Greetings once again from our nation’s capital. It has been six months since the last time I updated you, and while there has been a lot of churn, it sometimes seems there is very little forward progress. Things are beginning to break loose. The Trinity procurement contract was signed on 9 July and publicly announced on the 10th. Following on the successes with Roadrunner and Sequoia, Trinity will be an on ramp to future computing architectures.

There has in fact been a lot of progress in the ASC Program that will become more evident shortly. First, the new NNSA Administrator, Gen. Frank Klotz, has been confirmed by the
Senate, and is now at the helm. The Program has already had one opportunity to brief him
during his first week on the job, and expect to do so again in the near future. In addition to
Gen. Klotz, Chris Deeney’s position as Assistant Deputy Administrator for Research,
Development, Test, and Evaluation (NA-11) has been filled. Dr. Kathleen Alexander, who
previously headed the Work for Others program at DP and is a former staff member at both
Oak Ridge and Los Alamos National Laboratory started this past month. Dr. Alexander has
already had an opportunity to meet with the ASC Execs at a recent meeting in Washington.
ASC’s senior management structure is now in place for the rest of this administration.

On the budget front, ASC is continuing to recover from the very large sequestration
reduction in FY13. The current year has restored about half of the hit we took, and if the
President’s request for FY15 holds, ASC will be approximately back to where it was before
sequestration. So far things look promising. Part of the reason for optimism is the
continued interest in exascale computing and the reception our new ASC subprogram, the
Advanced Technology Development and Mitigation (ATDM) program has gotten from the
Congressional staffs and others. As we wait to see the final outcome from Congress on our
FY15 budget, we are well into the preparation of the FY16 President’s Budget. This will
occupy Headquarters the rest of the summer and into the fall.

ATDM grew out of the earlier Technology Mitigation Initiative (TMI), which many of you
worked on. TMI was intended to partially address the impact of changing HPC
architectures on our ability to perform our Stockpile Stewardship Mission in the absence of
a national Exascale Initiative. The ATDM subprogram is composed of two sub-elements,
one primarily focused on the development of new integrated design codes that will run
effectively on the evolving HPC architectures, and the other sub-element focused on the
hardware and environments that will be introduced over the next eight years. ATDM, in the
absence of a formal Exascale Initiative will remain a relatively small subprogram, roughly
comparable to PEM or V&V. The bulk of the Stockpile Stewardship mission will continue
as always. The current integrated design codes will continue to be adapted and extended by
the IC/PEM/V&V subprograms, and CSSE and FOUS will continue at roughly their
current level of effort.

With respect to Exascale, as I noted last time, Congress has directed NNSA to develop and
carry out a plan to incorporate exascale computing in the Stockpile Stewardship Program.
This is not an exascale program, but recognition that stockpile stewardship requires
exascale computing to be successful, and that it must be explicitly planned for.

The Secretary of Energy Advisory Board (SEAB), which I mentioned in my last Meisner
Minute, has completed review of the Department’s and the nation’s needs for exascale
computing and has developed a draft report. Secretary Moniz will use this report, as well as
many other inputs, in deciding if and how vigorously DOE will pursue exascale computing.
In the meantime, we will explore new numerical methods and code designs as we influence
technology developments through the ATDM subprogram.

Finally, Dan Orlikowski has returned to LLNL after spending a year on detail here at
Headquarters. We thank Dan for his hard work, and real contributions to the program. Two
new detailees have come on-board to assist the Headquarters staff, both from Sandia. Eric Strack will be primarily assisting the Defense Applications and Modeling program (IC, PEM, and V&V) and Sue Kelly will be primarily assisting Thuc Hoang on the CSSE and FOUS subprograms. Welcome Eric and Sue.

Trinity—NNSA ASC’s First Advanced Technology System

On July 10, 2014, the Department of Energy’s (DOE) National Nuclear Security Administration (NNSA) and Cray, Inc. announced a contract agreement for a next generation supercomputer, called Trinity, to advance the mission for the stockpile stewardship.

Managed by NNSA, Trinity is a joint effort of the New Mexico Alliance for Computing at Extreme Scale (ACES) at Los Alamos and Sandia national laboratories as part of the NNSA Advanced Simulation and Computing Program (ASC). Trinity will be used by Los Alamos, Sandia, and Lawrence Livermore national laboratories to run the largest and most demanding simulations to improve the understanding of predictive capability in stockpile stewardship, assuring the safety, security, and effectiveness of the U.S. nuclear deterrent.

Scheduled for delivery starting in mid-2015, Trinity will deliver at least 8 times greater application performance than Cielo, the current NNSA supercomputer sited at Los Alamos. Trinity is the first Advanced Technology System for the NNSA ASC Program and starts to implement the new NNSA ASC Computing Strategy for ASC Platforms, which is based on meeting mission needs and to help prepare the ASC Program for future architectural designs.

The Trinity supercomputer is designed to provide increased computational capability for the NNSA Nuclear Security Enterprise in support of ever-demanding workloads, e.g., increasing geometric and physics fidelities while maintaining expectations for total time to solution. Application performance and increases in geometric and physics fidelities are key drivers for Trinity.

The Trinity supercomputer is being provided by Cray, Inc. and is based on the Cray XC30 system architecture. Trinity will be a single system that contains both Intel “Haswell” processors and Intel’s next generation Xeon Phi processors (Knight’s Landing) to deliver over 42 peak Petaflop/s performance and will have an aggregate memory capacity of 2.11 Petabytes. The Haswell partition provides a natural transition path for many of the legacy codes running on the Cielo supercomputer.

The addition of the Knights Landing (KNL) processors, delivered in FY16, results in a significantly larger system than the current NNSA ASC platforms and provides application developers the opportunity to start the transition of ASC applications to next generation
supercomputing technologies and to run those applications at large scale. In order to effectively use the KNL processor to its full potential, the ASC code teams must expose higher levels of thread- and vector-level parallelism than has been necessary for the traditional multicore architectures. To help facilitate this transition, the Trinity Center of Excellence was established, with staff from the tri-Labs, Cray, and Intel.

Trinity introduces tightly integrated nonvolatile “burst buffer” storage capabilities. Embedded within the high-speed fabric are nodes with attached solid-state disk drives. The burst buffer capability will allow for accelerated checkpoint/restart performance and relieve much of the pressure normally loaded on the back-end storage arrays. In addition, the burst buffer will support novel new workload management strategies such as in-situ analysis, which opens a whole new space in which projects can manage their overall workflows.

Trinity also introduces advanced power management functionality that allows monitoring and control of power consumption at the system, application, and components levels.

The Trinity technical specifications and the request for proposals were developed as part of a joint effort between ACES and the National Energy Research Scientific Computing Center (NERSC), managed by the DOE Office of Science. This is the first joint effort between NNSA and the Office of Science on defining requirements for this scale of systems.

Trinity is named after the first nuclear weapons test conducted in New Mexico in 1945.

*Return to top*
**A New Take on Simulation Uncertainty**

The Quantification Methods (QM) project, part of the ASC Verification and Validation program element at LANL has driven the need for a new methodology for determining upper bounds on the uncertainties in simulations of engineered systems due to limited fidelity in the composite continuum-level physics models needed to simulate the systems. Current methodologies—parametric and model selection—cannot quantify this uncertainty. In particular, it has been shown that some models have no (or almost no) parameters to vary. Moreover, parameters in a physics model are generally provided to move between like processes (e.g., copper and aluminum) and model mean behavior. Therefore, parameters in physics models have little, if anything, to do with the generic approximations for a given physics model. It has been shown that generally no parameters are provided in regimes where there are no data to use to move between like processes. Finally, it has been shown that combining model selection and parametric methodologies lead to disjoint domains and, like parametric methodologies, provide, at best, a lower bound on the uncertainty. Nineteen physics models from five physics categories (plasma fusion, material damage, neutronics, material strength, and high explosives) have been used to illustrate these points.

An alternative method for determining uncertainties to address resource allocation questions has been proposed. The method, referred to as Physical Uncertainty Bounds (PUBs), obtains bounds on simulation uncertainties by first determining bounds on the physical quantities or processes relevant to system performance. Physics bounds are currently being constructed and use combinations of fundamental physics, experimental data, and numerical data. Some of the bounds currently being worked on are briefly discussed in the paper “Physical Uncertainty Bounds” [1], which will be released soon. A number of other papers containing much more detail on individual bounds are in progress.

In the words of the NRC committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification, *Uncertainty Quantification* should address the question: “How do the various sources of error and uncertainty feed into uncertainty in the model-based prediction of the quantities of interest?” [2]. By model-based prediction the committee is referring to a computational model that is used to simulate a real-world physical system (e.g., a bridge, protein folding, the climate), but they limit their focus to physics based and engineering models. Likewise, QM focuses only on computational models of engineered systems (e.g., a bridge, a flyer plate). The conclusions that have been drawn may pertain to other fields, but we make no such claim. In particular, QM addresses the uncertainty associated with a quantity of interest from a simulation of an engineered system due to using physics models to approximate physics processes in the simulation.

Therefore, the above quote from the committee on Mathematical Foundations of Verification, Validation, and Uncertainty Quantification [2] may be rephrased to more specifically define the *Uncertainty Quantification* question: How do the various sources of error and uncertainty due to using physics models to approximate physics processes feed into the uncertainty in the simulation-based prediction of the quantities of interest? Note
that the terminology “simulation-based” instead of “model-based” has been used. In the subject work, the word “model” means “physics model,” not “numerical simulation model.” There are two primary purposes for this research:

1. To determine resource allocation. In particular, for which physics processes and in what regimes would experimental evidence or better theoretical models reduce our uncertainty?

2. To quantify the uncertainty due to limited fidelity of physics models on our ability to predict experiments with no previous experiments to baseline against.

The reader with a background in uncertainty quantification might refer to this uncertainty as model-form uncertainty. Only a subset of model-form uncertainty as defined in [2] and [3] is addressed. In particular, in [2] the use of subgrid models that do not possess the full symmetry group of the Navier-Stokes equations is presented as an example of model-form error. In contrast, uncertainties associated with numerical methods are not considered in the subject work. The discussion is limited to physics models as defined above. Similarly, [3] includes numerical models (e.g., artificial viscosity) in the model-form uncertainty. Although, there are certainly valid arguments for including numerical models in a definition of physics models, that has not been included in the current effort. As stated in [2], “Methods for expressing model-form error, and assessing its impact on prediction uncertainty, are in their infancy compared to methods for addressing parametric methods.” It must be emphasized that this research addresses only a subset of the model-form error as described in [2] and [3].

QM proposes obtaining bounds on simulation uncertainties by first determining bounds on the physical quantities or processes relevant to system performance. These physical quantities/processes can be grouped into physics categories. It is these individual physical quantities and physics processes for which we have theoretical physics models, and these models collectively constitute a composite physics model for the given category. PUB consists of bounding these physics processes themselves as opposed to carrying out statistical analyses of the parameter sets of specific physics models (parametric methodologies) or simply switching out the available physics models (model selection methodologies). The difference between the proposed approach, which is based on bounding the physics processes, and parametric methodologies is mathematically defined in [1]. This approach constitutes a significant advance over traditional parametric methodologies that have been shown to provide at best lower bounds on the physics category uncertainties. The methodology advocated here has the potential to provide, for the first time, realistic bounds on the performance of engineered systems. The figure below illustrates the difference between traditional methodologies and PUBs for the Finite T Shear Modulus model Preston-Wallace [4]).
Uranium shear modulus versus temperature at $P = 0$; physics bounds in red. On the left, variations in the Preston-Wallace physics model parameter $\alpha$ are shown as a family of blue lines. On the right the uncertainty spanned with PUBs.

References


“Cramming” Sequoia Full of Jobs for Uncertainty Quantification

Just a few years ago, scaling simulation codes to run on thousands of cores was the pinnacle of supercomputing prowess. With the arrival of Sequoia, a 1.5 million core BlueGene/Q system, using a million cores is now commonplace. In April, over 50 application runs used the entire machine: 98,304 16-core nodes all “crunching” numbers in unison. Each parallel application run on Sequoia is called a job, and Sequoia was designed to run a small number of very large jobs. Over 900 jobs in April alone used 64,000 cores or more: more than the entire core count of LLNL’s next largest supercomputer, Zin. Sequoia users have discovered another interesting use for all those cores—running extremely large ensembles of small jobs as part of statistical studies in the area of uncertainty quantification (UQ).

Over the past decade, UQ has played an increasingly central role in NNSA’s stockpile stewardship program and in other national security missions. UQ techniques are used to
measure the sensitivity of simulations to input variability, thus helping scientists understand the likelihood of certain outcomes when some inputs are not exactly known.

Sequoia has more than enough compute power for UQ jobs, but large ensembles have stressed the machine in unanticipated ways. Specifically, Sequoia’s job management system was not designed to run more than 192 simultaneous jobs. The first attempts to run large UQ ensembles failed because the system resource manager could not scale to the job counts needed for UQ, but the Sequoia team at Livermore Computing (LC) rose to meet the challenge. Sequoia UQ users indicated that they wanted to run 10,000 jobs at the same time (while also confiding that a colleague was advocating running one million jobs). LC agreed to squash scaling bugs to push the limits as far as possible, and some of the early barriers were easy to overcome.

Linux, the operating system on Sequoia’s front-end nodes (FENs), has safeguards to prevent users from overwhelming the system by launching a large number of processes (colloquially called a “fork bomb”). Removing these safeguards allowed more than 192 jobs to be launched. At just over 512 jobs, the control system became so busy and unresponsive that another set of safeguards determined that the control system had hung and killed it before it could finish launching all of the jobs.

IBM quickly reproduced and fixed this scalability issue, allowing runs up to 3,000 simultaneous jobs before panicking the system. UQ users were then asked to try 1,500 simultaneous jobs—well below the tested limit. After two days of successful runs, all 1,500 jobs were assigned to a single batch-scheduling node, exhausting available memory. After the first batch-scheduling node failed, the batch system helpfully moved all of the jobs to the next working batch scheduling node, causing a cascade of crashes. Only after purchasing additional memory and installing 256 GB of memory in each batch-scheduling node were the UQ users able to push their jobs through. With the extra memory, the system would only panic after 3,000 jobs, and UQ users were able to push through more than 40,000 16-node jobs in April with this additional headroom.

One of the challenges of operating one of the world’s largest supercomputers is that the vendor cannot test the full-scale system in advance. Frequently, subtle bugs only manifest at scale, and LC staff must be capable of handling issues the vendor did not foresee. In this case, IBM could only test its system with up to 2,048 jobs. Working with IBM, LC staff determined the cause of the system panic. With this fix, users can now run over 20,000 simultaneous jobs on Sequoia.
Summary: Progression of Concurrent Job Thresholds

And what about that user who wanted to run one million jobs? Although the 10,000-job goal was surpassed, the system resource manager would need to be rewritten to scale further. Following a Sequoia status update, a possible solution surfaced. Simulations running on Sequoia use the Message Passing Interface (MPI) to send messages on the machine’s network. Using tools developed as part of Computer Scientist Todd Gamblin’s research on MPI performance measurement, Gamblin developed “Cram,” a tool that allows many small application instances to run inside of a single large job. Each instance runs a separate sub-problem with its own inputs and working directory. LLNL’s ARES code already implemented a similar capability at the application level, but Cram virtualizes the MPI interface and can be applied transparently to any application, without even modifying the source code.

Because Cram runs within a single MPI job, Cram can run one million jobs while only burdening the resource manager with the bookkeeping overhead of one. In May, LC was able to run 1.5 million simultaneous jobs on Sequoia using the tool—one job per core on the machine. Surprisingly, the main bottleneck was the creation of the cram file that describes to Cram the jobs that it should run. If done naively, creating the file requires 1.5 million command invocations on the front-end nodes, and this can take up to 54 hours. By adding a Python interface to Cram, the cram file could be created with a single Python script, and file creation time for a million-job run dropped to less than 4 minutes. Once that is done, it takes only 2 minutes to launch 1.5 million jobs from a Cram file on Sequoia. Users will now be able to use all the cores on Sequoia, no matter how many UQ jobs they need.

The team’s results were presented at ScicomP 2014.
**ParticlePack: Packing a Peck of Pickled Peppers**

A newly released version of LLNL’s ParticlePack shape-packing code includes features that have significantly enhanced the simulation of high explosives, propellants, and additive manufacturing for DOE, DoD, NASA, and their contractors.

ParticlePack is a preprocessing geometrical shape-packing code that generates objects for ALE3D, LLNL’s multiphysics, massively parallel, arbitrary Lagrangian–Eulerian simulation tool. ParticlePack is part of the ALE3D package that is distributed to government facilities and their contractors, and the code development is supported in part by the ASC Program.

ParticlePack was created to help define packs of similar objects (for example, grains, ball bearings, fuel pellets, manufacturing powders) for use in the ALE3D simulations. Today, the new version of the code is reducing problem setup time and improving the input accuracy for ALE3D.

“As scientists continue meshing very detailed, mesoscale objects, there are times they need to ‘inject’ material shapes into an existing background mesh,” explained Gary Friedman, developer of ParticlePack. “In ALE3D terminology, this is referred to as ‘shaping.’ This is where ParticlePack comes into play.”

ALE3D already has a full set of basic shaping capabilities, but ParticlePack adds the ability to define a set of particles that are tightly packed into a space and shaped into an ALE3D model. Although ParticlePack’s primary function is to shape objects into a background mesh, the code also provides an option to generate the particles as conformal mesh objects.

In addition, the code allows users to choose multifaceted grains, spheres, cylinders, or slabs for the particles and pack them into spherical, cylindrical, or rectangular enclosures. Users can define their particle size distribution using discrete size sets, Gaussian size distributions, or bimodal size distributions. Options are available to the user to set separations between particles. This is useful for modeling binder between high-explosive grains. Users can also specify material types based on particle volumes, particles with a coating layer, and particle defects. Particles can be randomly positioned in the pack or packed in a lattice pattern.

The new version of ParticlePack incorporates algorithms to generate shape files that take advantage of the massively parallel shape generation capabilities in ALE3D. “We now have the capability to generate and shape millions of particles in a couple of hours,” Friedman said.

Currently, ParticlePack is in the final stretch through a five-year plan to add enhancements that will be used in a wide range of high performance computing applications, benefitting DOE, DoD, and NASA. As part of the ALE3D modeling and simulation code, ParticlePack is helping scientists and engineers investigate the safety and performance of munitions and rocket motors, and pushing the limits of new manufacturing methods, such as 3D printing.
Scaling Studies for Simulation of Non-Equilibrium Flow using SPARTA on the Sequoia Platform

For the past two years, Sandia’s Computational Science and Engineering Science Centers have been collaborating on the development of SPARTA, a new three-dimensional scalable Direct Simulation Monte Carlo (DSMC) code. After this relatively short development time, SPARTA has been tested on Sequoia, Lawrence Livermore National Laboratory’s 98-thousand-node, 1.57-million-core super-computer. During a recent Dedicated Application Time (DAT), scaling and verification studies for SPARTA were performed on Sequoia, in addition to simulations of the Richtmyer-Meshkov (RM) gas instability. SPARTA performed extremely well on the scaling study. Figure 1 shows the strong (single curve) and weak (curve to curve) scaling results running problems of various sizes on one node up to the full 1.57 million cores. The largest calculation (not shown) used 3 trillion particles and 1 trillion grid cells. The data in the figure are for runs using 1 thread/core
(16 MPI tasks/node). Sequoia’s multi-threading capabilities were also successfully exercised by SPARTA with 2 and 4 threads/core, though memory/node becomes a limitation in this mode at the highest node counts.

![Figure 1. Scaling up to 96k nodes (times 16 cores).](image)

The Richtmyer-Meshkov (RM) gas instability simulation also ran successfully for 24 hours using all 1.57 million cores. This is the largest molecular-level, 3D-RM instability calculation ever performed with the DSMC method. The RM instability is a fundamental problem of fluid mechanics and is widely used for assessing the accuracy of continuum fluid dynamics codes. In Figure 2, the recent DSMC calculations of non-dimensional perturbation growth rates are shown to agree with the short-time linear theoretical result, and fall amid results from a variety of continuum codes in the non-linear, late-time regime, which typically include some, but not all relevant physics, such as non-equilibrium effects. DSMC includes non-equilibrium effects, which enables investigation of high Mach number and non-equilibrium mixing effects which continuum codes cannot accurately predict.
SPARTA will be an important tool for development of a predictive capability for hypersonic re-entry of a weapon and the resultant mechanical environments induced by turbulent reacting flow around the body. Key physical phenomena enabled by this capability include non-continuum flow at extreme altitudes and reacting gas dynamics associated with ablation. We also hope to leverage this DSMC capability for development of next-generation solution algorithms in a wide range of engineering simulation applications.

Pinch Weld Process Modeling to Improve GTS Reservoir Stem Weld Quality

At Sandia, we have been working to develop the capability to model the resistance pinch weld that is used to seal the tubing after pressurizing gas transfer system (GTS) reservoirs. In the pinch welding process, current is passed through electrodes that compress the stem tube with a prescribed force. The current resistively heats and softens the material, and as the tube is compressed, the inside diameter closes upon itself and forms a welded interface through a diffusion bond process. It has been experimentally observed that the quality of the weld depends on the initial properties of the stem tube. Our goal is to provide a computational tool that will enable optimization of the pinch weld process parameters to achieve a quality bond and eliminate cracks that have been observed in non-optimal processes. The optimal parameters, such as the applied current, number of current cycles, or applied force would vary based on initial tube geometries and material properties.
Three-way coupled electrical-thermal-mechanical simulations were performed in Sierra with inputs based on experimental measurements. A strain rate and temperature dependent, internal state variable constitutive model accurately captures the material response up to melt. The simulations exhibit experimentally observed features such as the lateral flow of material shown in the image below. Due to the large deformations in the process, it is necessary to re-mesh the 3D deformed geometry, map the material and displacement variables to the new mesh and restart the analysis to allow simulation of the full closure of the weld. After the new capability is fully validated using experimental data, it will then be used to propose modifications to the pinch weld process to improve bond quality.

Lateral flow of material in pinch weld process modeling: (left) cross-section image from an interrupted pinch weld for drawn tubing; (right) results of process modeling simulation.

Predictive Science Academic Alliance Program II

In June 2013 the National Nuclear Security Administration (NNSA) announced the selection of six academic centers of excellence whose primary focus will be on the emerging field of predictive science. The six universities were selected either as a Multidisciplinary Simulation Center (MSC) or as a Single-Discipline Center (SDC) and include

- University of Florida, Gainesville, Fla., “Center for Compressible Multiphase Turbulence,” an SDC
- University of Illinois-Urbana-Champaign, Champaign, Ill., “Center for Exascale Simulation of Plasma-Coupled Combustion,” an MSC
- University of Notre Dame, Notre Dame, Ind., “Center for Shock Wave-processing of Advanced Reactive Materials,” an SDC
- Stanford University, Stanford, Calif., “Predictive Simulations of Particle-laden Turbulence in a Radiation Environment,” an MSC
• Texas A&M University, College Station, Texas, “Center for Exascale Radiation Transport,” an SDC

• University of Utah, Salt Lake City, Utah, “The Uncertainty Quantification-Predictive Multidisciplinary Simulation Center for High Efficiency Electric Power Generation with Carbon Capture,” an MSC

The PSAAP II centers will develop the science and engineering models and software for their large-scale simulations utilizing methods of verification and validation and uncertainty quantification, with an additional focus on extreme-scale computing. The goals of these disciplines are to enable scientists to make precise statements about the degree of confidence they have in their simulation-based predictions and prepare to exploit new generations of computing technology.

Center for Compressible Multiphase Turbulence
[University of Florida]

The overarching goals of the Center are threefold: to radically advance the field of compressible multiphase turbulence (CMT) through rigorous first-principle multiscale modeling; to advance very large-scale predictive simulation science on present and near-future platforms; and to advance a co-design strategy that combines exascale emulation with a novel energy-constrained numerical approach. Petascale (working toward exascale) simulations of instabilities, turbulence and mixing in particle-laden flows under conditions of extreme pressure and temperature are being performed to investigate fundamental problems of interest to national technological leadership.

The chosen CMT demonstration problem for the initial phase of our research activities is an explosive dispersal of solid particles that consists of a detonator connected to a charge, which runs across the length of a cylinder filled with iron powder. The mid-section of the cylinder is occupied by a matrix enclosing a cloud of inert spherical glass particles. Simulating the evolution of the mid-section of the cylinder is the objective. See Figure 1 for preliminary simulations. The uncertainty budget will drive the overall development of models, software, simulation, and calibration/validation experiments.

Figure 2 shows a concept diagram of our Exascale emulation project within CMT. It is based on a coarse-grained simulation approach that is analogous to rapid virtual prototyping. Exascale emulation will provide an accurate first-order approximation for performing design-space exploration on extreme-scale, future-gen systems before (and complemented by) time-consuming detailed simulation/emulation. As shown

Figure 1. Experiment (top) and simulation (bottom) at early times of a cylindrical charge with inert particles.
in the figure, the project is divided into three major tasks. Task A is Behavioral Emulation Object (BEO) modeling, calibration, and validation of application kernels (App BEOs) and systems (architectural BEOs). A BEO is a model that is used as a surrogate of the real (existing) or notional (future) application or architecture. Task B is testbed experimentation and benchmarking used to calibrate and validate the developed models (BEOs). Task C is development of the behavioral emulation platforms. Currently, we are exploring the use and development of three platform types: discrete-event simulator, a symmetric multi-processing simulator on many-core computer, and hardware emulator on a reconfigurable computer.

![Figure 2. Concept diagram for Exascale emulation for future-gen systems.](image)

XPACC: The Center for Exascale Simulation of Plasma-Coupled Combustion [University of Illinois at Urbana-Champaign]

The University of Illinois is ramping up activities in its new ASC PSAAP II Center for Exascale Simulation of Plasma-Coupled Combustion (XPACC).

The predictive science goal of the center uses plasmas to fundamentally boost the performance and efficiency of turbulent combustion. Radicals produced in plasmas can accelerate burning by short-circuiting standard chemical pathways; electric fields affect flame stability by accelerating charged chemical species within thin flame fronts; and plasma-induced Joule heating affects both flow, via thermal expansion, and chemistry, via temperature. Efforts are centered on predicting the plasma-mediated sustained-ignition of a turbulent gaseous fuel jet in a cross flow. Low-dimensional, physics-targeted experiments will be used for selection and calibration of models in the full-system simulations.

The Center will employ an existing simulation code (*PlasComCM*) that was developed at the University of Illinois and has been used in a range of flow-physics applications. It employs locally structured, globally unstructured overset meshes, which can provide high-resolution discretizations without the mesh generation challenge of globally structured meshes. This code has been used extensively on present-day petascale systems for physics-detailed simulations of compressible turbulence phenomena, including jet noise and high-speed fluid-structure interactions, and has been shown to scale beyond 100K processor
cores. However, the incorporation of ever more physics-faithful combustion and plasma models, along with the multiple simulations needed for uncertainty quantification, will require efficient simulations on next-generating computing platforms.

PlasComCM will be used as an exemplar of modern extreme-scale applications. The Center will develop new programming approaches that can be applied to well-structured existing applications to achieve better performance and scalability. The techniques include data structure and code optimizations for memory efficiency and exploitation of high-throughput computing elements and automatic load balancing to address performance irregularities in compute systems, algorithms, and problems.

These and other tools and techniques will be designed to work on applications with similar needs, and the Center will actively engage the community to ensure the success of these performance-oriented programming systems. To facilitate this, XPACC will host a series of open Workshops on Exascale Software Technologies (WEST), held near NNSA labs in order maximize interactions with NNSA personnel. Workshop themes will be developed to track the evolving needs of extreme-scale computing for physical systems.

![Dielectric barrier discharge (DBD) plasma source, coaxial with fuel jet, (b) glow discharge, and (c) 400M mesh point detailed simulation for modeling flame stabilization by DBD.](image)

Center for Shock Wave-processing of Advanced Reactive Materials (C-SWARM) [University of Notre Dame]

The development of controlled microstructures is a primary goal in designing novel materials with unique properties. The main mission of the Center for Shock Wave-processing of Advanced Reactive Materials (C-SWARM) is to predict shock conditions under which new materials can be synthesized. This processing generates high temperatures and pressures that can lead to new materials being created. Such material transformations are governed by a plethora of physics, mechanics, and chemistry that tests our understanding of microstructure-property-relations and our capacity to tune materials at will. The goal of C-SWARM is deployment of verified and validated computational simulations to predict the properties and dynamics of this complex system, with quantified uncertainty.
Heterogeneous reactive materials (HRMs) are one of the best examples of a multiscale problem where Exascale computational resources can have a transformative impact. These materials acquire enhanced properties after shock wave processing due to the very large pressures and strain rates they experience. C-SWARM represents a coordinated effort combining multi-resolution simulations, a transformative computational execution model, and experiments to analyze the chemo-thermo-mechanical response of HRMs.

Polydisperse materials with different chemical composition and disparate thermo-mechanical behavior exhibit a complex response that is difficult to capture using conventional theories. To model such materials, C-SWARM employs a multiscale strategy that solves adaptively phase-averaged macro-continuum equations using an Eulerian algorithm with constitutive equations locally provided by well-resolved micro-continuum simulations of the multi-phase mixture using a Lagrangian algorithm. The macro- and micro-continuum codes will utilize an advanced execution model, ParalleX. ParalleX enables data mining of runtime information to dynamically adapt the codes’ demands to resource availability. Furthermore, Active System Libraries will be implemented to ensure that the macro- and micro-continuum algorithms can be represented in an effective and hardware-independent fashion.

The integrated V&V/UQ program provides a platform for computational model verification, validation and propagation of uncertainties. The emphasis of C-SWARM is on quantifying the predictive ability of the multiscale simulations in an efficient manner. A key component of the center is a series of carefully co-designed experiments and data-driven simulations (with quantified uncertainties) to enable meaningful and rigorous comparison of simulation predictions with experimental results.

The project is administered within the College of Engineering at the University of Notre Dame. Prof. S. Paolucci, is the Principal Investigator and Director. The research is a collaborative effort between the University of Notre Dame, Indiana University, and Purdue University.
Boulder, the University of Texas at Austin, and the State University of New York at Stony Brook.

The project, “Predictive Simulations of Particle-laden Turbulence in a Radiation Environment,” will investigate the effect of radiation on particle motion in a turbulent airflow. This is a poorly understood physical process that can open new opportunities for efficiency gains in solar thermal receivers with applications in energy conversion and chemical splitting of components in chemical plants.

The objective in such systems is to achieve high temperatures with minimal losses to the environment. Conventional solar-receivers collect focused sunlight primarily via a solid surface, which then conductively heats the target fluid. One main drawback is that achieving high mean temperature in the fluid requires local heating of the solid surface to even higher temperatures, which can result in significant radiation losses. Particle-based receivers present a potential remedy to this issue by allowing more uniform and volumetric absorption of radiation by the working fluid. Given most fluids are transparent to radiation, absorbing particles are needed to intercept high-energy solar rays and allow local transfer to the fluid mixture. However, the three-way coupled physics of particle transport, fluid dynamics, and radiation, presents additional engineering challenges, which are mostly unexplored. For example, fluid motion in such systems naturally involves turbulence, and while turbulence helps global mixing of mass and heat, it induces preferential concentration of particles: local turbulent vortices can centrifuge out particles to zones of local shear, and lead to heating non-uniformities and reduction of the overall efficiency. Additionally, temperature inhomogeneity leads to local fluid expansion altering the turbulence structures. Insights into design and optimization of such systems require careful investigations of the physical interactions between particle, radiation transport and fluid turbulence.

The Center will focus on simulations at an unprecedented level of fidelity by accessing the largest supercomputers in the U.S. This will help develop and ultimately demonstrate predictive science simulations on next-generation Exascale systems - computers that can perform a quintillion floating-point operations per second—expected to become available around 2020. The current trend in supercomputer systems is to increase the number of computing cores and as a consequence this leads to data movement across extensive networks. Moreover, diverse and specialized computing units (CPUs, GPUs and other accelerators) and multiple memory banks with different performance will coexist creating a truly complex hardware and network system. Achieving high performance requires new programming models that enable computational scientists to develop algorithms and
software tools without intimate knowledge of the underlying architecture details. A key feature of the Stanford PSAAP II Center is a strong partnership between computational and computer scientists. Domain Specific Languages (DSLs), such as the Stanford-developed Liszt, are the key ingredient to enable computer programs to identify and recover from faults while providing flexibility in data management and efficiency in mapping algorithms on the most appropriate computing units.

The project will also concentrate on uncertainty analysis, allowing researchers to quantify errors and uncertainties in the simulations and, therefore, to determine how much confidence can be placed in the results. A dedicated experimental campaign will be undertaken alongside the computational work to help understand and identify these uncertainties and to provide overall validation data to assess the predictive ability of the physical models and software tools developed within the project.

In addition to the research effort, new graduate-level classes on computational science, multiphase flows, radiation modeling and high-performance computing will be introduced at Stanford. Workshops on uncertainty analysis, multiphysics and exascale challenges will also be organized to expose the broader community to the research efforts at the Center.

The agreement continues a 15-year history of strong collaboration between NNSA laboratories and Stanford University, including the Advanced Simulation and Computing (ASC) and PSAAP programs. For more information on Stanford’s PSAAP II project, please visit the Center’s website http://exascale.stanford.edu or contact the Director, Gianluca Iaccarino (jops@stanford.edu).

---

Center for Exascale Radiation Transport (CERT)
[Texas A&M University]

The Center for Exascale Radiation Transport (CERT) was created through a research grant from the NNSA PSAAP II. CERT is focused on the development of computational techniques for efficiently simulating thermal radiation transport (propagation) using extreme-scale or exascale computers together with the development of predictive science techniques to quantify uncertainty in simulated results. CERT is led by Texas A&M University. The University of Colorado and Simon Fraser University are also participating in this effort.

Exascale computers planned for the future will consist of many millions of processors, and be capable of executing on the order of $10^{18}$ floating-point operations per second. The fastest computers currently in existence execute roughly $10^{16}$ floating-point operations per second and use enormous amounts of power. In order to achieve affordable operating costs, exascale computers must consume far less energy than current computers. Computing on exascale-scale machines will be very different from computing on existing machines because of this low-power requirement. The entire concept of the “cost” of computational
algorithms will change and algorithms must have the capability to detect erroneous computation and either correct it or tolerate it in some quantifiable manner.

An unusual aspect of the CERT predictive science effort is the use of neutron experiments as a surrogate for thermal radiation experiments. The use of neutrons will actually enable us to improve the predictive science associated with our simulations far beyond that which could be achieved with thermal radiation experiments. CERT will develop exascale computer science algorithms, exascale adaptive transport algorithms for both improved accuracy and numerical error estimation, multiscale physics models relating to transport in embedded voids and small cracks, and exascale multilevel preconditioning techniques. Computer science research will include the development of methods for fault tolerance and the development of performance models that include the impact of iterative methods on parallel efficiency. Below is a graph comparing the weak scaling performance of our latest transport solution algorithm with the predictions of the associated performance model. It can be seen that our algorithm scales with roughly 60% efficiency from 1 to 384,000 cores of the Sequoia machine at LLNL, and that the performance model agrees very well with the actual performance.

![Parallel Efficiency vs. Core Count](image)

The Carbon-Capture Multidisciplinary Simulation Center (CCMSC) [The University of Utah]

CCMSC has the mission of using exascale UQ-predictive simulation science to rapidly design and deploy a new technology for secure electric energy generation; namely, a high efficiency advanced ultra-supercritical oxy-coal power boiler. This overarching problem integrates a group of multidisciplinary scientists and engineers from The University of Utah; the University of California, Berkeley; and Brigham Young University. The CCMSC is organized into three teams: the Exascale Team, the Predictive Science/Physics Team, and the Validation/Uncertainty Quantification (V/UQ) Team.
We use hierarchical validation to obtain simultaneous consistency between a set of selected experiments at different scales embodying the key physics components (large eddy simulations, multiphase flow, particle combustion and radiation) of the overarching problem. We extrapolate the uncertainty obtained from the V/UQ of the sub-scale, sub-physics analysis to a prediction of the full-scale boiler that is simultaneously consistent with all of the experiments and with all of the validation metrics of our validation hierarchy. CCMSC starts with an existing proven computational platform (UintahX) and sequentially moves to multi-petaflop and eventually exascale computing. We will accomplish this transformation with three software infrastructure components: (1) the exascale runtime system, (2) TASC, a Transparent Abstractions for Scalable Computing, representing a high-level, portable “assembly language” for scientific computation with transparent abstraction by using a sub-Turing, embedded domain-specific language, and (3) the data management and visualization infrastructure for dealing with large data and for connecting that data to the visualization and data analysis components.

The expected impact is a demonstration of using exascale computing with V/UQ to more rapidly deploy a new technology for providing low cost, low emission electric power generation. To this end the CCMSC has established a collaborative agreement with Alstom Power to jointly use this exascale technology to contribute to the design of a new 350-MWe oxy-coal boiler. The exercise of developing the tools for this overarching problem will produce: (1) exascale computing software that will be regularly released through open-source licensing, (2) tools for V&V/UQ for use with other large applications with expensive function evaluations and sparse/expensive experimental data, and (3) new advances in computational fluid dynamics, multiphase reacting flow and radiative heat transfer. We also anticipate an impact on the educational process of a new generation of students educated in exascale predictive science.

A volume rendered image of a CCMSC prediction of an Alstom oxycoal boiler. The image shows local concentrations of the 90-micron coal particles in the boiler. The particles concentrate in local regions due to variation in local Stokes number. The simulation was performed on Titan using 1 million core hours.
ASC Salutes Bob Ballance

Bob Ballance’s background and interests place him in the intersection of system design, system administration, operations, user support, vendor relations, and application development—a busy locale where the traffic is heavy, collisions are frequent, and life is interesting. Significant system-shaping projects, like bringing up Red Storm, managing Cielo as part of the SNL/LANL ACES alliance, or coalescing Sandia’s NW and Institutional Computing strategies around the “Mission Computing Council” have all benefited from Bob’s participation. But it is his attention to detail and to how teams think, like defining categories for on-going data transfer testing so that various Tri-Labs sub teams could move forward as a single cohesive team, that best captures his contributions.

Since joining Sandia in 2003, Bob’s most significant contributions to the ASC program have resulted from nurturing close collaborations among the three labs: the Tri-Lab resource management effort, Tri-Lab shared system management via the Enhanced Priority Request (EPR) team, management of Tri-Lab system resources (Red Storm, Cielo, Trinity), work supporting the ACES partnership with Los Alamos, and interactions with users from across the Labs. “My job is mostly linguistics,” he once replied to a manager. “It’s about getting the team to agree to a common language so that we can move forward together.” Bob’s excellent work was recognized through his promotion to Distinguished Member of Technical Staff in 2010.
As a practicing computer scientist he has served multiple tours of duty—graduate student and professor, researcher and developer, entrepreneur and consultant, manager of high-tech teams—all efforts that have helped to prepare his assignments at Sandia. Since receiving his Masters Degree from the University of Michigan in 1978, he has published research in programming environment design, programming language syntax and semantics, object-oriented software engineering, and system monitoring. He received a Ph.D. in Computer Science at UC Berkeley in 1989. As a graduate student, he worked at HP Labs where he wrote the device driver for HP’s original prototype flatbed scanner and helped to design and write the first compiler for HP’s RISC-based computer architecture. The latter assignment melded compiler design with computer architecture, which served him well when he served as Acting Manager for Sandia’s Scalable System Architectures department where he contributed to the ASC Advanced Architectures Test-bed program.

Bob’s work often leads back to the community: in the early 2000’s he helped to establish the Linux Clusters Institute (LCI), an organization centered on helping organizations to start using and operating Linux clusters in education, research, and business. Later, as a member of the core team for Red Storm, he developed close working relationships both within Cray and throughout the Cray system user community where he helped to found the Cray XTreme Systems Group, now a part of the Cray Users Group. In 2014, he was recently elected Secretary of the Cray User’s Group Board of Directors. He also regularly serves on the Technical Program Committee for of the International Conference for High Performance Computing, Networking, Storage and Analysis, commonly called SuperComputing.

At Sandia, Bob is currently engaged in defining the Usage Model and organizing the operations teams for the ACES Trinity System, slated to be deployed at Los Alamos in 2015, and advising the Commodity Technology Systems procurement (CTS-1). Still exploring, he’s actively involved in applying data-intensive computing at Sandia and in automating the process of system reporting for the Tri-Labs. His software packages have been adopted by the Tri-Labs for usage reporting on Sequoia and Cielo.

At home, Bob is an avid baker (sourdough), cook (Szechuan, Spanish, Italian), hiker, rock-climber, and juggler (balls, clubs, and very occasionally lit torches), and he just celebrated 35 years of marriage to his wife, Kathy.

Recent Publications

Los Alamos National Laboratory


**Sandia National Laboratories**


LA-UR-14-26576