S
N
L

H
P
C

2018

SANDIA NATIONAL LABORATORIES

HIGH PERFORMANCE COMPUTING

INVESTING IN OUR NATIONAL SECURITY

THERE IS MORE TO THESE PAGES THAN MEETS THE EYE

THERE IS MORE TO THESE PAGES THAN MEETS THE EYE

Unlock the secrets of High Performance Computing using SNLSimMagic, an augmented reality application developed at Sandia National Laboratories.

1. *Download the SNLSimMagic application on your iOS device*
2. *Use your phone to scan images with the associated icon to watch the content come to life*

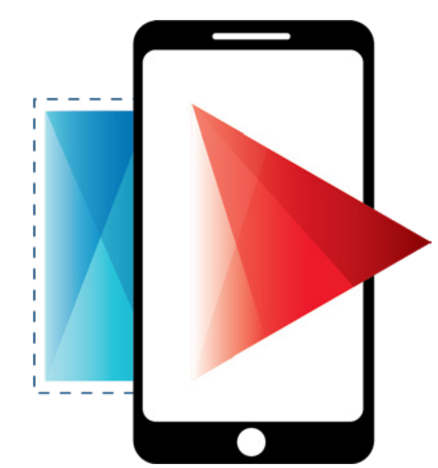


TABLE OF CONTENTS

Not a day goes by without Sandia making a major contribution to the most important national security issues. Our scientists and engineers do it all with a breathtaking range of skills.

Among our most important capabilities, and one with the broadest impact, is High Performance Computing (HPC). Current research efforts are leading to exascale computation, the next giant leap in supercomputer technology. Sandia is on the front lines of this exciting new frontier. Our country's weapons program requires computing resources of the highest speeds generating enormous data files from modeling and simulation of complex systems.

Sandia has a long history of contributions to HPC. Our innovative computer architectures and advanced simulation and modeling software and algorithms have earned international recognition, and have been used in high-consequence U.S. operations and by industry leaders to gain a competitive edge in the marketplace.

Sandia has a continuing commitment to provide modern computing infrastructure and systems in support of all our national security missions. We recently expanded Building 725 and designated it as the new and future home of HPC systems at the Labs. Astra, one of the first supercomputers based on the Arm processor architecture, is the first occupant.

NNSA/Advanced Simulation and Computing office announced that Astra, the first of a potential series of advanced architecture prototype platforms, will be part of the Vanguard program. Vanguard will evaluate the feasibility of emerging HPC architectures as production platforms to support NNSA's mission to maintain and enhance the safety, security, and effectiveness of the U.S. nuclear stockpile.

Astra, with a theoretical peak of more than 2.3 petaflops, is easily the most powerful system ever installed at Sandia and is just one of the groundbreaking HPC programs under way at the Labs.

With our computing resources and expertise, Sandia continues to provide exceptional service in the national interest and leadership in cutting-edge computing innovation.

The projects described in this edition of the HPC Annual Report will open your eyes to an amazing world of discovery.

Dr. Stephen Younger
Director, Sandia National Laboratories

Wallace Bow
Manager, R&D Science and Engineering

AUTONOMY FOR HYPERSONICS

A new arms race is emerging among global powers: the hypersonic weapon. Hypersonics are flight vehicles that travel at Mach 5 (five times the speed of sound) or faster. They can cruise in the atmosphere, unlike traditional exo-atmospheric ballistic missiles, allowing stealth and maneuverability during midflight. Faster, lower, and stealthier means the missiles can better evade adversary defense systems. The United States has experimented with hypersonics for years, but current investments by Russia and China into their own offensive hypersonic systems may render U.S. missile defense systems ineffective. For the United States to avoid obsolescence in this strategically significant technology arena, hypersonics—combined with autonomy—needs to be a force multiplier.

Achieving an autonomous hypersonic flight vehicle that can intelligently navigate, guide, and control itself and home-in on targets ranging from traditional stationary systems to targets that are themselves hypersonic vehicles—with all the maneuverability that this entails—may sound far-fetched. But to Sandia's Autonomy for Hypersonics (A4H) team, this dream is now one step closer to reality.

The A4H team is the winner of Sandia's first "Mission Campaign," a new type of transient Laboratory Directed Research and Development (LDRD) investment area championed by Labs leadership to build an innovative portfolio of projects that tackle future national security needs. The A4H Campaign will research and develop autonomous system technologies that will significantly enhance the warfighting utility of hypersonic vehicles by exploiting artificial intelligence (AI) to shorten mission-planning timelines and enable adaptive targeting decisions. This new capability will provide in-flight flexibility to adapt hypersonics to conditions such as:

1. Newly designated targets (e.g., engagement of secondary targets after the successful engagement of primary targets by leader vehicles)
2. Evolving terminal engagement conditions (e.g., mobile target with changing geodetic location)
3. New threats (e.g., emerging countermeasure capabilities presented during flight)
4. Unexpected flight conditions (e.g., adapting to sensor information that indicates a notable departure in environment and vehicle properties)

SENSE-THINK-ACT

Autonomous systems are characterized by the use of a closed-loop "SENSE-THINK-ACT" operation to achieve their desired goals. Today, hypersonic flight vehicles can and do ACT on their own. However, they require up-front, rules-based programming and they rely solely on GPS for their terminal accuracy, which significantly limits their operational relevance. A4H aims to give these systems the ability to SENSE and THINK so that they can extract information from their surroundings and intelligently adapt to changing circumstances and elusive targets. Closing the SENSE-THINK-ACT loop will provide onboard intelligence that significantly improves the ability to engage diverse targets in contested environments.

High Performance Computing (HPC) will be a key player in the success of A4H and will address the need for hypersonic engagement strategies beyond simulated scenarios. Current hypersonics rely on long-lead planning with scripted modeling and simulation to provide mission plans tailored to specific circumstances; this is a time-consuming and significant research challenge. Sandia's HPC capabilities will be instrumental in overcoming this challenge, utilizing billions of calculations to find the best tactics and solutions for this kind of modeling and simulation.

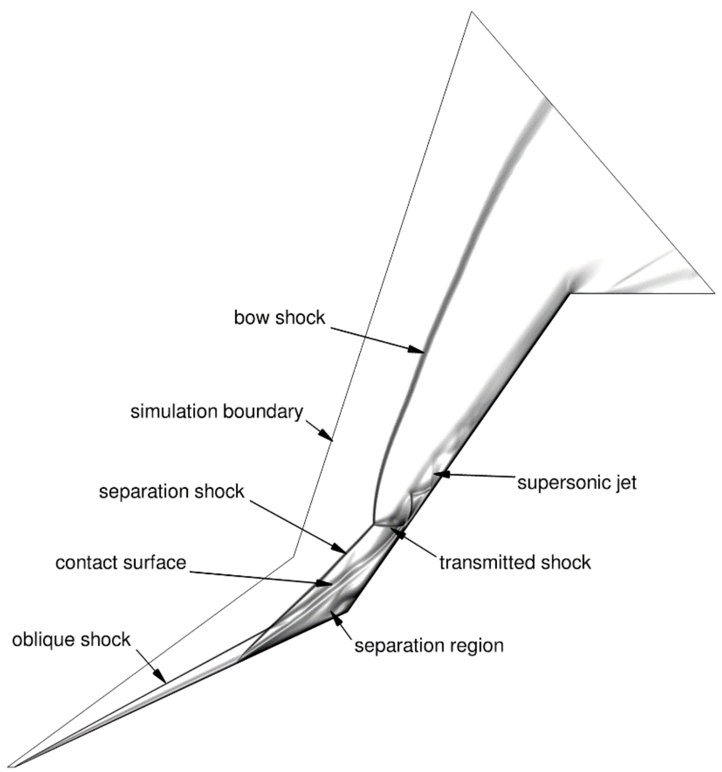
HPC will help A4H leverage modern machine learning (ML) and AI to enable quick-turnaround mission planning and analysis and to allow rapid extrapolation and generalization from previously analyzed scenarios to new and unpredicted environments. Through groundbreaking autonomy research in adversarial learning using the modeling and simulation power of Sandia's HPC, the A4H Campaign will enable transformational hypersonic capabilities that allow the United States to regain and retain leadership in this strategically significant arena, strengthen national security, and inform what is possible for next-generation projects.

AUTONOMY CAPABILITIES

Hypersonics represents one of the hardest—if not THE hardest—national security problem spaces for autonomy. The need to carry out sophisticated operations in high-consequence, adversarial environments represents a primary difference between national security and commercial applications of autonomy. Building a strong foundation in high-consequence autonomy will position Sandia to support other national security needs. The developed autonomy capabilities will extend to other Sandia flight and sensing systems in support of missions such as airborne intelligence, surveillance, and reconnaissance (ISR), space resilience, and contested space. They will also extend to other classes of hypersonic systems (e.g., air-breathing vehicles), other advanced flight systems (e.g., maneuvering reentry), and adverse-environment autonomous systems in general.

Paul Crozier
Manager, R&D Science and Engineering

SIMULATING VEHICLE ENTRY AND AIRFLOW WITH SPARC

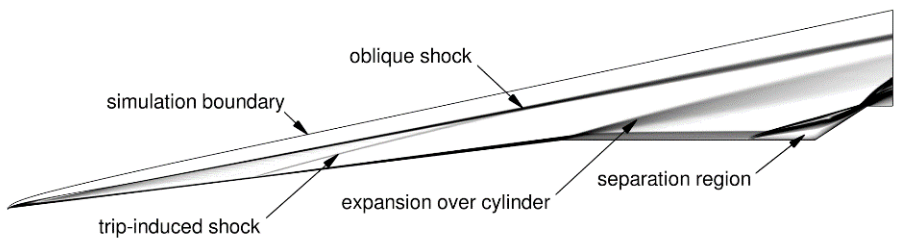


SPARC simulations capture important flow physics features that impact the design and operation of hypersonic vehicles.

The occasional rumbles and booms of impact hardware testing that penetrate Sandia's landscapes are the result of extensive preparation and teamwork. They represent the meticulous culmination of time, budgets, and brains—all colliding into a span of several seconds.

Data from such tests is used to predict how Sandia's high-risk, high-consequence products will behave upon deployment. Unfortunately, these tests cannot perfectly simulate all the environments that the items will be subjected to. For example, they cannot reproduce the exact thermal and structural elements of atmospheric reentry. These elements are primarily analyzed in virtual environments, usually with custom-built programs. While some entities (e.g., Sandia, NASA) have created and utilized modeling and simulation codes before, the Sandia Parallel Aerosciences Reentry Code (SPARC) is the first to be compatible with next-generation hardware for the exascale computing era.

Developed primarily by the Aerosciences and Computational Thermal and Fluid Mechanics Departments, SPARC is a compressible computational fluid dynamics (CFD) code that simulates aerodynamic forces and heat transfer effects. Generated data helps researchers better understand the effects of enormous pressure and heat loads placed on vehicles during their reentry trajectory. SPARC is capable of simulating both vehicle entry and airflow around the reentry body, including reactive gas flow and ablation. Sandia's Kokkos programming model allows SPARC to achieve performance portability.



Numerical Schlieren visualization of turbulent flow simulated over the HIFiRE-1 geometry using SPARC.

Beyond its utility for reentry analysis, SPARC has also enabled the development of additional hardware and software components developed at Sandia. For example, SPARC has become a primary application used to demonstrate the utility of Advanced Technology Development and Mitigation (ATDM) software components developed across multiple centers. Additionally, SPARC's use of Kokkos has allowed subject matter experts (SMEs) to consider how to leverage Kokkos for other programs, such as SIERRA. Similarly, observing how SPARC operates with hardware components will allow researchers to analyze how legacy applications will adapt to next generation hardware, including Advanced Technology System 2 (ATS-2). It will also provide high-fidelity training data for Autonomy for Hypersonics (A4H), yielding information that can be used for trajectory analysis in A4H.

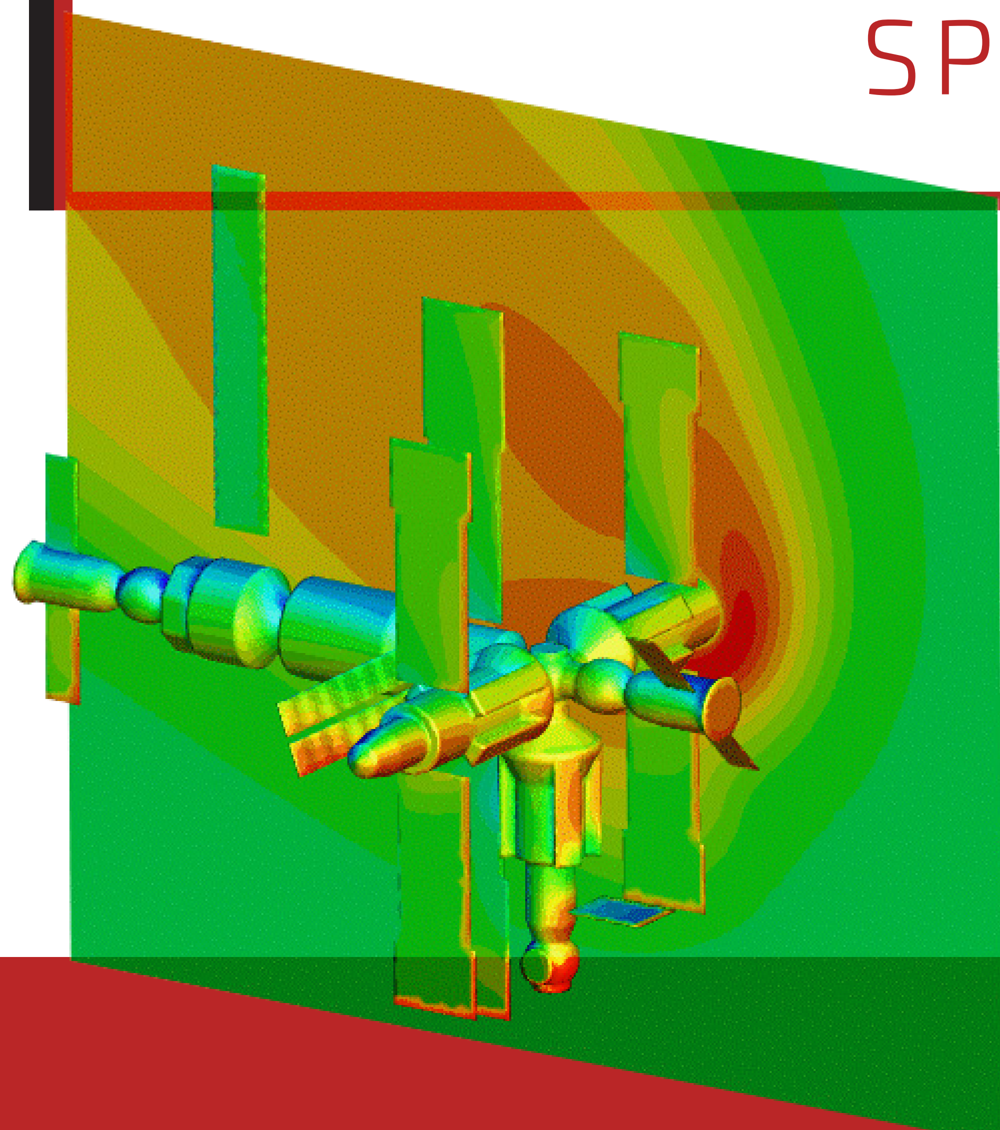
This year, SPARC began its validation process in earnest, with results being compared against experimental ground-test results. Thus far, efforts to establish code credibility are favorable. Scalability tests and validation against flight-test data in FY19 will showcase SPARC's capability to fully harness the power of next-generation platforms, including flight-test modeling for virtual reentry-trajectory simulation.

The capability to virtually model an entire flight test has many advantages. If the results of virtual flight tests can be trusted with as much confidence as those yielded during traditional tests, then SMEs will be able to make more informed decisions regarding vehicle design. Additionally, a greater variety of test conditions can be applied to simulated flight tests, including miniscule variations, which can be examined more easily and cost-effectively than traditional tests. This capability allows SMEs to draw data from more samples while understanding the precise nuances that yield certain results. SPARC can thus be used as a design tool, a qualification tool, an analysis tool, and a troubleshooting tool.

While Sandia's supercomputers cannot be heard across the desert, their whirs and hums remain continual, and SPARC is providing confidence that, while not as uproarious as a traditional test, the melody of supercomputers will be sustained for years to come.

Michail Gallis + Stan Moore
R&D Science and Engineering, Mechanical Engineering + Computer Science

PUSHING SUPERCOMPUTER LIMITS WITH SPARTA



Fluid flow around the Mir space station modeled with SPARTA. The plane shows the temperature profile, with the surface colored according to the heat flux.

Scan photo to unlock more information

Fluid flows are typically modeled by assuming continuum flow and solving the Navier-Stokes partial differential equations—a method known as computational fluid dynamics (CFD). However, for low density (rarefied) gases, such as those found at high altitudes in Earth's atmosphere, the continuum assumption breaks down, and a more fundamental method of modeling gas flows with particles must be used. In the Direct Simulation Monte Carlo (DSMC) method of molecular gas dynamics (MGD), computational particles move, reflect off boundaries, and collide like real molecules. DSMC has typically been applied to hypersonic reentry at high altitudes; however, increasing computational power and a recently developed highly scalable DSMC code have allowed Sandia researchers to apply DSMC to systems that would have been impossible just a few years ago.

Exascale computing, despite its 50x leap over current petascale computing capabilities, does not guarantee that using larger and more refined CFD simulations will produce a corresponding leap in the accuracy of simulations. Traditional CFD can only be as accurate as the underlying equations being solved. Subject matter experts (SMEs) realized that there may be shortcomings in those equations and that including additional physics found at the molecular level could improve the accuracy of CFD simulations. Researchers postulated that these difficult problems would need to be examined on a smaller, more fundamental scale via MGD as embodied in the DSMC algorithm, especially when it came to analyzing instabilities and turbulence.

In 2011, the Stochastic Parallel Real Time Gas Analyzer (SPARTA) code was developed at Sandia as a joint effort between the Thermal/Fluid Component Science and Multi-Scale Science Departments. The SPARTA code uses the highly scalable message-passing interface (MPI) domain decomposition method to run on the world's largest supercomputers. SPARTA has run on the entire Sequoia supercomputer at Lawrence Livermore National Laboratory (LLNL) using 1.57 million cores simulating over a trillion computational particles. Additionally, SPARTA was the first science code to run on the entire Trinity platform at Los Alamos National Laboratory (LANL). The Trinity platform includes both Intel Haswell CPUs and many-core Intel Knight's Landing Xeon Phis, which required a heterogeneous approach to use both types of resources effectively. SPARTA ran on over 19,000 Trinity nodes using more than 1.2 million MPI ranks. SPARTA is also being prepared for future exascale computing platforms by incorporating the Kokkos performance portability capability, also developed at Sandia. Kokkos allows the same C++ code to run on GPUs and multi-threaded CPUs without requiring the developer to write target-specific code, such as CUDA.

DSMC and SPARTA have allowed Sandia researchers to study problems beyond Sandia's mission as well. For example, SPARTA has been applied to investigation of the Columbia space shuttle disaster, satellites orbiting Mars, and concepts for a new probe to sample Venus' upper atmosphere.


Walt Witkowski + Angel Urbina
Senior Manager + Manager, R&D Science and Engineering

VERIFICATION AND VALIDATION

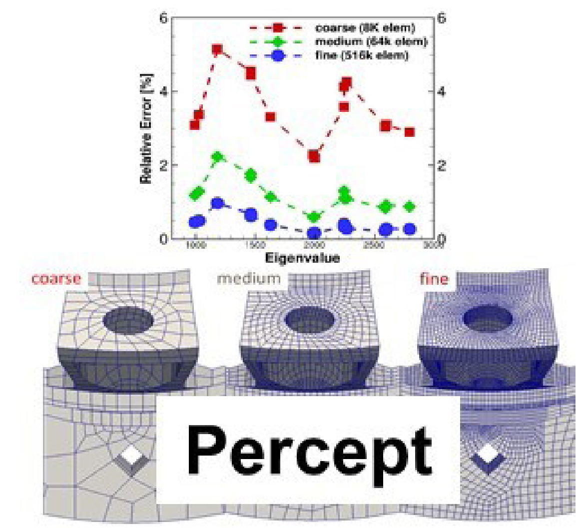
To err is human. Such a mantra is arguably essential for a contented life. It reminds us that, because of constant variability, not all of our experiences will be perfect ones, and thus persistence is more vital than precision. While it is a viable attitude for our personal lives, sometimes exactness is crucial—such as when dealing with products that must always function when needed, and never otherwise.

When the United States discontinued nuclear testing in 1992, the frontier of computational simulations was introduced to supplement qualification testing activities. New Advanced Simulation and Computing (ASC) credible simulation capabilities allowed subject matter experts (SMEs) and decision-makers to continue to verify and revise designs. Computational simulations allowed SMEs to explore previously untested scenarios and qualify previously unmeasurable aspects of their research, namely situations that simply were not possible to observe through traditional field testing (e.g., catastrophic accidents on transport aircraft).

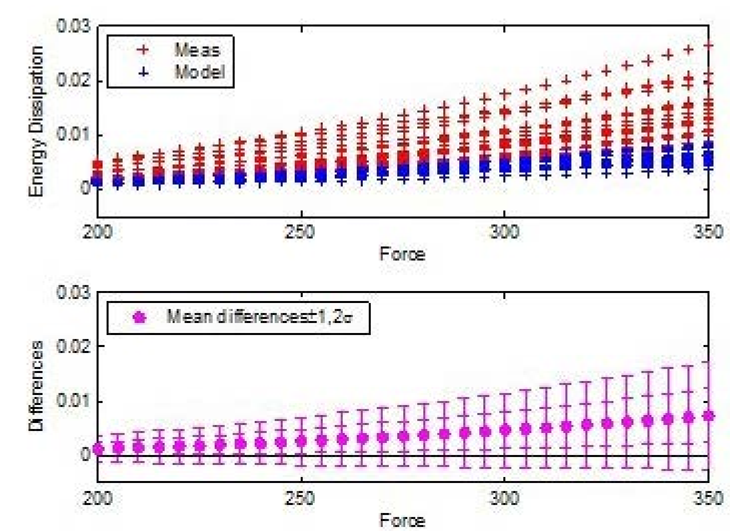
Despite their advantages, SMEs continued to recognize that computational simulations are approximations, adhering to the adage that, while all models are wrong, some remain useful. While the mathematics and algorithms applied to the ASC models were relatively sound, the platforms on which they were run introduced a new generation of verification challenges. Despite their revolutionary and appealing nature, SMEs had to verify how trustworthy ASC simulation capabilities truly were, especially if reliance on them was increasing.

Sandia's Computational Simulation Group develops verification and validation (V&V) philosophies, methodologies, and tools to help code developers and analysts investigate codes and models for ASC platforms. Verification activities explore whether codes have been written and implemented as intended by determining if their algorithms function and converge as expected. Uncertainty quantification (UQ) activities allow SMEs to provide a mathematical description of inputs and parameters to understand the response variability in a computational simulation, thus helping to assess model validity for its intended use. V&V activities focus on the quantities of interest obtained, directly or indirectly, from computational simulations so that SMEs can ultimately understand the nature of errors yielded, including their characteristics and magnitude. This knowledge helps them determine potential solutions for errors, specifically by providing a fundamental point at which to test the pedigree of the solutions.

In recent years, quantification of margins and uncertainty (QMU) has gained increased attention from Sandia's nuclear deterrence customers. V&V plays an important role in QMU, as it informs the uncertainty aspects of QMU work. Additionally, the ability to apply modern V&V activities to key disciplines such as thermal and solid mechanics, structural dynamics, and electrical and radiation science increases Sandia's research scope into areas that nuclear deterrence customers wish to see increased QMU calculations. V&V also benefits significant finding investigations (SFIs): high-visibility/high-risk events that require a high level of scrutiny and technical rigor. The contribution of V&V to SFIs includes the formality in assessing the technical process used to verify and validate a computational simulation and/or the quantification of uncertainty.



Uniform mesh refinement performed with Percept toolkit in SIERRA. Percept includes a capability that can refine very large meshes in parallel.



Model validation comparing CompSim results and experiments.

While V&V partners have yet to be embedded within every Sandia project team, they are currently working with analysts who support life extension programs and principal stockpile systems and components. Over time, V&V SMEs hope to join even more projects to provide consistency as well as a standard approach for design and assurance of design throughout the laboratories.

Failure is unacceptable when lives are at stake. However, assuring perfection is exceedingly difficult—arguably impossible—when variability is inevitable and flawless uniformity simply does not exist. Sandia SMEs must constantly balance between the need for exactness and the need to assure performance between variability. Thanks to V&V, the ability to claim both is increasing. Sandia's methods continue to render exceptional service in the national interest not only with the tangible items produced, but also the methods used to create and verify them, especially as some Sandia V&V staff also serve on national standards committees.


Sarah Kiewig + Jaideep Ray
R&D Science and Engineering, Computer Science

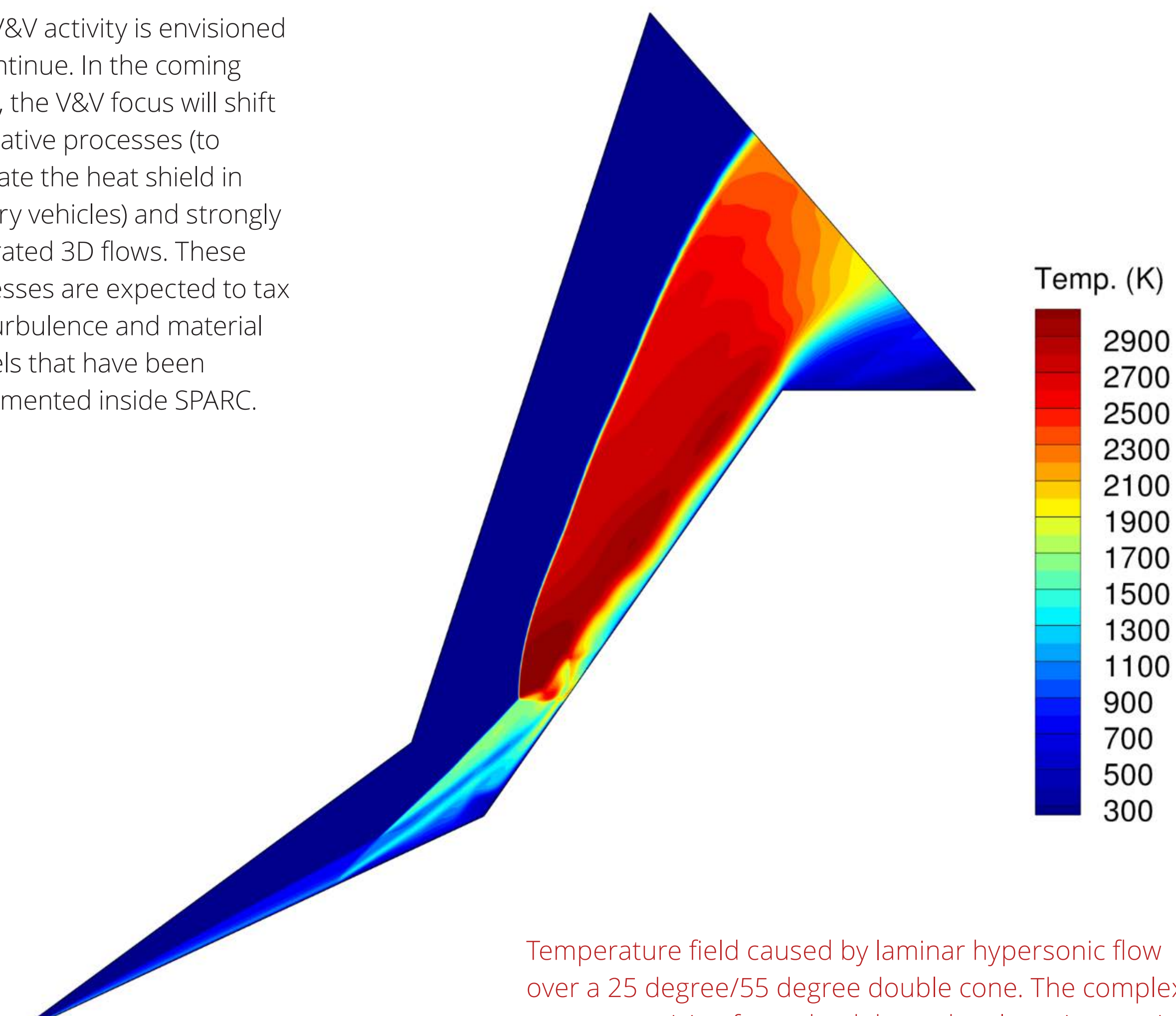
V&V FOR HYPERSONIC VALIDATION

Guilty until proven innocent. A self-proclaimed “group of doubters,” the mechanical and aerospace engineers at the Verification and Validation (V&V), Uncertainty Quantification (UQ), and Credibility Process Department are evaluating computational fluid dynamics (CFD) models that are used to simulate hypersonic airflows (6–20 times the speed of sound) around reentry vehicles. The department’s UQ effort operates under the premise that model predictions and their uncertainties need to be assessed to understand the bounds of their applicability. One way to achieve this is to compare model predictions against corresponding experiments. To that end, researchers examine thousands of medium-scale simulations (hundreds to thousands of central processing unit [CPU] hours per invocation), seeded with different inputs and flow environments, to automatically assess their likelihood of being able to reproduce experimental measurements. This endeavor lies at the heart of their effort to verify the Sandia Parallel Aerodynamics and Reentry Code (SPARC).

Under the Advanced Technology Development and Mitigation (ATDM) program, Sandia is engaged in restructuring its simulation software so that researchers may efficiently work with the novel High Performance Computing (HPC) architectures envisioned in the next decade. At times, the restructuring is quite extensive, requiring researchers to ensure that the accuracy, robustness, and fidelity of the physics and chemistry encoded in the software has not been impaired. Comparing with experiments is the simplest means of doing so. Unfortunately, the extreme environments targeted by SPARC simulations are difficult to reproduce inside conventional test environments (e.g., wind tunnels), and experimental data is sparse or replete with uncertainty. Thus, a computational model’s inability to match experiments may be due to approximations inherent in the computational model, shortcomings of the experimental data, or errors introduced during software modification. Performing this attribution, largely an exercise in numerical analysis and statistical inference, is the core task of the SPARC V&V team.

The physical and statistical models in SPARC are currently being validated against experiments in which flow is sufficiently energetic to cause oxygen and nitrogen to react together. These aero-thermodynamic phenomena occur in both laminar and turbulent flows. Recently, research has uncovered situations where experimental data is likely suspect, and others where assumptions, which were needed to fill gaps in experimental data, were shown to be incorrect. Such a thorough testing of hypersonic models is rare in published literature and is enabled by the abundance of HPC resources that Sandia has made available (both institutional and Tri-Lab, accessed via the Advanced Technology Computing Campaign).

This V&V activity is envisioned to continue. In the coming years, the V&V focus will shift to ablative processes (to simulate the heat shield in reentry vehicles) and strongly separated 3D flows. These processes are expected to tax the turbulence and material models that have been implemented inside SPARC.



Temperature field caused by laminar hypersonic flow over a 25 degree/55 degree double cone. The complex structure arising from shock-boundary layer interaction at the junction of the two cones is clearly evident.


John Korbin + David Peterson
R&D Science and Engineering, Mechanical Engineering

CREATING ASSURANCE WITH COMPUTED TOMOGRAPHY

Critical safety devices, such as airbags in passenger vehicles, have recently been scrutinized for manufacturing defects. Assurance that a certain lot testing of the components used in these products, such as a lot of pyrotechnic generators, had passed a series of safety and qualification tests would provide some measure of reassurance. But imagine being told that the very vehicle safety system that you drove off the lot underwent and passed its own test that pledged a specific performance envelope for safe use.

Sandia's computed tomography (CT) work, performed by the Computational Shock Physics Department, is doing just that. Thanks to CT, subject matter experts (SMEs) can scan and create digital twins of individual components for qualification purposes. Whereas many acceptance tests take only one sample component for testing, CT scan-based approaches subject each individual item to their own test scenario. High Performance Computing platforms allow for these replicas to be exact copies of their "real-world" counterparts at the infinitesimally microscopic scale (tens of micrometers). Using image segmentation, Sandia's digital models are four orders of magnitude higher fidelity than similar modeling scans. SMEs achieved this capability by finding a scalable algorithm to apply to the scanning process. With it, scaling limits that previously applied to image segmentation (the previously laborious process of turning the CT data into useful material volumes) are removed.

Even when a part is flawlessly designed, variance between individual units can still surface during the manufacturing phase. Precisely calibrated machines, including 3D printers, will not create exact replicas each time, and even miniscule differences between individual parts could cause variance in performance (an issue known as geometric uncertainty). Because changes between items can occur at the manufacturing level, testing prototypes or representative samples does not necessarily provide exact data on performance. In addition, representative testing does not necessarily yield total surety of performance, but rather performance boundaries that define a certain margin of acceptable risk for an item. For example, representative samples can help qualification test personnel assert that a certain item has a 3% chance of performing outside of its intended performance window. Additionally, qualification testing frequently necessitates the sacrifice of some parts, thus increasing costs.

CT-based simulated methods, on the other hand, create assurance that is not bound by a lot error percentage. CT also allows researchers to pinpoint potential errors in items, which can vary between individual items of the same type. Moreover, digital twins can be subjected to many different test scenarios many times, sometimes examining a single component through thousands of runs. CT does not eradicate the need for statistical analysis; instead, it provides higher confidence for statistical analysis, minimizing the performance error window to fractions of a single percentage. Finally, because no physical parts have to be sacrificed, CT-based methods create substantial cost savings. CT continues to create certainty across Sandia, and may soon transition to the corporate world as well.

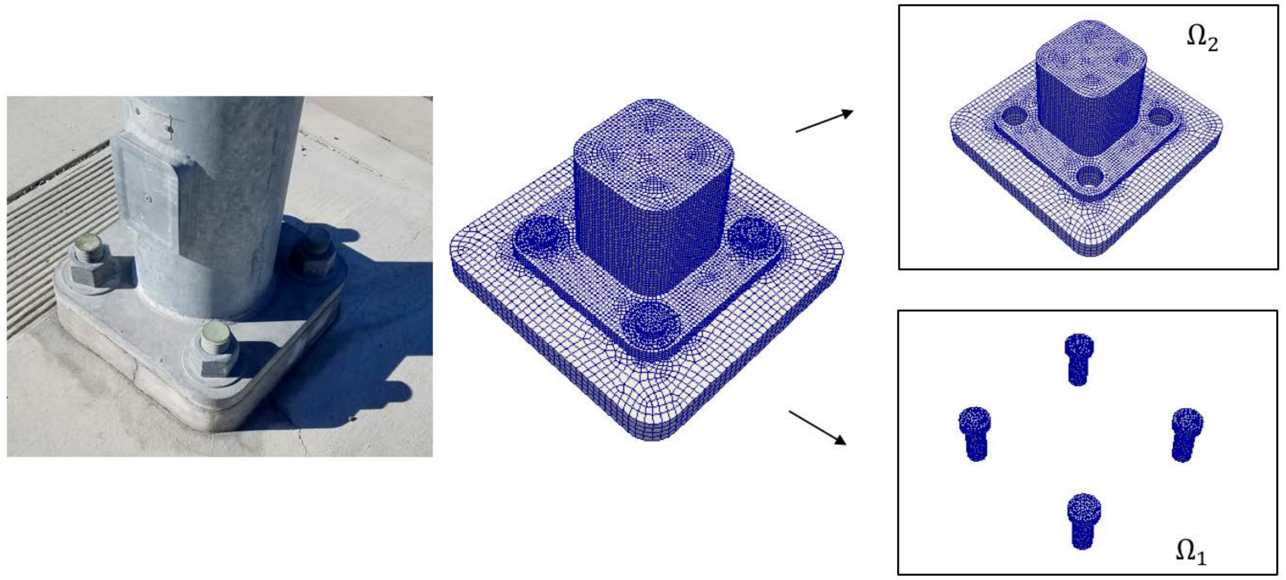
Left to right: CT scans of Nintendo Entertainment System (1983), Super Nintendo Entertainment System (1990), and Nintendo 64 (1996) gaming console controllers. Potential failure points increase as products evolve in technical complexity.

Alejandro Mota + Irina Tezaur
R&D Science and Engineering, Mechanical Engineering + Computer Science

CONCURRENT MULTISCALE COUPLING IN FINITE DEFORMATION SOLID MECHANICS

It's easy to overlook the small and simple steps required for success while we work on achieving our personal goals; however, when it comes to component testing, single parts or failure points are often the culprit for devastating outcomes. Nonetheless, it can be difficult to explore all such possible failure scenarios for any given test, especially since they tend to occur during interactions between small-scale (e.g., bolts, screws, welds) and large-scale (e.g., components, full engineering systems) elements.



Example of a bolted joint geometry coupled iteratively with the Schwarz Alternating Method.

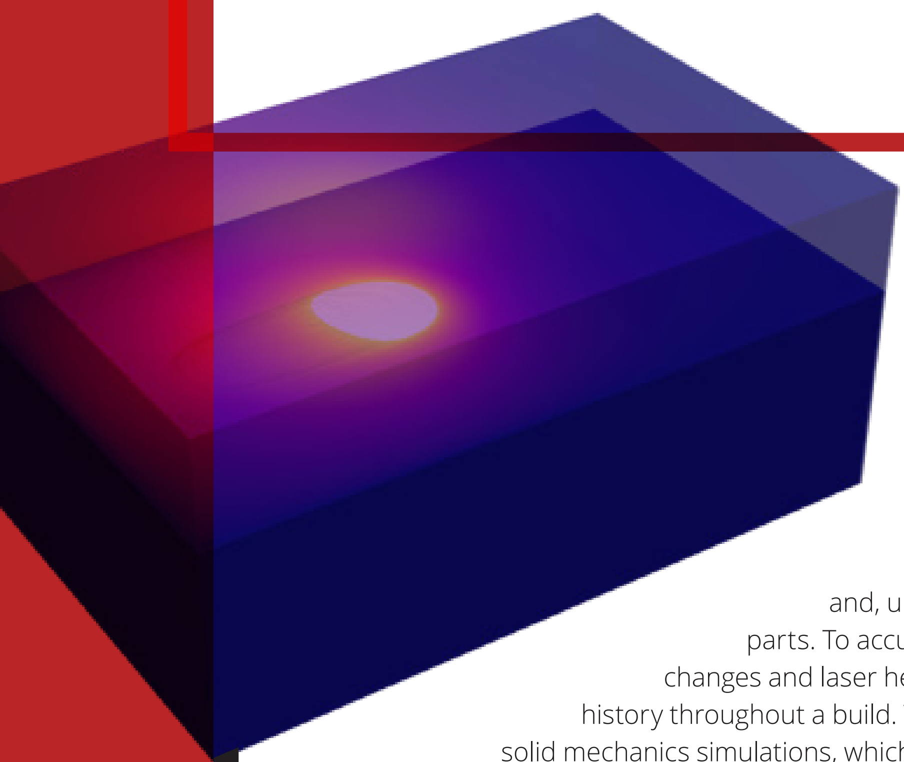
Sandia researchers have found a means to explore as many configurations as possible during component testing through the application of the Schwarz Alternating Method to computational modeling. This method is based on a straightforward concept: solutions to partial differential equations (PDEs) in simple domains can be used to iteratively piece together a solution in a more complex domain. Sandia researchers have adapted the method for nonlinear finite deformation inelastic problems and created an efficient, non-intrusive implementation that makes it possible to explore the interaction of phenomena at different scales. Using High Performance Computing (HPC) for such efficient computations, the group is able to test many different types of configurations virtually to predict the behavior of small-scale parts that can cause the failure of an entire system, thus enabling substantial savings in both funds and computational resources.

The Schwarz Alternating Method allows for concurrent multiscale coupling of regions with different meshes, element types, and levels of refinement. The meshes need not be conformal, and coupling different material models is possible provided they are compatible in the overlap region. This coupling allows for a "plug-and-play" framework in which meshes for different components can be exchanged without the need to re-mesh a complex geometry consisting of a variety of scales (a task that can take upwards of several months).

The new method has been prototyped on a number of quasistatic and dynamic problems within Sandia's HPC Albany/Laboratory for Computational Mechanics research code. Researchers have begun to apply the method to problems of interest to production (e.g., multi-scale problems with fasteners), with promising results. The next step is to implement the method within Sandia's SIERRA/SM production code, thus enabling analysts to simulate complex multiscale engineering systems possessing millions of degrees of freedom.

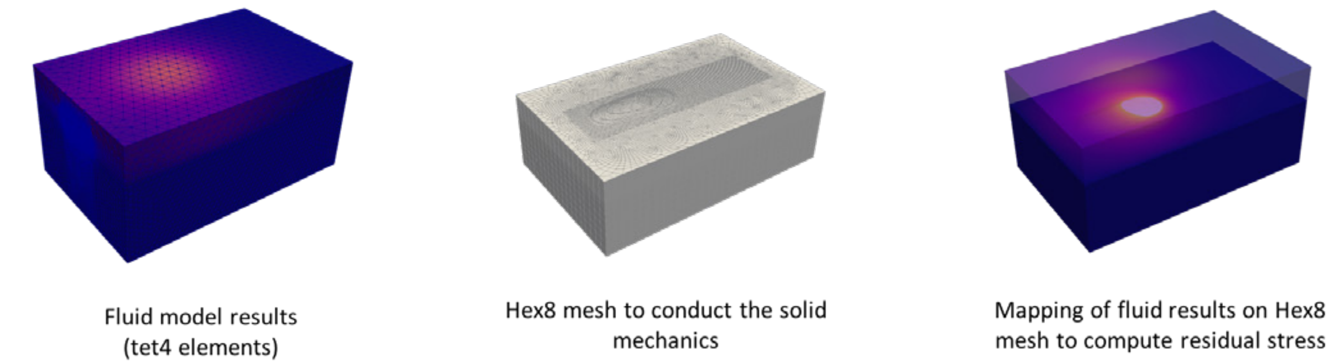
Lauren Beghini
R&D Science and Engineering, Mechanical Engineering
Contributing researchers: Mike Stender, Mike Veilleux, Kurtis Ford, Brad Trembecki, Sam Subia

ADDITIVE MANUFACTURING PROCESS MODELING IN SIERRA



Additive manufacturing (AM) is a capability that streamlines material fabrication and shaping into a single process. One widely used AM process is the laser engineered net shape (LENS) manufacturing process, which uses a laser to melt and add powdered metal to surfaces. Simulation of the AM process is important to gain insight into material behavior and, ultimately, performance of additively manufactured parts. To accurately simulate the AM process, surface topology changes and laser heat input are tracked to provide a detailed thermal history throughout a build. These thermal histories are then used to inform solid mechanics simulations, which can predict the evolution of residual stresses, distortions, and the microstructure within a part.

One area of study during FY19 has focused on integration of AM part-scale models with other Sandia AM mesoscale models. Researchers have been able to leverage their previous work by creating a new thermal mechanical part-scale modeling methodology that couples high-fidelity fluid models of melt behavior with SIERRA solid mechanics modeling. By using the high-fidelity fluid model, the surface topology and melt pool dynamics are accurately captured for radiation and convection, providing a detailed thermal history throughout the build. Then, to communicate between the models, the high-fidelity thermal field data is mapped onto a hexahedral mesh (in SIERRA Adagio) as an input to the solid mechanics simulations, which are capable of predicting the evolution of residual stresses and microstructure within a part.

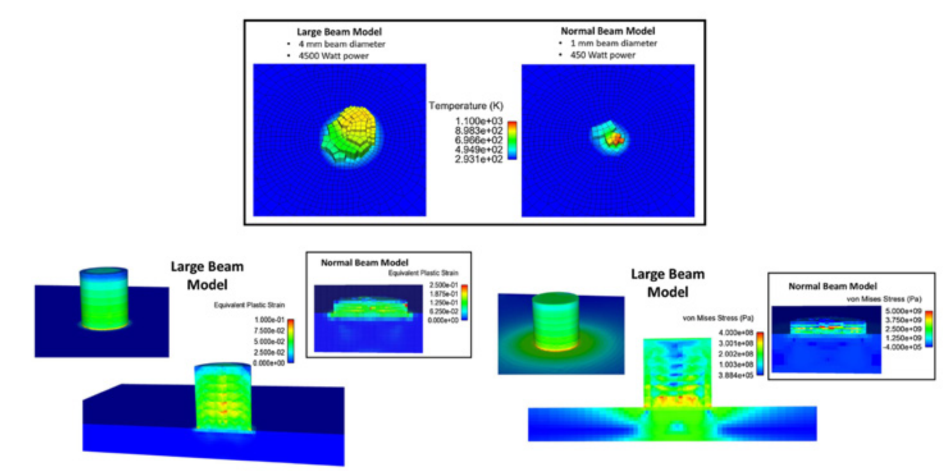


Demonstration of mapping high-fidelity fluid model results onto solid mechanics input mesh for coupled approach

One of the primary challenges for process simulation is optimizing computational throughput. AM simulations are extremely costly due to long build times, small melt pools, high thermal gradients, and the continuously evolving mesh topology. This year, researchers on a Laboratory Directed Research and Development (LDRD) Grand Challenge project, led by Allen Roach, improved the robustness and speed of modeling the LENS AM process. They employed a “lumped laser method” to improve computational efficiency, where the spot size of a laser was increased and tuned to significantly reduce run time while still capturing the general trends in residual stress and distortion behavior.

Meanwhile, the researchers are working in collaboration with The University of California, Davis on validation of the methodology for button builds with different laser scan patterns (spiral in, spiral out, and cross-hatch) and variations in base plate size. These activities measure residual stress using both the contour method and slitting activities by cutting apart the specimens and measuring the out-of-plane deflections to back-compute the state of stress. Preliminary results are promising.

Future work includes improvements in the robustness of the coupled simulations, enhancements to the solid mechanics constitutive model (to include near-melt plasticity and capture melt behavior), developing increased confidence in these models through continuous validation, and ensuring usability both within and outside Sandia. LENS AM work may also enable researchers to engineer a residual stress state to design improved parts that could be “born qualified” and ready to put in a weapon system.



Comparison of temperature profiles, plastic strain, and residual stresses for “lumped laser method” and normal beam model for 304L stainless steel build


Brian Naughton + David Maniaci
R&D Science and Engineering, Materials Science + Aeronautical Engineering

WIND ENERGY TECHNOLOGY

At first glance, it does not seem like the enormous wind turbines that dapple the nation's landscapes need much improvement—after all, few common structures mirror their grandeur while simultaneously instilling a sense of our own smallness. Especially when one recalls the form and function of the comparatively modest predecessor of the wind turbine (the windmill), it appears even more apparent that our ability to harness and utilize wind power has reached its structural and technical apex. The Department of Energy Wind Energy Technologies Office (WETO), however, is proving otherwise.

SpinnerLidar instrument installed on a SWiFT turbine. The instrument produced the wake measurement data used for wake dynamics analysis.

Scan photo above to unlock more information

WETO subject matter experts (SMEs), including those in Sandia's Aerosciences and Wind Energy Technologies Departments, aim to improve both the technologies and the methods used for wind power analysis. Their work enables the accelerated development of wind energy technology and thus the potential for wind to become a more cost-effective, practical, and sustainable energy source for the nation. SMEs hope for wind energy to make up 35% of the power supply in the United States by 2050.

Recent HPC activities have primarily supported the field of wake dynamics: the characterization and analysis of wind flows generated within a wind plant. On the smaller scale, turbines create wakes that increase turbulence around the wind plant, thus generating a lower speed resource for the other turbines to ingest (even micro-scale measurements of turbulence can affect turbine blades and reduce the efficiency of turbine output and increase structural damage). Using HPC platforms, SMEs are able to model turbulence flows and assess potential solutions regarding optimal turbine design, placement, and operation. Beyond the wind farm, the high-fidelity models can be expanded to investigate the impact of regional weather on the wind plant production, acoustic propagation to reduce impacts to local communities, and the interaction between neighboring wind farms.

Due to the high computational cost of high-fidelity wind plant models, WETO SMEs aim to use multilevel, multifidelity uncertainty quantification (UQ) methods to propagate uncertainties using a suite of wind models. Such multilevel multifidelity methods would allow SMEs to utilize a framework of computational models to estimate production uncertainty in future wind plants. Additionally, measurement data and modeling components could, over time, allow SMEs to perform more efficient and effective testing, atmospheric modeling, high-fidelity wind plant modeling, and controls research to improve wind plant efficiency and reliability.

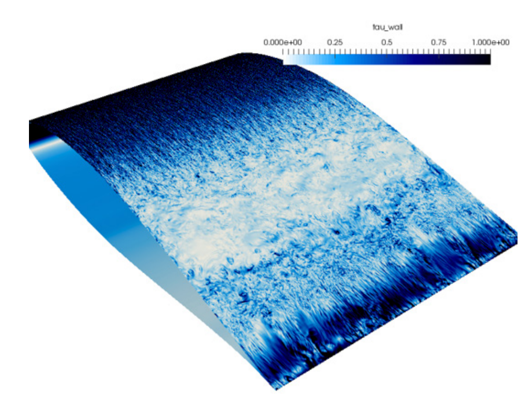
Despite its rapid progress and numerous applications and capabilities, current petascale computing resources limit the scope of what can be explored in wind energy technology. Thus, WETO SMEs have launched the ExaWind Initiative. The next generation computing platform for wind farm simulations, ExaWind aims to adapt the Sandia-developed computational fluid dynamics (CFD) code Nalu to leverage the capabilities of next-generation HPC exascale architectures. By the time the hardware is operational, SMEs will be prepared to simulate full wind farms, orders of magnitude more than today's analysis allows. Researchers hope to start using these capabilities between 2021 and 2023.

The potential for ExaWind to enhance the present and prepare for the future will enable the nation to develop the tools and technologies required to optimize and therefore generate even more wind power. It would allow us to retain energy independence as well as garner diversity in our energy portfolio by reducing our dependence on fossil fuels. Such modeling is also showing its potential to improve weather-forecasting models. Investments in the fundamental science and computing capabilities now will lead to the reliable, cost-effective, and secure energy technologies for the nation well into the future.

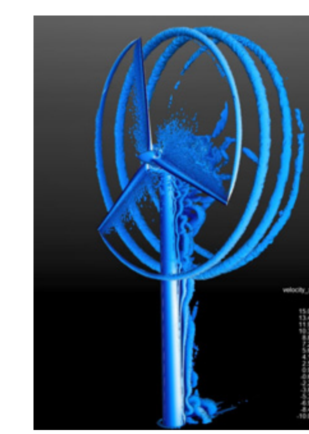
INSIDE THE SWiFT FACILITY

In addition to refinement of wind turbine design, SMEs also perform system-level improvements, such as optimizing the design and operation of an entire wind farm. Using the Sandia Scaled Wind Farm Technology (SWiFT) facility at the Texas Tech University National Wind Institute Research Center in Lubbock, WETO SMEs are also able to analyze more nuanced aspects of wind power improvement, including the influence of atmospheric conditions on the wind resource, the interaction between wind turbines in a farm through complex flow, and the interaction of the wind farm with other generation sources and the grid. The wide range of length and time scales as well as the variety of physical phenomena involved in these processes require the resources of modeling and simulation to sufficiently explore technology improvement opportunities.

The SWiFT facility supports a variety of research and development objectives, including the discovery of fundamental aerodynamic and structural phenomena, demonstration of new wind energy technologies, and the generation of high-resolution experimental data to support the development and validation of high-fidelity models used for simulations on HPC resources. Because testing a large number of full-scale turbines is extremely costly and time-consuming, it would take upwards of several years to acquire the data required for effective analysis, significantly hindering the pace of innovation and technical development. Thanks to HPC modeling platforms, however, larger numbers of tests in different iterations can be simulated at an accelerated rate.



Turbulence modeled on a turbine blade section with Nalu.



Wind turbine velocities modeled with Nalu.

(Credit: Matt Barone, Stefan Domino, and Chris Bruner)

Scan photos above to unlock more information

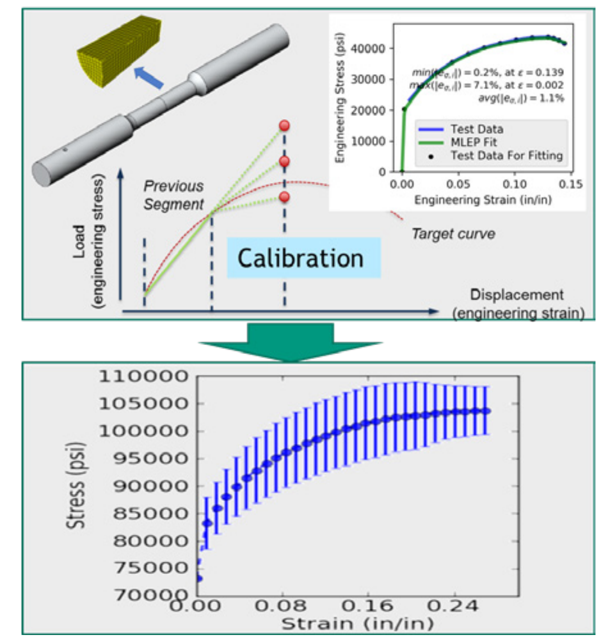
George Orient
R&D Science and Engineering, Mechanical Engineering

B61-12 NOSE BOMB SUBASSEMBLY

A robust analysis workflow developed by the B61-12 Solid Mechanics (SM) team was used to predict nose bomb subassembly (NBSA) fuzing performance.

SM modeling informs B61-12 qualification evidence for meeting impact fuzing requirements. Uncertainty quantification (UQ) is a key element of model credibility for high-consequence engineered systems. UQ helps researchers determine how known and characterized input uncertainties, such as constitutive behavior (Figure 1), signal-processing parameters, and numerical factors, influence predicted system performance. For highly nonlinear and potentially discontinuous system responses, it is impossible to define worst-case deterministic conditions, and UQ is required to bound predicted performance variations.

FIGURE 1



Constitutive model calibration process quantifies an input uncertainty.

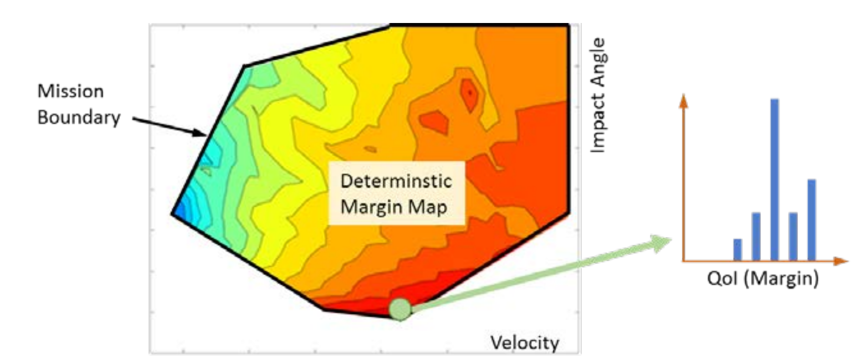
A configuration-controlled automated analysis workflow was used to construct a map of key system responses over the impact velocity/impact angle space for water target impact. The same analysis workflow was driven by Dakota to perform probabilistic validation assessment of regions of the mission space where deterministic margins are predicted to be low (Figure 2). Probabilistic sensitivity measures have identified the most significant factors influencing fuzing performance. Additionally, validation evidence comparing experimental observations and model predictions contributed to the credibility of the SM modeling for predicting impact fuzing and margin.

The modeling, including sensitivity and uncertainty analyses, will be included in the upcoming B61-12 NBSA Qualification Evaluation Release in mid-FY19. The current effort required approximately 1,500 instances of the SIERRA NBSA model each running on 1,000 processors for 5–15 hours, which was executed on the Trinity Advanced Simulation and Computing (ASC) platform to evaluate uncertainty. An additional 1,500 instances have been executed to assess performance and margins for multiple target compliances.

The custom software infrastructure used for this effort was groundbreaking, as it involved performing UQ on a full system mechanical model. The next goal is to democratize this process. An ASC Level 2 milestone is planned for FY19 to demonstrate a reliable, resilient, and user-friendly analysis workflow engine technology in Sandia Analysis Workbench, which will enable ModSim teams to routinely perform similar uncertainty-informed margin assessments in upcoming weapon qualification programs.

FIGURE 2

V-gamma contour map of fuzing time for water target illustrates (a) qualification evidence provided to NBSA product team and (b) validation evidence accounting for uncertainties.



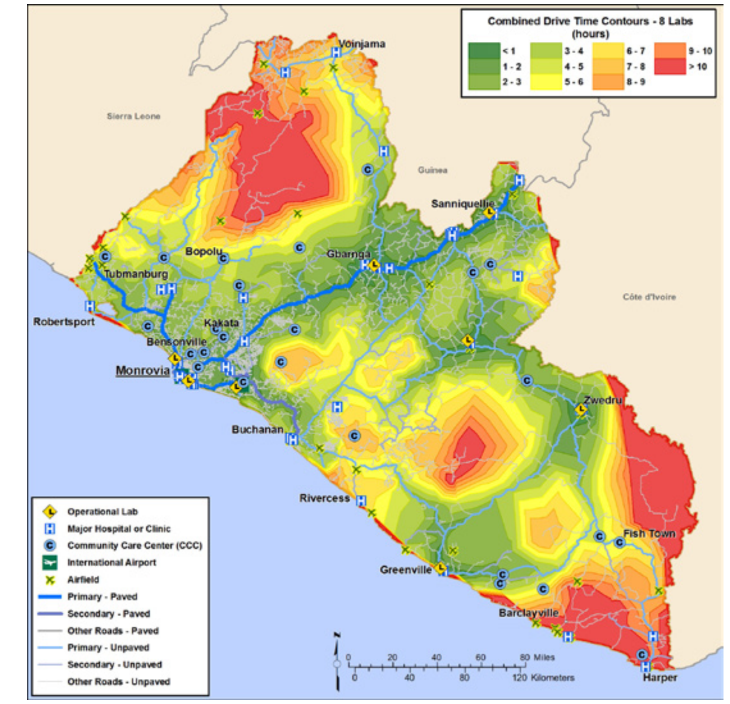
Patrick Finley
R&D Science and Engineering, Computer Science

TRANSPORTATION MODELING AND GLOBAL HEALTH

A patient in the United States has been diagnosed with Ebola. Fear and panic spreads across the country, and hospitals are inundated with hundreds of people—some infected with the highly contagious disease and others not. Blood tests are needed for positive diagnoses, but the diagnostic labs are overwhelmed with blood samples to test, and staff are overworked and stressed. Infected people need to be quarantined and treated, but it's hard to find rooms to quarantine so many patients. Sick people who need triage and regular care for other emergencies are afraid to go to hospitals for fear of Ebola, which has a 50% fatality rate. And since hospitals are so overwhelmed, sick people often stay home, infecting heathy people around them; the United States is now in the grips of a full-blown Ebola outbreak.

Sandia's high performance computers simulated such a nightmare scenario recently, and with good reason. An Ebola outbreak in the United States could be devastating if hospitals are not prepared. When an Ebola outbreak in West Africa became a global concern in 2014, health advisers were alarmed at the length of time it took to properly diagnose infected people. In rural areas in Liberia, for example, blood samples from ailing people would be sent to a laboratory for testing, but the closest lab was hundreds of miles away through difficult and sometimes impassable roads. In more urban areas, blood samples would be sent to nearby labs, but those labs were often already overburdened by the sheer volume of samples to test. Staff at some treatment centers were unaware that a lab a little farther away might have the capacity to take in more samples. Meanwhile, undiagnosed infected people were unknowingly spreading the disease to many others around them, worsening the outbreak.

The U.S. Defense Threat Reduction Agency (DTRA) and Centers for Disease Control and Prevention (CDC) posed a serious question: how do we improve blood-sample transportation routes in Liberia to ensure that samples taken from ill people are tested as quickly as possible, ensuring a proper diagnosis and faster treatment? Sandia scientists, already experts in transportation modeling for nuclear materials, quickly swarmed on this problem.



Map of drive times to operational labs in Liberia

The Sandia Ebola response team immediately set out to collect data from the region using available maps and local information, and transformed the raw data to GIS maps. Then, applying Sandia transportation routing algorithms, the team identified the optimal routes to get blood samples to the best laboratory for testing, even if that lab was not geographically the closest. The models also showed the best possible locations for mobile diagnostic laboratories that would better support the very rural regions that were most affected by the Ebola outbreak.

The team relied on the immense computational power of the Red Sky high performance computer to verify that the new sample transport routes were not only fast, but also minimized the chance for transmission of the Ebola virus among Liberian patients in hospitals and communities. After building a computer model of the healthcare system of the entire country, the team used that model to show how many individuals would contract Ebola under different scenarios. They then ran the model millions of times on Red Sky, using every conceivable combination of disease propagation factors and control measures. The power of this massively parallel parameter study enabled Sandia researchers to define confidence metrics for Ebola cases prevented by the new transport routes, and to ensure that gains would hold up even if the pattern of disease spread had changed dramatically.

Sandia was uniquely suited for this project due to its experience in global health security combined with its computer modeling capabilities, and the efforts of its transportation mapping and validation using Red Sky helped ensure faster diagnoses. When the outbreak was contained, the model also showed which mobile test labs were best to be decommissioned and which would be beneficial to remain in their locations should another outbreak occur.

The outbreak in Africa sent red flags to researchers who were concerned about an outbreak occurring in the United States. Using similar computing models running on Sandia HPC systems, the team studied a simulated outbreak and utilized Veteran's Affairs (VA) hospitals around the country in their simulation. They established which hospitals would best be used for triage, which to use for quarantine, and which to use for regular emergency care. The success of this study led the VA and CDC to re-evaluate crisis management plans for such a scenario. In the years since this HPC-based global health modeling capability was pioneered in Liberia and Sierra Leone, it has been adapted to improve the efficiency of healthcare delivery and infectious disease control in many other developing countries in Africa, the Middle East, and Southeast Asia.

Speed, scalability, stability, cost-optimization, and security—key attributes required for the high performance computers and systems completing the complex scientific and research tasks demanded by Sandia's national security mission. But who creates the unique technical and functional environments that enable High Performance Computing (HPC) innovation, and ultimately, mission success? In this edition of Sandia's HPC Annual Report, we not only celebrate the groundbreaking research developed on these platforms, but also the teams that provide the hardware, operating systems (OS), software codes, and data transfer and storage capabilities that combine to enable HPC achievement.

The following pages highlight some of the innovative techniques and specialized algorithms Sandia infrastructure teams deploy to fully leverage the speed and capabilities of our HPC platforms: the Multiscale Science department that collaborates with open source coders from across the DOE complex to create and support today's petascale and tomorrow's exascale machines; and the Sandia Interdisciplinary Machine Learning Research (SIMLR) team that uses machine learning in a manner that provides an accelerated path to results for highly complex and time-consuming engineering problems.

And in other important HPC infrastructure news, Sandia will soon complete a new building which will host Astra, a new National Nuclear Security Administration/Advanced Simulation and Computing (NNSA/ASC) system that will be the world's most powerful Arm processor-based supercomputer and the fastest high performance computer at Sandia. The 15,000 square-foot building will also be Leadership Energy and Environmental Design (LEED) Gold-certified, with novel energy- and water-saving technologies.

Sandia continues to stand at the forefront of HPC thanks to the dedicated people who not only manage the complex systems of today, but imagine, design, and implement systems to handle the needs and capabilities of tomorrow. Without a doubt, HPC will play a critical role in national security and scientific discovery for the foreseeable future.

Mark Sellers
Associate Labs Director of Mission Assurance

Scan photo to unlock more information

John Noe
Manager, R&D Science and Engineering




BUILDING 725 EXPANSION

Scan photo above to unlock more information

In October 2017, Sandia broke ground for a new computing center dedicated to High Performance Computing (HPC). The east expansion of Building 725 was entirely conceived of, designed, and built in less than 18 months and is a certified Leadership Energy and Environmental Design (LEED) Gold design building, the first of its kind for a data center in the State of New Mexico. This 15,000 square-foot building, with novel energy and water-saving technologies, will house Astra, the first in a new generation of Advanced Architecture Prototype Systems to be deployed by NNSA and the first of many HPC systems in Building 725 East.

LEED is the most widely used green building rating system in the world and quantifies building performance in terms of energy and water use. Building 725 East boasts innovate low and renewable energy consumption using onsite water and solar farms. Because HPC systems require massive cooling capabilities, Building 725 East dissipates heat energy through a 90% liquid/10% air thermocouple cooling system that will save several million gallons of water per year. It shares a non-load bearing, movable wall with the original Building 725, for an additional 20,000 square foot expansion if or when needed. Natural lighting from north-facing windows will further reduce energy needs for interior lights, and the infrastructure is in place to boost electrical service as demand increases in future.

Sandia has long been considering options for expanded computer facility space, but has been constrained by rising costs and decreased real estate in Sandia's main tech area. Even as funding was approved to pursue new supercomputers as early as the year 2000, the Computer Center management began preparing ideas and alternatives for providing appropriate space for an as-yet-to-be-designed system. While initial estimates for renovating the Building 880 Computing Annex came close to $9 million, constructing Building 725 was less than $5 million in 2003, and the 2017-2018 expansion expansion was accomplished for under $10 million. Building 725 East is a Closed Area (vault-type room), which can support classified processing up to Secret Restricted Data.

Astra will be installed in Building 725 East later this year. The new addition reflects Sandia values by incorporating efficiency, the use of advanced technology, and a commitment to sustainable design principles.

Simon Hammond
R&D Science and Engineering, Computer Science

ASTRA

Microprocessors designed by Arm are ubiquitous in automobile electronics, cell phones, and other embedded applications, but until recently they have not provided the performance necessary to make them practical for High Performance Computing (HPC). Arm processors are designed to offer a reduced instruction set, thereby reducing implementation complexity and allowing for increased processing speeds.

Astra—one of the first supercomputers to use processors based on the Arm architecture in a large-scale HPC platform—will arrive at Sandia in late 2018. The system will be built from 2,592 servers (nodes), each with two Cavium Thunder X2 Arm processors. Each Thunder X2 processor will offer 28 cores and four-way hardware threading and will be more than 100 times faster than a modern cell phone.

NNSA announced that Astra, the first of a potential series of advanced architecture prototype platforms, will be deployed as part of its Vanguard program. Vanguard will evaluate the feasibility of emerging HPC architectures as production platforms to support NNSA's mission to maintain and enhance the safety, security, and effectiveness of the U.S. nuclear stockpile.

One of the important questions Astra will help answer is how well the peak performance of this architecture will translate into real performance for mission applications. Astra's theoretical peak will be more than 2.3 petaflops, equivalent to 2.3 quadrillion flops, or calculations, per second. While being the fastest machine is not one of the goals of Astra or of the Vanguard program in general, three of Astra's nodes will be roughly equivalent to the ASCI Red computer, Sandia's original teraflop class system.

The story behind Astra goes back several years, when NNSA asked the Tri-Labs (SNL, LLNL, and LANL) to investigate an alternative architecture for nuclear weapons modeling and simulations technology. Because of Sandia's history in the development of these kinds of supercomputing machines, Sandia took the lead to find the best solution to NNSA's request. Sandia offered one option that was quite different than the likes of Intel or IBM systems: an Arm-based architecture that was previously considered non-tenable because it was not considered to be robust enough to handle petascale-sized operations.

Sandia was instrumental in selecting the hardware of Astra as well as the software to run it. Astra's design includes 36 compute racks (9 scalable units, each comprised of 4 racks with 72 compute nodes per rack) that are cooled by a closed-coupled air-to-liquid system. The system software stack was specifically built by Sandia's dedicated software engineers and in part utilizes best-in-class Tri-Labs software to increase implementation efficiency. For the first year or so, Astra will be available to application teams from the NNSA Tri-Lab user community and other selected teams for unclassified computing and testing; the system will migrate to classified computing after this initial phase.

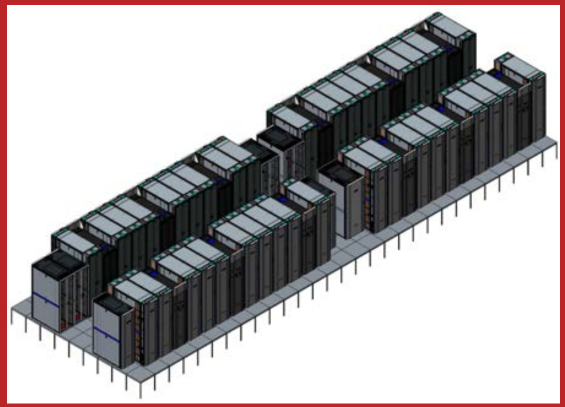
Astra will be installed at Sandia in an expanded part of Building 725 that originally housed the innovative Red Storm supercomputer. The Astra platform will be deployed in partnership with Westwind Computer Products Inc. and Hewlett-Packard Enterprise. Like Red Storm, Astra will require a very intimate collaboration between Sandia and commercial partners, in this case, Westwind, HP Enterprise, Arm, and Cavium, as well as with the wider HPC community to achieve a successful outcome.

NNSA VANGUARD PROGRAM

Astra will be deployed as the first of a potential series of advanced architecture prototype platforms as part of the Department of Energy's National Nuclear Security Administration (NNSA) Vanguard program. The Vanguard program was conceived to address inconsistencies and gaps in high-performance software and hardware development across the Tri-Labs (SNL, LLNL, and LANL).

After Astra is deployed, the Vanguard program will continue to evaluate the feasibility of emerging HPC architectures as production platforms to support NNSA's mission to maintain and enhance the safety, security, and effectiveness of the U.S. nuclear stockpile. Researchers at each of the NNSA Tri-Labs, as well as at HP Enterprise, select universities, and some vendors will have access to Astra during its first year of operation to evaluate its performance.

HPE Apollo 70 Cavium TX2 Node



Astra Data Center Layout

BEHIND THE SCENES OF HIGH PERFORMANCE COMPUTING

The computational power of High Performance Computing (HPC) is almost beyond our comprehension when we consider that 5 quadrillion computations can happen in a matter of seconds, or that machine learning is changing the way everything works. But none of that happens in a vacuum, and the teams behind the scenes—the developers of the hardware, operating systems, data transfer protocols, and the applications themselves—are the unsung heroes of a world where faster is better and you'd better hope there's no bug in the software or the hardware to slow you down. HPC is most successful when all these aspects work together seamlessly. The stories that follow are a tribute to the hardworking teams behind the scenes.

Ron Brightwell
Manager, R&D Science and Engineering

WHAT IS THE OPERATING SYSTEM FOR AN HPC?

Every computer requires an operating system (OS), from your desktop at home, your laptop at work, your cell phone, and even the computer in your car. But the OS of your PC is not the same as the OS of your cell phone. Likewise, the OS of HPC is considerably different than the OS of desktop computers. Ron Brightwell, manager of Scalable System Software at Sandia, leads a team of researchers who have helped shape the OS of Sandia's HPC centers.

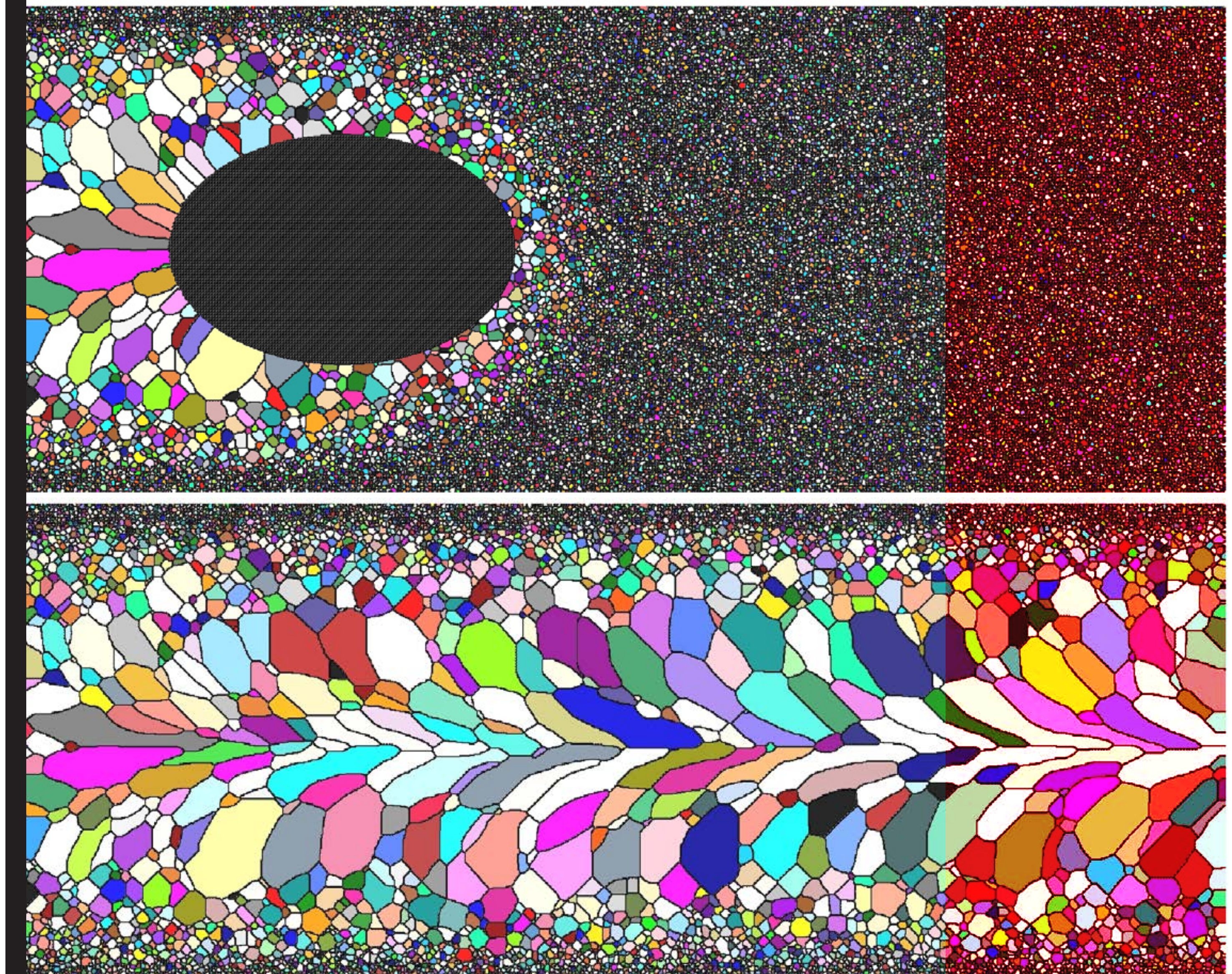
The Scalable System Software department traces its roots back to the early days of distributed-memory massively parallel processing (MPP) systems of the late 1980s. During this time, Sandia established the viability of MPP systems, such as the nCUBE-10 and the Intel Paragon, in solving mission-critical applications using modeling and simulation. The department grew out of the need to design, develop, and deploy more efficient system software focused on meeting the performance and scalability demands of these applications running on the largest and fastest computing systems in the world.

The group became firmly established in the early 1990s when researchers at Sandia partnered with the University of New Mexico to develop a customized system software environment based on a lightweight compute node OS designed specifically for large-scale, distributed-memory, message-passing machines. This initial lightweight kernel environment was successfully deployed on several large production systems at Sandia and eventually evolved into the OS that ran on the compute nodes of the world's first general-purpose parallel computer to achieve teraflops-speed: the Intel ASCI Red system.

System software research and development activities provide the software foundation that enables the scaling and performance of applications to unprecedented levels. Brightwell's team has performed pioneering work in lightweight OS and scalable runtime systems for some of the world's largest computing platforms. Core areas of competency include lightweight OSs, multi-threaded runtime systems, high-performance interconnect APIs, parallel I/O and file systems, and scalable system management infrastructure software.

It is widely accepted that future extreme-scale parallel computing systems will require alternative methods to enable applications to maintain current levels of uninterrupted execution. As the component count and speed of future systems continues to grow, the likelihood of a failure impacting an application grows as well. Researchers in the Scalable System Software department are exploring strategies to increase the resilience and reliability of future extreme-scale systems. As parallel computing architectures and applications have continued to evolve, the department has expanded into several other system software areas around operating systems, but has continued to focus on addressing the needs of extreme-scale systems and applications.

Steve Plimpton
R&D Science and Engineering, Computer Science



An early and final snapshot from an HPC simulation of an arc-weld spot (ellipse) passing over the surface of a 3D block of polycrystalline metal to create a weld between the upper and lower halves. The passage of the spot melts the metal, which then recrystallizes into large grains (colored). The shape of the grains and the strength of the weld depend on parameters like spot size/shape/velocity and material properties that can be varied in the simulation. This is work by John Mitchell and Veena Tikare (1400).

OPEN SOURCE SOFTWARE

Just like the software for your cell phone is different than the software used on your laptop or desktop machine, so too the software for HPC systems is different than other software. Fully leveraging the tremendous speed of an HPC platform requires exploiting parallelism at multiple levels. Thus, specialized algorithms and code are required to run HPC models and simulations efficiently.

Due to the growing complexity of HPC applications and the supporting system software, no one team handles it all. Software developers at Sandia join with developers across DOE and around the world to create and support open source software for today's petascale and tomorrow's exascale machines. At Sandia, open source codes are used to tackle large-scale modeling problems across nearly all its mission spaces, including national security challenges. The exponential growth in HPC computing power over the past three decades has fueled this impact. Consider that the first gigaflop machines (10^9 calculations per second) were available for use at Sandia around 1990. The first exaflop machines (10^{18} calculations per second) will be deployed by DOE in the early 2020s, with the power of one billion gigaflop machines.

Steve Plimpton, a computational scientist at Sandia's Center for Computing Research (CCR), has been working on open source HPC simulation codes for the past 15 years. "It's a highly collaborative process which produces software that is much more capable and robust than we could ever create ourselves," he says. "It's fun and challenging to work in an open source environment, where no one person is expert in all the aspects of a complex application, and users give instant feedback on code we release and are always asking us for new capabilities."

One example of a successful open source code that Sandia sponsors is the LAMMPS molecular dynamics simulator (http://lammps.sandia.gov) for materials modeling, which has grown over two decades to encompass hundreds of models for solid-state and soft matter. Depending on the resolution of the model used, length and time scales—from the atomic to the mesoscale to the continuum—can be simulated. Sandia researchers have used LAMMPS to model lipid membranes, entangled polymer melts, nanoparticle composites, energetic materials, granular powders, and even the seasonal agglomeration and melting of the Arctic ice pack.

Plimpton says, "We try to design HPC applications so that anyone inside or outside Sandia can easily contribute good ideas and code that can be quickly tested by large user communities. This includes finding and fixing bugs, adding models or features that enable new science, and creating ways to couple codes together for multiphysics or multiscale modeling."

NON-TRADITIONAL SUPERCOMUTING

The new age of supercomputing has arrived—graphics processing units (GPUs) provide a substantial advantage over traditional hardware. Today, supercomputers like Summit—the world's most powerful supercomputer at Oak Ridge National Laboratory—are still in high demand for extremely fast computing power. Supercomputing relies on CPU processing, with the eventual goal of transitioning HPC codes to leverage GPUs. By investing in GPU systems, Sandia is embracing this new age and enabling researchers to explore problems in energy, advanced materials, artificial intelligence (AI), and nuclear weapons with a new high-speed computing paradigm.

As part of this new computing paradigm, teams at Sandia are using these GPU systems to approximate traditional simulation models with machine learning. These trained machine learning models may be used in place of traditional supercomputing code modules for an accelerated time to solution. Here we highlight two teams pioneering non-traditional supercomputing at Sandia: a team that builds the hardware, and a team that researches and develops machine learning algorithms.


Craig Ulmer + Jerry Friesen
R&D Science and Engineering, Computer Systems

CARNAC FOR EMULYTICS™

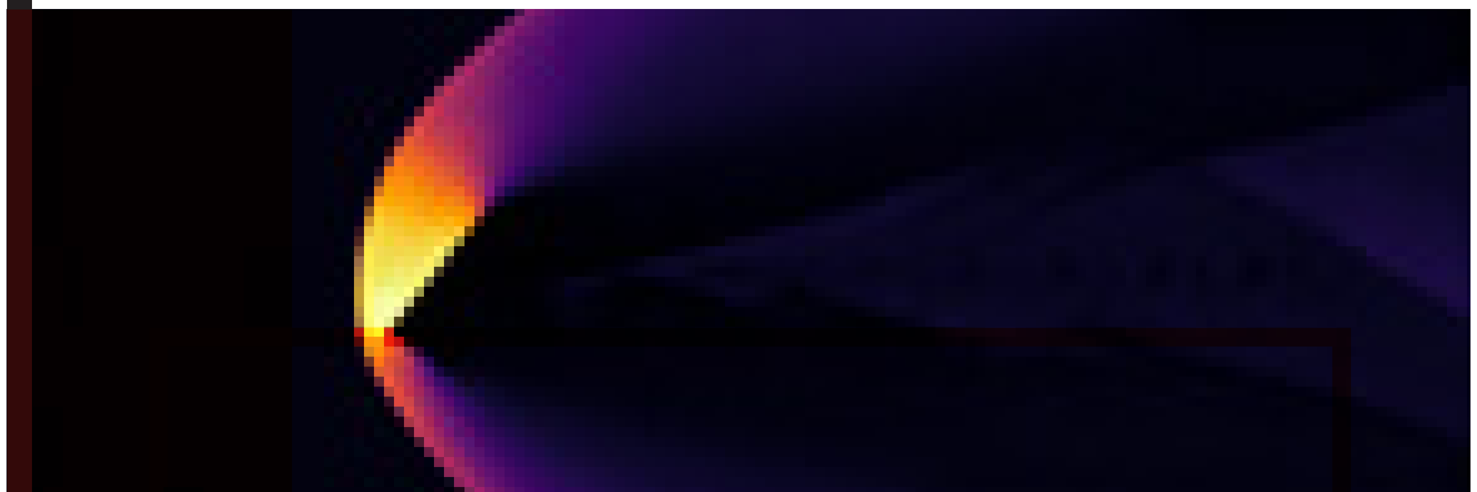
Carnac, located at Sandia's California site, is an institutional cluster for Emulytics that provides security researchers with resources to model enterprise computer networks and evaluate how resilient they are to attacks. While multiple Emulytics cluster computers have been built at Sandia, Carnac is the first system that was developed as an institutional resource that can be shared among different groups with disparate requirements.

A key challenge in deploying Carnac was creating an environment where a user could reserve a number of bare-metal computers from the cluster and then install a custom operating system for performing an Emulytics experiment. After becoming frustrated with the shortcomings of existing solutions, the Scalable Modeling & Analysis team blended together Sandia software (igor) with an open source cluster management package (cobbler) to provide a platform that all the Sandia Emulytics teams could use.

A side benefit of this provisioning software is that it can also be used to adapt the Carnac platform to serve other, non-Emulytics workloads during idle times. Because it's not uncommon for the job queues on other clusters to get swamped a week before a major conference or milestone deadline, it's extremely useful to be able to borrow idle nodes from Carnac and run a High Performance Computing (HPC) software stack for a short period of time to help users get a few extra cycles in before their deadline.

The next goal for the team is to architect a solution that allows Emulytics users to also take advantage of high-performance data analytics (HPDA). Prior to the acquisition of Carnac, the team deployed the Kahuna cluster to support HPDA research. This platform provides approximately 100 TB of distributed NVMe storage and 1.5 PB of traditional storage, and enables researchers to warehouse and analyze large datasets. Emulytics experiments generate large amounts of data that are being thrown away because teams don't have ways to store and analyze them. Combining Carnac and Kahuna provides a way to analyze data in greater detail, improving Sandia's ability to understand the results of experiments.

Charlie Snider, Carianne Martinez,
Kevin Potter, Matthew Smith, Emily Donahue
R&D Science and Engineering,
Computer Science & Information Technology



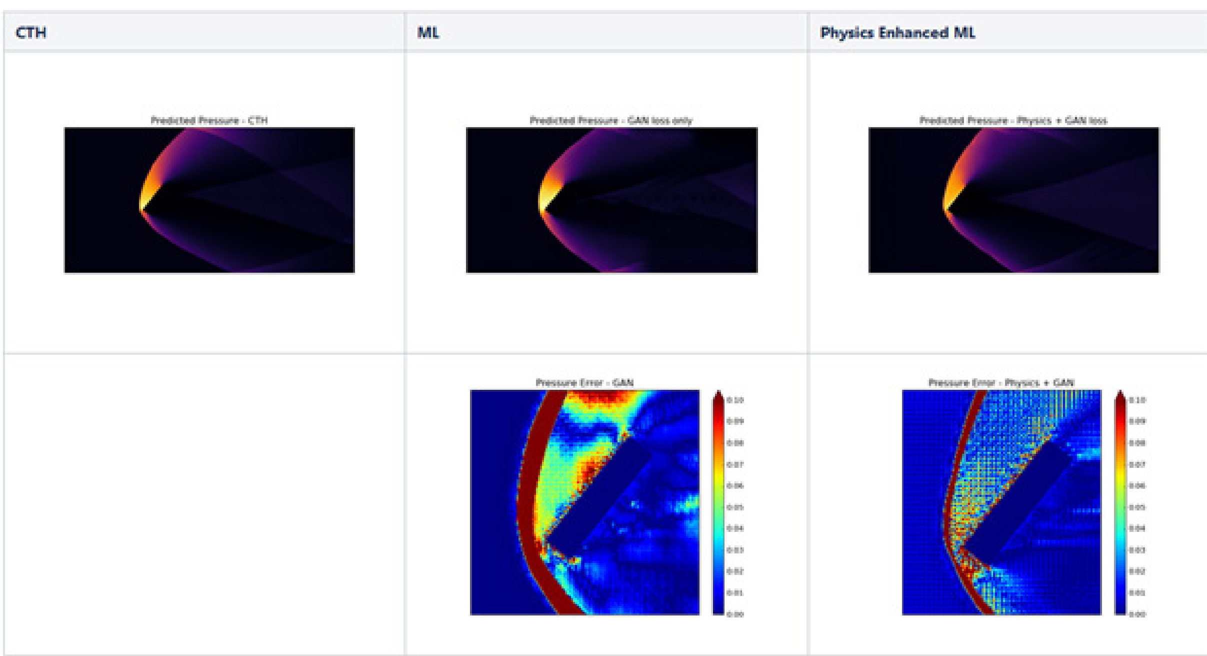
MACHINE LEARNING FOR CODE ACCELERATION

Due to recent advances, such as the onset of graphics processing units (GPUs), deep learning (a subset of machine learning) has become a viable technique for solving complex problems in many fields. Inspired by the human brain, deep learning mimics neural systems with computers that "learn" to perform a task without being specifically programmed for that task. In healthcare, for example, radiologists are trained to look for signs of lung cancer in patients and generally review hundreds of CT scans per day, which can often lead to mistakes in diagnoses. Alternatively, trained machine learning models can be utilized to review those same scans and identify discrepancies in lung tissue, without being specifically programmed to identify lung cancer.

At Sandia, machine learning is being put to the test on numerous multimission projects across the Labs. The Sandia Interdisciplinary Machine Learning Research (SIMLR) team leverages machine learning to increase the performance of highly complex and time-consuming engineering solutions. Starting with traditional supercomputing codes, the SIMLR team creates algorithms that learn models to approximate these codes quickly on GPU systems.

In traditional HPC computing, code is required to take the inputs and calculate the outputs. But given a large distribution of input and output pairs, the SIMLR team is able to use machine learning to predict output from new input without using the original code, and in doing so, accelerate the path to solution.

Sandia Laboratories Director Dr. Stephen Younger, in a 2018 All-Hands meeting titled "Creating Our Future," listed his top five "Big Ideas" to shape our future, including "transforming our use of supercomputers to leapfrog exascale technology." To this goal of shaping the future, the SIMLR team is augmenting many of Sandia's missions with machine learning to produce results in milliseconds rather than minutes.



Comparison of predicted pressure on a held out (unseen test not provided in the training set) rigid body rectangle traveling at Mach 5 from right to left at a single instance in time. The top row, from left to right, is the CTH result (one of Sandia's physics code bases), a machine learning (ML) result, and an improved ML result with a physics-driven loss function. Along the bottom row is the relative error comparing the ML result in the row directly above with the CTH result. CTH produces this output in 2.5 minutes; the ML inference time is less than a second.

Chris Saunders
Information Systems Architect

THE HARDWARE OF SMALLER CLUSTERS

Chris Saunders and the three technologists working with him are in high demand from Sandia's deep learning teams, and they're kept busy by building new clusters of computer nodes for researchers who need the power of supercomputing on a smaller scale. Sandia researchers working on Laboratory Directed Research & Development (LDRD) projects, or innovative ideas for solutions on short timeframes, formulate new ideas on old themes and frequently rely on smaller cluster machines to help solve problems before introducing their code to larger HPC resources. These research teams need an agile hardware and software environment where nascent ideas can be tested and cultivated on a smaller scale.

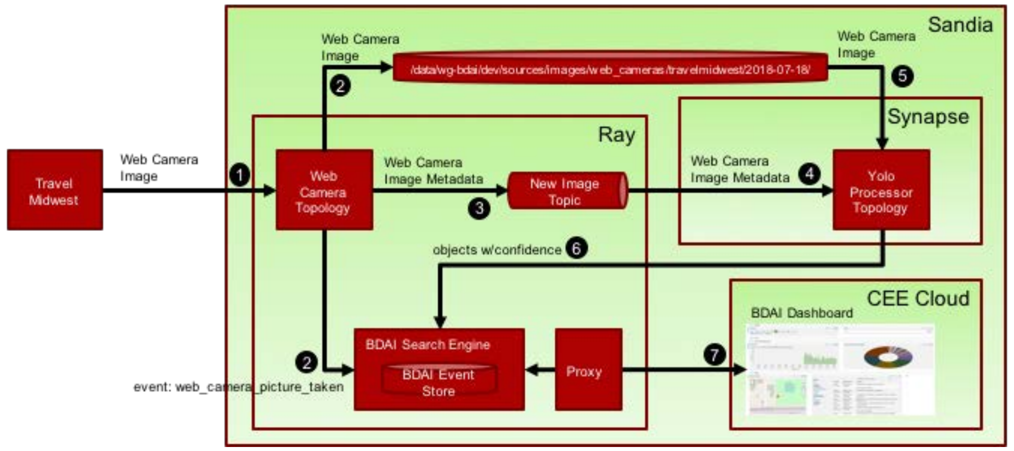
Saunders and his team at Sandia's Science and Engineering Computing Environments are successfully enabling this research by creating pipelines for emerging code—from the Cloud, to containers, to virtual machines—that build the right environment quickly to help teams solve their problems in a matter of days rather than months. While the larger HPC sources are available, it's these smaller clusters that can rapidly build a foundation for teams to build on for later development on larger systems.

YOU ONLY LOOK ONCE

As an example, Sandia's 6300 Capability Stewardship is a team that uses the smaller Sandia Ray cluster for its Big Data for Actionable Intelligence (BDAI) project. The goal of BDAI is to provide a shared understanding of all data sources collected, which includes tweets, web-camera images, MapQuest incidents, and construction-moratorium data. These wildly different data sources are normalized into a common frame of reference so that meaningful comparisons of events in space, time, and context are possible. Sandia's Ray cluster allows a distributed streaming processing platform to efficiently consume, process, and analyze this geospatial data. The result is a quick analysis—literally in seconds—of the raw data collected.

The Capability Stewardship team can rapidly build connectivity pipelines as a service to their customers, creating the infrastructure, the platform, and the software the customer needs at an accelerated pace, without the cost of running the larger HPC supercomputers. At least as far as deep-learning teams go, smaller is sometimes better.

BDAI WEB CAMERA OBJECT DETECTION



Susie McRee
Software Systems Engineer

DATA TRANSFER TOOL

Scan article title to unlock more information

Sandia has been developing and supporting data transfer tools for over 20 years and has the expertise to take DOE into the extreme scale era. In looking at exascale and beyond (Extreme Scale Computing), data sets can be thousands of TB in size, a single file can be in the 100 TB range, and billions of files are expected. Huge bursts of data need to be transferred, even today.

While data archiving is not often thought about, it is an integral part of the full data management path when data is generated on HPC systems. In order to move generated data to its final resting place (data archive) or to transfer between file systems, a capable data transfer tool is required.

The Sandia Data Management Team provides users with five data transfer tools to meet their needs. Currently, none of the five available tools is appropriate to handle file transfers for exascale and beyond. Some tools are well over 20 years old and are very fragile, meaning that changing code in one place can easily break code elsewhere. Some code is also very complex. Sandia's goal is to replace the five data transfer tools with a new tool currently in development: the Friendly Extensible Transfer Tool, or FrETT.

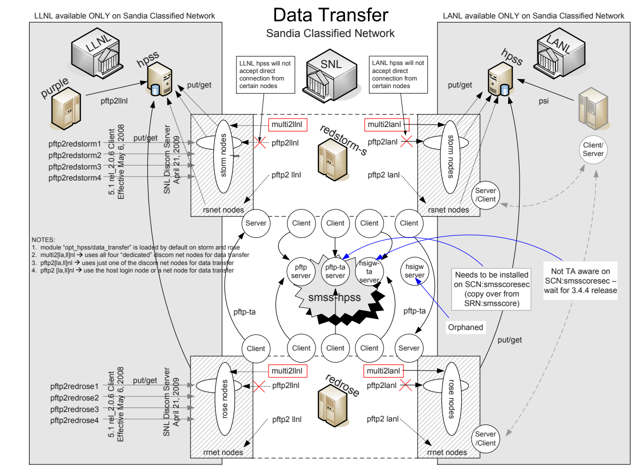
FrETT is a cross-platform (Linux, Windows, Mac) data transfer tool and will be deployed with a Graphical User Interface (GUI), Command Line Interface (CLI), and web application. It is end-point agnostic—allowing for flexibility—and modular, which means any site can create their own module to be used with FrETT. If other sites are interested in that module, it can be incorporated back into a FrETT release.

Sandia worked with Los Alamos National Laboratory (LANL) and Lawrence Livermore National Laboratory (LLNL) to deploy the same data transfer tools at each site so users would be familiar with the same tools. The result is data transfer among the Tri-labs that has been significantly simplified. The Data Management Team's goal is to meet the DOE exascale requirements for file transfer and to reduce the number of data transfer tools used between the Tri-labs from three to one.

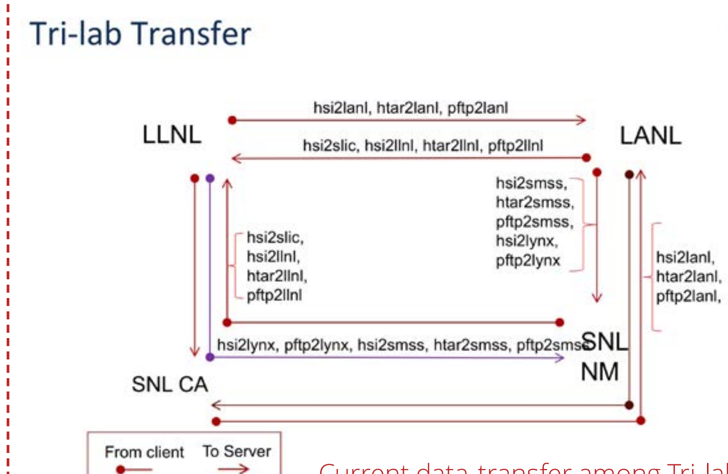
Sandia gathered requirements from LLNL, LANL, the National Energy Research Scientific Computing Center (NERSC), Oak Ridge National Laboratory (ORNL), and Argonne National Laboratory (ANL), which cover nuclear weapons and the Office of Science. ORNL is currently testing the new tool. FrETT, including the CLI and GUI, is planned for release in late 2018 with limited availability within Sandia. After thorough vetting, FrETT will be released with general availability in 2019 and be deployed to the DOE in late 2019 or early 2020. After release, Sandia will focus on creating a FrETT web application.

Sandia recognizes the need for a data transfer tool for extreme scale processing for the DOE complex. Collaboration with other DOE labs will ensure that DOE requirements are met, and FrETT will be available to other government agencies and may be released outside government agencies in the future. Sandia is also currently working with other DOE labs on extreme-scale storage and metadata-tagging (system and user) solutions.

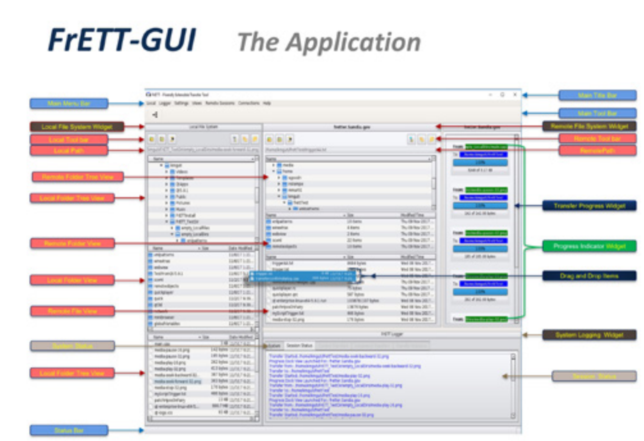
DATA TRANSFER EVOLUTION



Data transfer between Tri-labs was convoluted and complex.



Current data-transfer among Tri-labs is simpler, but still uses five transfer tools.



Future FrETT GUI. Simple, easy to use, one stop for all.

CREDITS

John Noe
Manager, HPC Development

Yasmin Dennig
Editor

Laura Perlichek and Whitney Lacy
Contributing Writers

Charles Vane and Meg Davidson
Contributing Editors

Stacey Long Reynolds
Designer

Regina Valenzuela
Video

Daniel Gomez
SNL Sim Magic

Sandia National Laboratories

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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. SAND2018-10962 R