and applications.

In addition to such special events, a group of 14 IBM and NVIDIA personnel collaborate with Livermore application teams through the COE on an ongoing basis, with several of them at the Livermore campus full or part time. Neely notes that the COE's founders considered it crucial for vendors to be equal partners in projects, rather than simply consultants, and that the vendors' work align with and advance their own research in addition to Livermore's. For instance, Livermore and IBM computer scientists plan to jointly author journal articles and are currently organizing a special journal issue on application preparedness to share what they have learned with the broader HPC and application development communities.

Application teams are also sharing ideas and experiences with their counterparts at the other four major Department of Energy HPC facilities. All are set to receive new leadership-class HPC systems in the next few years and are making preparations through their own COEs. Last April, technical experts from the facilities' COEs and their vendors attended a joint workshop on application preparation. Participants demonstrated a strong shared interest in developing portable programming strategies—methods that ensure their applications continue to run well not just on multiple advanced architectures but also on the more standard cluster technologies. Neely, who chaired the

Livermore's Center of Excellence (COE), jointly funded by the Laboratory and the multilaboratory Advanced Simulation and Computing (ASC) Program, aims to accelerate application preparation for Sierra through close collaboration among experts in hardware and software, as shown in this workflow illustration.



meeting, says, "Most of the teams in attendance need to run and support users on multiple systems, across either NNSA laboratories or the Leadership Computing Facilities of the Office of Science. This joint workshop was designed to examine the excellent work being done in these COEs and bring the discussion up a level, to learn how we, as a community, can achieve the best of both worlds—performance and portability." The consensus at the meeting was that the COE is indeed the right vehicle for exploring solutions to the daunting challenges they all face. Neely adds, "It may be too early to declare victory, but the COEs are working well."

A Smashing Development

Preparing an application for a new HPC system is an iterative process requiring a thorough understanding of the application and the new architecture. For instance, Sierra will feature more memory and a more complex memory hierarchy in the nodes than does Sequoia. Making full use of Sierra's features requires developers to carefully manage where data are stored across the memory hierarchy and to develop algorithms with a far greater degree of parallelism—the organization of tasks such that they can be performed in parallel rather than serially. Hitting performance targets on Sierra calls for 10 to 100 times more parallelism in applications than is present today. Developers must also identify the tasks best handled by the GPUs—which are optimized to efficiently perform the

A "Hit Squad" for Troubleshooting

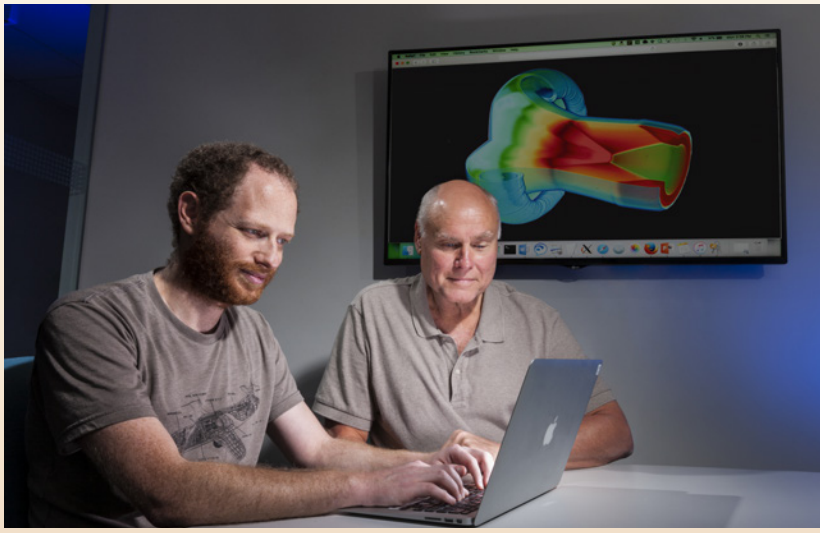
In preparing Livermore's scientific applications to run on Sierra, everyone involved contributes something unique. Application developers understand the needs of their users and the performance characteristics of their applications. Vendors have in-depth understanding of the complex architecture of the coming machine, the systems software that will run on it, and the tradeoffs made during design and development. In-house experts on Livermore's Advanced Architecture and Portability Specialists (AAPS) team provide yet another essential contribution—robust expertise in developing and optimizing applications for next-generation architectures.

"The AAPS team pulls together exceptional talent in several key areas," says AAPS lead Erik Draeger. "We have computational scientists with years of hands-on experience in achieving extreme scalability and performance in scientific applications. Together, we have multiple Gordon Bell prizes. Our computer scientists have deep technical knowledge in the latest programming languages and emerging programming models. This breadth gives the team a clear sense of what is possible in theory and where things can get tricky in practice."

AAPS team members act as a troubleshooting "hit squad," according to Institutional Center of Excellence (iCOE) lead Bert Still. At any given time, team members are working with multiple application teams to provide hands-on support and advice on code preparation tailored to the unique characteristics of the code involved. Naturally, says Draeger, the biggest challenges arise with codes poorly suited for future architectures. Nevertheless, rewriting an application from scratch is rarely an option because of its size (sometimes millions of lines of code) and the effort already expended to create and enhance that code (sometimes years or even decades). In such instances, the AAPS team uses tools such as mini-apps (compact, self-contained proxies for real applications) to help determine precisely how much code needs to be rewritten and how to do so as efficiently as possible. (See *S&TR*, October 2013, pp. 12–13.)

The team's mandate also includes documenting and transferring knowledge about common challenges and successful solutions, particularly regarding strategies for portability—the ability of an application to run efficiently on different architectures with minimal modification. The biggest portability success thus far with Sierra may be developing the copy-hiding application interface (CHAI), an effort led by developer Peter Robinson, with AAPS team support. Draeger explains, "CHAI is an abstraction that allows programmers to easily write code that minimizes data movement and runs well on a variety of architectures. CHAI solves some of the key problems of writing code for both performance and portability."

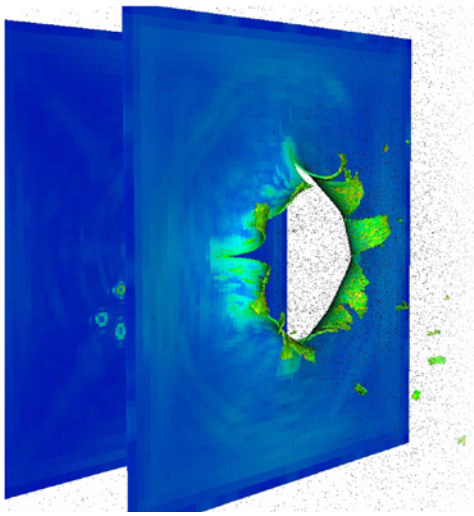
The AAPS team is currently funded as part of Livermore's COE for Sierra, but Draeger hopes the team will endure well beyond the new system's arrival. He says, "High-performance computing is not likely to get any easier in the coming years, and we need to work together and share our experiences as much as possible if we're going to continue to be leaders in the field. AAPS is one important mechanism to achieve that goal."



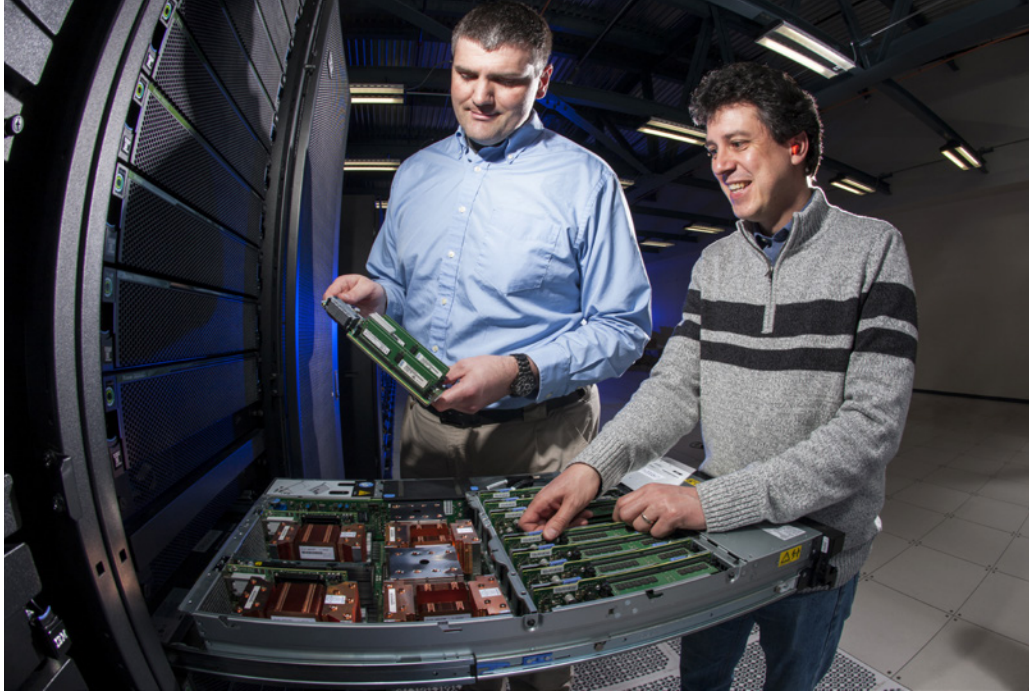
Ian Karlin (left), a member of Livermore's Advanced Architecture and Portability Specialists team, works with Tony DeGroot, part of the ParaDyn application team. (Photograph by Randy Wong.)

same operation across a large batch of data—and those to be delegated to the more general-purpose CPUs. At the same time, programmers will try to balance Sierra-specific optimizations with those designed to improve performance on a range of HPC architectures.

The iCOE team preparing the application ParaDyn readily acknowledges its challenges. ParaDyn is a parallel version of the versatile DYNA3D finite-element analysis software developed at Livermore beginning in the 1970s to model the deformation and failure of solid structures under impact. (See *S&TR*, May 1998, pp. 12–19.) The application performs well on today's HPC architectures but is proving difficult to prepare for Sierra. "We were fortunate in the past because our coding style worked well on Cray and today's HPC machines, including Sequoia," says Tony DeGroot. "Previously, we didn't have to make as many changes as other teams did to convert from machine to machine, but now we have to work harder to convert than some other teams do."



ParaDyn simulates the deformation and failure of solid structures under impact, such as the hypervelocity impact through two plates shown here. The ParaDyn team is presently exploring how to maintain the strong performance that the application achieves on CPUs while optimizing some aspects for Sierra's GPUs.



Adam Bertsch (left) and Pythagoras Watson examine components of the Sierra early-access machine, which is being used to help prepare codes to run on the full version of Sierra as soon as it comes online. (Photograph by Randy Wong.)

As with most teams, the ParaDyn crew began by assessing how the data move in the application and how work is divided up among its algorithms—the self-contained units that perform specific calculations or data-processing tasks. Edward Zywicz explains, "Unlike most physics codes, which have tens of millions of similar or identical chunks of a problem that can be processed in the same way, we have a large number of dissimilar groups. GPUs work best with large numbers of similar items. So the nature of our problems dictates some performance challenges." The ParaDyn team found the primary bottleneck to be data movement between GPUs and their high-speed memory. Consequently, the team is reorganizing and consolidating data and will need the help of the ROSE preprocessor (see *S&TR*, October/November 2009, pp. 12–13) to reduce the need for data movement. This strategy saves time and energy, the most precious resources in computing.

Pinpointing an application's performance-limiting features and testing the solutions are far more difficult when the hardware and systems software involved are still under development, as with Sierra. Hybrid CPU–GPU clusters at Livermore are a good stand-in but require far less parallelism to achieve good performance than does the more architecturally complex Sierra. Another crucial difference is that the hybrid clusters require the explicit management of data movement between CPUs and GPUs, which Sierra will not. DeGroot says, "These hybrid clusters help us identify bottlenecks and learn about our code and its performance, but they do not represent exactly how the code will perform on Sierra." To overcome this shortcoming, Livermore in late 2016 acquired two small-scale "early-access" versions of Sierra—one for ASC computing and another for other work. (These are separate from the permanent counterpart to Sierra being built by the M&IC Program.) These realistic stand-ins feature CPUs, GPUs, and networking components only one generation behind those of Sierra.

Heartening Results

Cardioid, a sophisticated application that simulates the electrophysiology of the human heart, is also being prepared for Sierra through iCOE. Developed to run on Sequoia by Laboratory scientists and colleagues at the IBM T. J. Watson Research Center, the powerful application

can, say, model the heart's response to a drug or simulate a particular health condition. (See *S&TR*, September 2012, pp. 22–25.) David Richards, one of Cardioid's developers, is leading the readiness effort. He explains, "Cardioid was highly optimized for Sequoia. We took advantage of features in Sequoia's CPU, memory system, and architecture to maximize performance. We are now trying to implement a more portable and maintainable design." What Cardioid may lose in performance will be more than made up for in improved features for researchers. Richards says, "Sierra will be six times more powerful in terms of raw flops than Sequoia. Even if we trade a little performance for flexibility, we will still be able to run more simulations, and run them faster, on Sierra." Potential collaborators have already expressed great interest in running immense, highly computationally demanding ensembles of Cardioid simulations, indicating the pent-up demand for a machine such as Sierra that can run more simulations simultaneously.

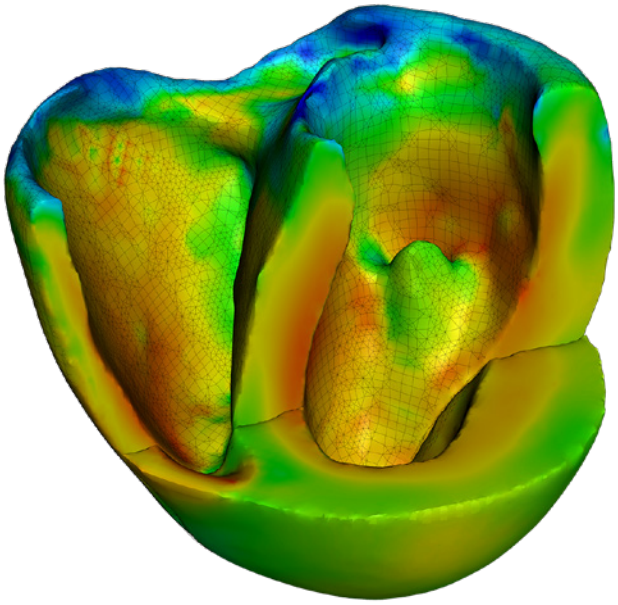
Cardioid has two major components developed specifically for Sequoia—a mechanical solver and a high-performance electrophysiology code. With help from iCOE, the Cardioid team intends to improve the portability and maintainability of the mechanics solver using components created and maintained at the Laboratory, such as the open-source MFEM finite element library. The team is also re-architecting the electrophysiology code for greater ease of use. Currently, the computational biologists and other researchers who use the code but lack advanced programming skills struggle to make complex modifications to the model. However, the new system will enable users to input their models and data in a familiar fashion. The system will then automatically translate that input into the format of OpenMP—a popular programming model that enhances the portability of parallel applications—and then produce a version of the code optimized to run on Sierra.

Code preparation is already netting promising results. When Cardioid team member Robert Blake optimized a representative chunk of the code for GPUs using OpenMP at a hackathon, he achieved 60 percent of peak speed for 70 percent of the code—a strong result. The Cardioid team also identified important issues with OpenMP that they are working to resolve in collaboration with the model's standards committee.

Looking Ahead

Thanks to the close and coordinated efforts of COE teams, vendors, users, and developers, a mission-relevant selection of Livermore's science applications will be ready when Sierra and its M&IC counterpart come online. All involved are fully confident that their preparatory efforts will enable scientists to make the most of the machines right from the start. Still says, "To further scientific discovery, we want to maximize use of the machines and minimize downtime for the five or so years of their operational lifespan."

Given the rapid evolution and obsolescence of computing hardware, the Laboratory is already planning for the generation of HPC systems to follow Sierra. Expected sometime around 2022, these systems will likely constitute another major leap in performance, up to exascale levels (at least 10^{18} flops), or 10 times the speed of Sierra. "We have a two-pronged approach for exascale," notes Neely, who is also involved in exascale planning. "We will attempt to carry forward as long as possible with existing codes. At the same time, we are exploring what we would do if untethered from existing designs to take advantage of the new architecture." (See *S&TR*, September 2016, pp. 4–11.) By increasing application portability,



Developed by Lawrence Livermore and IBM to run on Sierra's predecessor, Sequoia, the highly scalable and complex Cardioid code replicates the heart's electrical system—the current that causes the heart to beat and pump blood through the body. Sierra is expected to enable Cardioid users to run larger ensembles of high-resolution simulations than was ever possible on Sequoia.

Livermore's twin COEs aim to extend the lifetime of many of Lawrence Livermore's immense, complex, and mission-relevant applications into the exascale era ahead.

—*Rose Hansen*

Key Words: Advanced Architecture and Portability Specialists (AAPS), Advanced Simulation and Computing (ASC) Program, Cardioid, Center of Excellence (COE), central processing unit (CPU), copy-hiding application interface (CHAI), graphics processing unit (GPU), high-performance computing (HPC), IBM, institutional Center of Excellence (iCOE), Linux cluster, Multiprogrammatic and Institutional Computing (M&IC) Program, node, nonvolatile random-access memory, NVIDIA, OpenMP, ParaDyn, petaflop, ROSE, Sequoia, Sierra, T. J. Watson Research Center.

For further information contact Rob Neely (925) 423-4243 (neely4@llnl.gov).