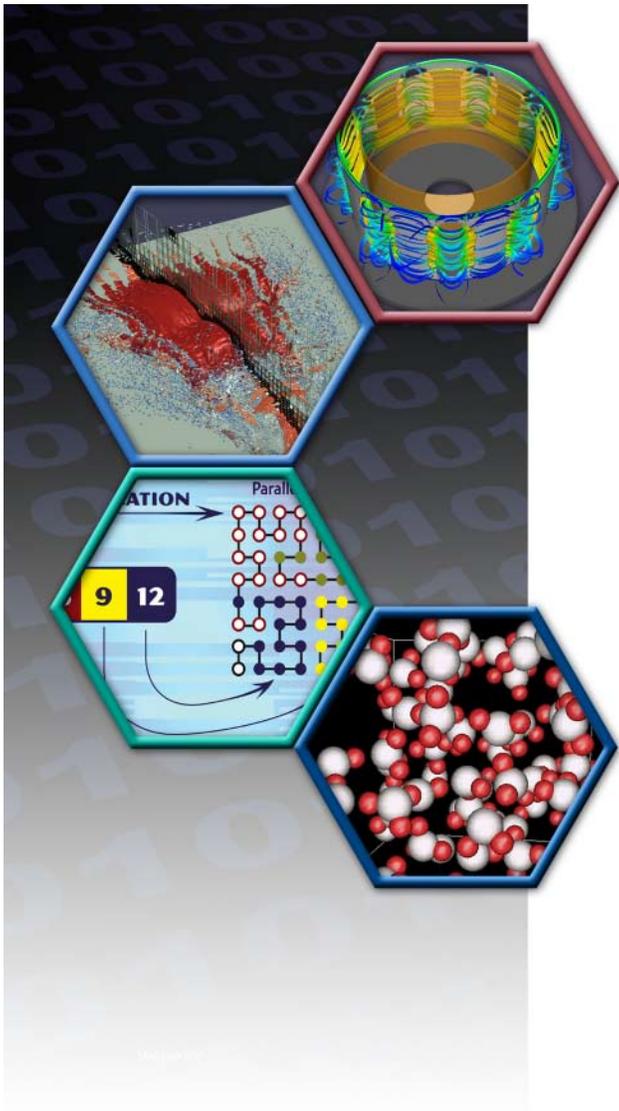


COMPUTATIONAL  
SCIENCE RESEARCH  
**HIGHLIGHTS**





## ON THE COVER:

Shown top-to-bottom:

1. 3D magnetic field generated by electrical current running through an idealized plasma liner and returning through a slotted current return as commonly found in the Z-Machine. The presence of slots perturbs the magnetic field topology, impacting the physics occurring inside. High-resolution versions of this simulation require 60 seconds to solve on 50 processors. Prior to the development of the new multigrid solver in ML, this simulation was not feasible due to convergence failures of the more standard linear solvers.
2. Simulation of a high velocity metal ball striking a metal brick.
3. Allocation Algorithm with Hilbert Space-Filling Curve and Span Minimization.
4. Obtaining an accurate description of water dynamics requires very long (10s of ps) molecular dynamics simulations on large supercells. This requires an XC functional that is both fast and accurate. The picture is a snapshot of a DFT-MD simulation using AM05 for 64 water molecules.



SAND 2009-2614P

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

---

# COMPUTATIONAL SCIENCE RESEARCH HIGHLIGHTS

JAMES S. PEERY  
DIRECTOR  
COMPUTATION, COMPUTERS,  
INFORMATION AND MATHEMATICS

ISSUED MAY 2009



## *FOREWORD*

This inaugural publication contains research highlights from the Computation, Computers, Information and Mathematics Center (CCIM) at Sandia National Laboratories. CCIM and its close partners at Sandia provide national leadership in High Performance Computing (HPC) and create technology and solutions for many of our Nation's most demanding national security challenges. Computing permeates all corners of Sandia. Most importantly, the nuclear weapons program — the centerpiece of the National Nuclear Security Administration — depends critically upon advanced modeling and simulation to ensure the continued safety and reliability of the U.S. nuclear stockpile. But every national security mission at Sandia — from energy to homeland security to nonproliferation — relies on advanced computing.

The vignettes in this publication introduce some of the ongoing work within CCIM. Collectively, they describe a portfolio that spans the spectrum from fundamental research to state-of-the-art applications. Vertical integration of algorithms, computer architectures, and applications-rooted uncertainty quantification to provide predictive capabilities for modeling complex phenomena is fundamental to our ability to contribute to national needs. Much of our work involves enabling technology for modeling physical and engineering systems; however, we also have extensive research programs in decision support, informatics, and discrete modeling.

The computing research enterprise at Sandia is closely tied to its broader set of missions and strategies. Thrusts within our strategic plan include national leadership in large-scale informatics, extreme computer architectures, and in the transformation of engineering practice through the use of advanced modeling and simulation technologies. Application focus areas include nuclear weapons, cyber surety, and energy and CO<sub>2</sub> challenges, which include climate modeling, alternative energy technologies, and improvements to the power grid. We also serve as stewards of important capabilities for the Nation in high strain-rate physics, electric modeling, scientific visualization, mesh generation, and computational materials. Many of these thrust, focus, and stewardship areas are touched upon in this publication.

This publication is the first in what will be a yearly series. It is organized according to several crosscutting themes: algorithms and mathematics, enabling tools for predictive simulation, computer architectures and systems, and application impacts. It is not intended to be complete, but rather to illustrate the progress made within, and among, many organizations at Sandia to serve the national need through advanced computational science.

We hope you find these topical reports inspiring and educational and welcome your feedback.

***James S. Peery, Director***

*Computation, Computers, Information and Mathematics*

# Contents

## Foreword

|  |       |
|--|-------|
| <b>I - Algorithms and Mathematics</b> .....  | I-1   |
| New Partitioning Algorithms Reduce Communication Overload in Parallel Computing.....                               | I-2   |
| Optimal Scheduling of Error Correction Brings Quantum Computing One Step Closer.....                               | I-4   |
| Mathematical Algorithms Enable Unprecedented Partial Differential Equation Solutions.....                          | I-6   |
| Time Integration Algorithms Enable Complex Simulations.....  | I-8   |
| Lagrangian Tracer Particles in Eulerian Simulations Help Materials Make a Clean Break.....                         | I-10  |
| Achieving Accurate Material Interfaces Through Patterned Interface Reconstruction.....                             | I-12  |
| New Framework Supports Robust Multi-physics Coupling.....  | I-14  |
| Scalable Solution Techniques Developed to Address Damage to Semiconductor Devices.....                             | I-16  |
| Modeling Multi-Material Systems with the Extended Finite Element Method.....                                       | I-18  |
| Tolerant Geometry Tools Make Possible the Generation of Contiguous<br>Meshes for Sloppy Assembly Models.....       | I-20  |
| Automated Topology Adjustments for Sweeping Enables More Accurate<br>Simulation of Complex Geometry.....           | I-22  |
| Advances in Mesh Optimization Create More Robust and Accurate Simulations.....                                     | I-24  |
| Geometric Capture Methods Provide New Tools for Hexahedral Mesh Generation.....                                    | I-26  |
| New Algebraic Multi-grid Capabilities Model Stretched Meshes and Anisotropic Problems.....                         | I-28  |
| Algebraic Multi-grid Algorithms Scale to Thousands of Processors.....  | I-30  |
| New Algorithms Improve Performance on Next Generation Multi-core Computers.....                                    | I-32  |
| Scalability-Cubed: Algorithms, Capability Development, and Interoperability<br>via the Trilinos Framework.....     | I-34  |
| Tensor Decompositions Provide New Methods for Analyzing Critical Data Mining Applications.....                     | I-36  |
| Advanced Methods Allow Processing and Searching of Very Large Document Collections.....                            | I-38  |
| New Techniques Support the Calibration of Model Parameters Under Uncertainty.....                                  | I-40  |
| New Computational Methods Enable Novel Nanoparticle Simulations.....   | I-42  |
| Advanced Algorithms Enable Unique Modeling of Nanoparticles in Solution.....                                       | I-44  |
| Sandia Operations Research Reduces Drinking Water Contamination Risks.....   | I-46  |
| New Algorithms Improve Water Quality Warnings.....   | I-48  |
| CoreSim Enables Larger Logistics and System-of-Systems Simulations.....  | I-50  |
| Biological Inspiration Motivates New Network Protocols.....  | I-52  |
| Embedded Sensitivity Analysis Enables High Impact Predictive Simulation.....                                       | I-54  |
| Peridynamics Unifies the Mechanics of Continua, Fracture, and Statistical Mechanics.....                           | I-56  |
| Redefining Chemical Compound Space for More Efficient Materials Design Algorithms.....                             | I-58  |
| <b>II - Enabling Tools for Predictive Simulation</b> .....   | II-1  |
| Dakota Toolkit Enables Large-Scale Optimization and Uncertainty Quantification for Sandia’s Missions.....          | II-2  |
| Automatic Differentiation Enables Accurate Derivative Evaluations for Predictive Simulation.....                   | II-4  |
| Meshing and Geometry for Modeling and Simulation with CUBIT.....   | II-6  |
| Structural Simulation Toolkit Allows Early Insight into Performance of<br>Microscale Supercomputer Simulation..... | II-8  |
| Expediting Device Modeling Using Verilog-A Compact Model Compiler.....   | II-10 |
| Titan toolkit Makes Large-Scale Analytics Possible.....  | II-12 |
| Extending ParaView for Fragment Analysis on Extremely Large Data.....  | II-14 |

**III - Architectures & Systems** ..... III-1

- The New Moore’s Law and Sandia’s Role in the Future of Computing ..... III-2
- Comparison Reveals Relative Strengths of Multithreaded and Distributed Memory Architectures ..... III-4
- Enhancing Next Generation Interconnection Networks for High-Performance Computing ..... III-6
- Operating System Jitter Can Make Applications Nervous ..... III-8
- Portals Networking Software Powers Petaflops Performance ..... III-10
- Catamount Lightweight Operating System Maximizes Multi-Core Compute Resources  
While Minimizing Power ..... III-12
- Allocation Algorithms Maximize Throughput in Parallel Supercomputers ..... III-14
- Enabling Large-Scale Informatics Using Novel Architectures: Pythia and X-caliber ..... III-16
- Kitten Lightweight Kernel Makes Multi-Core Processors Purr ..... III-18
- The Alliance for Computing at Extreme Scales (ACES) Targets Production Capability  
Computing for NNSA ..... III-20

**IV - Driving Applications** ..... IV-1

- Unprecedented Scale Semiconductor Modeling Conducted on Massively Parallel Platforms ..... IV-2
- New Solution Methods for Multiple-time-scale Resistive and Extended Magnetohydrodynamics ..... IV-4
- Rigorous Code Verification Drives Major Improvements in  
Computational Shock Magnetohydrodynamics ..... IV-6
- Developing Scalable Preconditioners for Parallel Electrical Simulation ..... IV-8
- Predicting Defect Physics for Radiation Damage in Semiconductor Devices ..... IV-10
- Enabling Quick and Fast Predictive Multi-Scale Modeling and Simulations ..... IV-12
- Uncertainty Quantification Methodology for Damaged Semiconductor Devices ..... IV-14
- Addressing Aging Problems Through Enhanced Low Dose-Rate Sensitivity (ELDRS)  
and other Radiation Effects Calculations ..... IV-16
- Circuit Based Framework Improves Modeling and Simulation of Latent  
*Mycobacterium Tuberculosis* ..... IV-18
- Developing Two-Fluid Models for Simulating Singlet Oxygen Generator Reactors ..... IV-20
- Developing Flexible Multi-Paradigm Modeling and Simulation Environments for Complex Planning ..... IV-22
- Evaluating Impacts of Arctic Climate Change on National Security ..... IV-24
- Large-Scale Atomistic Simulation Developed to Study Stockpile Materials Chemistry ..... IV-26
- Advanced Uncertainty Quantification Methods Enable Understanding of Impact of  
Radiation Environment on Nuclear Weapons ..... IV-28
- Effectively Modeling Z-machine Wire Array Implosions ..... IV-30
- An Advanced Tool for Plasma Simulation ..... IV-32
- Community Climate System Model Development for Petascale Supercomputers ..... IV-34
- Informatics Enables Data-Driven Discovery and Prediction ..... IV-36
- Verification and Validation Methodology for Computational Cognitive,  
Behavioral, and Social Models ..... IV-38

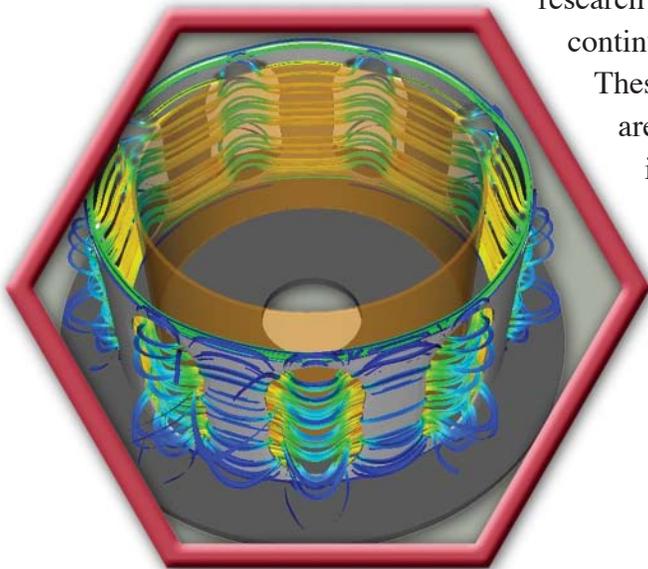
**Afterword**

# ALGORITHMS AND MATHEMATICS

The development of computational algorithms, and the mathematics underlying those algorithms, lies at the historical heart of the Computation, Computers, Information and Mathematics (CCIM) Center 1400. Massively parallel high performance computing (HPC) hardware, and the associated system infrastructure required to use, operate, and manage it is necessary for all contemporary leading-edge computational science, whose core continues to rest on innovative algorithms. For example, the experience of NNSA's Advanced Simulation and Computing (ASC) program over the past 15 years has been that algorithm advances have had a greater impact on HPC capabilities rather than on hardware advances. There is a close connection between the development of sophisticated and productive algorithms and a thorough understanding of increasingly complex and demanding HPC architectures. Thus, algorithm and architecture research within CCIM constitute two points of the triangular partnership that fundamentally expresses our work, the third point being driving applications. Broad classes of algorithm research and development continue to typify CCIM algorithm research themes, including parallel linear algebra, high-performance discretizations of important nonlinear partial differential equations, a large class of discrete algorithm

research for more general applications, both discrete and continuous optimization, and uncertainty quantification.

These algorithms have intersected with key applications areas within Sandia, as this collection of Highlights illustrates.



# NEW PARTITIONING ALGORITHMS REDUCE COMMUNICATION OVERLOAD IN PARALLEL COMPUTING

Contact:

Erik Boman, Scalable Algorithms, [egboman@sandia.gov](mailto:egboman@sandia.gov)

*As part of research supported by the CSCAPES (Combinatorial Scientific Computing and Petascale Simulations) SciDAC institute, we have developed new sparse matrix partitioning algorithms that use a 2D partitioning... We believe our data partitioning will be useful in a variety of algorithms, not just matrix-vector multiplication.*

Load balancing data among processors is crucial to the scalability of parallel codes. Software tools such as Sandia's Zoltan provide partitioning algorithms that compute parallel data distributions. The data distribution determines the work partitioning, and so is central to load balancing. For matrix computations, the data partitioning is typically done in a one-dimensional (1D) fashion; that is, the matrix is partitioned by rows (or columns, but not both). As part of research supported by the CSCAPES (Combinatorial Scientific Computing and Petascale Simulations) SciDAC institute, we have developed new sparse matrix partitioning algorithms that use a 2D partitioning as shown schematically in Figure 1 and for an application matrix in Figure 2. These 2D partitionings reduce the communication requirement substantially compared to the 1D approach.

We have studied a particularly important kernel in scientific computing, sparse matrix-vector multiplication, which is the crux of many iterative solvers. A new algorithm based on nested dissection (recursive substructuring) has been developed [1]. Empirical experiments show that the method clearly outperforms 1D partitioning and is competitive (in quality) with other proposed 2D methods that have been deemed impractical since they are too expensive to compute. In contrast, our method takes similar time to compute as traditional graph or hypergraph

partitioning. On a test set of sparse matrices from diverse applications like finite element computations, circuit simulation, and text processing (informatics), we observed an average reduction in communication volume of 15% for symmetric matrices but up to 97% reduction in extreme cases (Figure 3). The largest gains were for applications with highly irregular structure, like electrical circuit models, informatics, and matrices from constrained optimization. We believe our data partitioning will be useful in a variety of algorithms, not just matrix-vector multiplication.

Our new partitioning algorithm is currently being implemented in the Isorropia package (supported by the Advanced Simulation and Computing Program [ASC]). Isorropia provides partitioning and load-balancing services to Trilinos, together with Zoltan.

## Reference

1. Boman, E. G., and Wolf, M. M. (2008). *A Nested Dissection Approach to Sparse Matrix Partitioning for Parallel Computations*, SAND 2008-5482J, Sandia National Laboratories, Albuquerque, NM.

*This work has been supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program.*

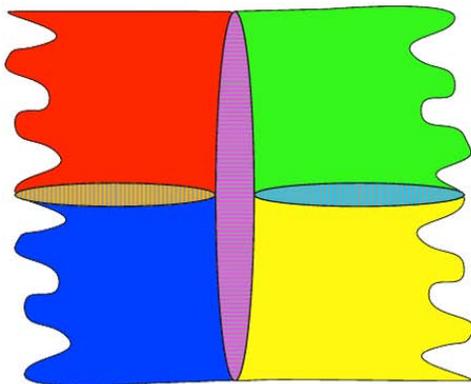


Figure 1(a)

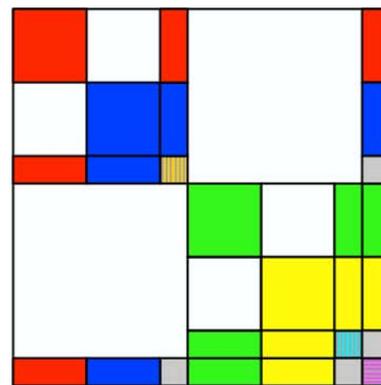


Figure 1(b)

Figure 1. (a) Mesh or graph partitioned into four regions (one for each processor), denoted by different colors. (b) The corresponding partitioned matrix, with a suitable ordering.

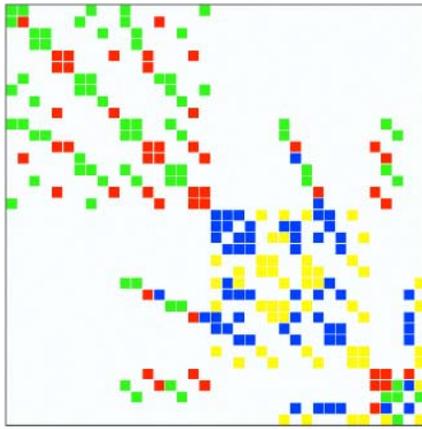


Figure 2(a)

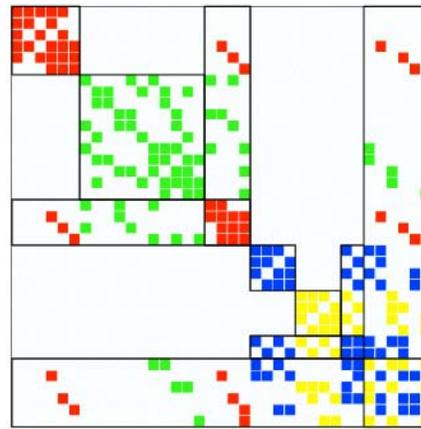
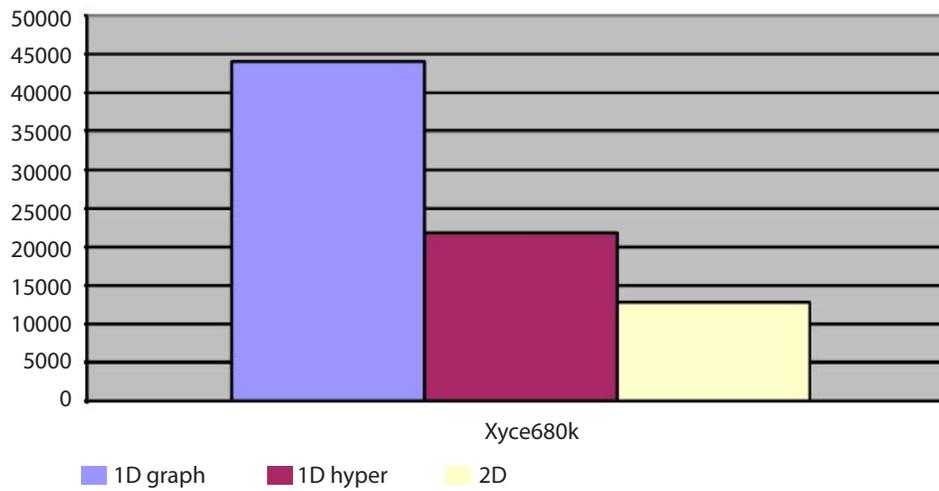


Figure 2(b)

**Figure 2.** (a) A small, sparse matrix from a DNA electrophoresis model partitioned into four parts with our new nested dissection algorithm. Figure 2(b) shows the same matrix permuted to expose the recursive structure.



**Figure 3.** The communication volume for a sparse matrix used to simulate a circuit with Xyce, for 64 processors.

# OPTIMAL SCHEDULING OF ERROR CORRECTION BRINGS QUANTUM COMPUTING ONE STEP CLOSER

Contacts: Robert Carr, Discrete Math and Complex Systems, [rdcarr@sandia.gov](mailto:rdcarr@sandia.gov)  
 Anand Ganti, Advanced Networking Integration, [aganti@sandia.gov](mailto:aganti@sandia.gov)  
 James Levy, Mixed Signal ASIC/SoC Products, [jelevy@sandia.gov](mailto:jelevy@sandia.gov)  
 Cynthia Phillips, Discrete Math and Complex Systems, [caphill@sandia.gov](mailto:caphill@sandia.gov)  
 Jean-Paul Watson, Discrete Math and Complex Systems, [jwatson@sandia.gov](mailto:jwatson@sandia.gov)

*Researchers have found optimal initial schedules to implement error correction for quantum computers... Efficient error correction improves both the computer throughput and the maximum tolerable gate error rate.*

Researchers have found optimal initial schedules to implement error correction for quantum computers. Quantum computers, where information encoding and processing are guided by the principles of quantum mechanics, could solve problems that are theoretically and practically intractable for conventional computers. For example, quantum algorithms for factoring and general search outperform the algorithms for these problems on conventional computers. Quantum computers could also be an excellent platform for quantum physics simulations.

The quantum-computing analog of a conventional bit is a qubit, which is manipulated by quantum gates. Sandia's Quantum Information Science and Technologies LDRD project is building a single physical qubit and designing a logical qubit. A logical qubit uses multiple faulty physical qubits to encode a single error-corrected logical qubit.

Error correction is a critical component for proper operation of a quantum computer. It is impossible to isolate the physical qubits, in this case implemented as double quantum dots, from the physical world. The interaction of the qubits with their surroundings, neighboring qubits, and control itself results in the accumulation of errors. Physical qubit errors accumulate constantly and continuously, unlike the discrete flip errors possible in conventional bits.

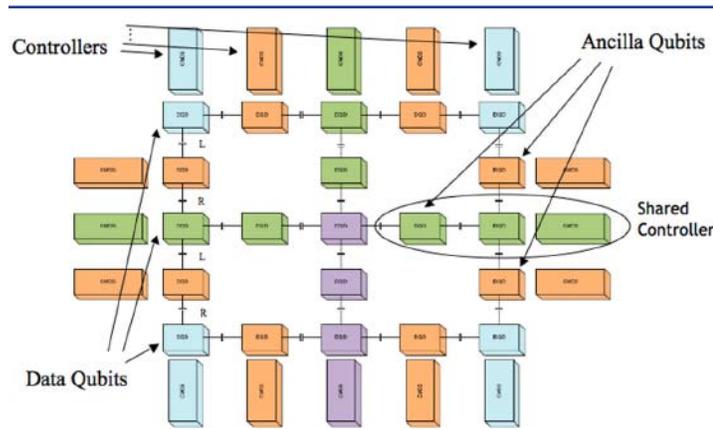


Figure 1. Logical qubit architecture.

Error correction must occur continuously during a quantum algorithm. It will dominate the operations on a quantum computer, and may account for 95% or more of the total operations. Efficient error correction improves both the computer throughput and the maximum tolerable gate error rate.

The precise nature of the error correction depends upon the specific architecture. The initial architecture for the LDRD is shown in Figure 1. The cubes represent qubits. There are 9 data qubits, the outer and middle qubits of each row and column. There are 12 ancilla qubits, which interact with data qubits in 2-qubit gates. Since observing (measuring) a qubit destroys its quantum information, error correction circuits measure only the ancilla qubits to infer the states of the data qubits. The rectangular boxes on the outside of the array of qubits are controllers. The coloring shows the pattern of controller sharing among qubits. For example, there are three purple qubits sharing a single controller.

The initial logical qubit architecture uses a Bacon-Shor error correction code. This involves implementing a circuit composed of quantum gates: one for each row and one for each column as shown in Figure 2. The operations are of different types, different lengths, and may involve one or two qubits.

The schedule assigns a start time to each operation. Operations must obey a partial order of precedence constraints. For each data qubit, the schedule must order the row/column block for each qubit. That is row/column operations cannot interleave, except for limited commuting at the boundary. Each controller can only send one type of signal at a time, so all qubits sharing a controller must always be doing the same thing, or be idle. Two-qubit operations must be synchronized and obey constraints on left/right quantum dot participation. The goal of the schedule is to minimize idle time. This involves the makespan (total length of the schedule) for data qubits and the idle time within a circuit for ancilla qubits.

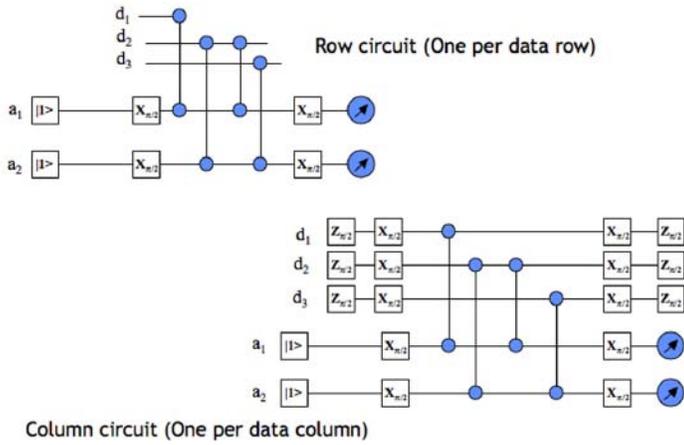


Figure 2. Row and column circuits to perform Bacon-Shor error correction.

We have computed optimal schedules using integer programming for both this initial architecture and a 13-qubit architecture that reuses ancilla. Figure 3 shows an optimal schedule for the architecture in Figure 1 that continuously

runs a Bacon-Shor error correction code. This circuit creates a quantum memory continuously holding its state.

Future error-correction methods that better control errors during the error correction itself will require more complex, more constrained schedules. Computing these schedules might require parallel integer programming, using the CCIM Parallel Integer and Combinatorial Optimizer (PICO) code, in combination with constraint programming techniques.

**Reference**

Levy, J. E., Ganti, A., Phillips, C., et al. (2008). *The Impact of Classical Electronics Constraints on a Solid-State Logical Qubit*, SAND 2008-7683C, Sandia National Laboratories, Albuquerque, NM.

*This work's original support was from the SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program.*

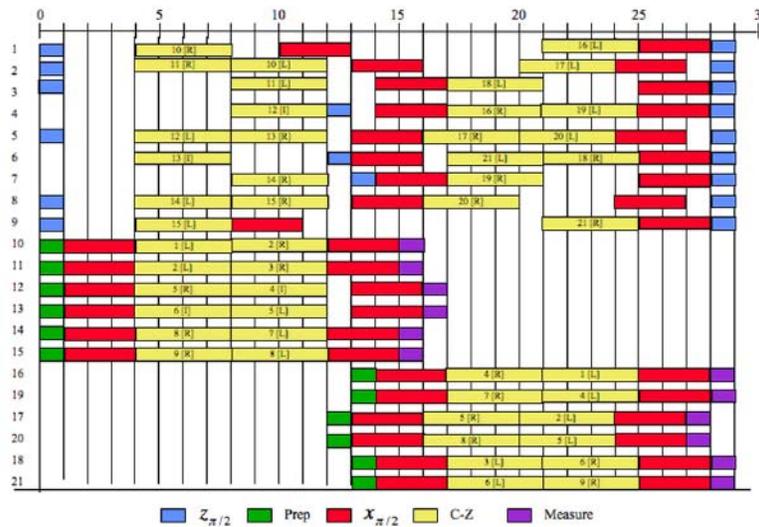


Figure 3. An optimal schedule for the 21-qubit architecture.

# MATHEMATICAL ALGORITHMS ENABLE UNPRECEDENTED PARTIAL DIFFERENTIAL EQUATION SOLUTIONS

Contacts: Pavel Bochev, Applied Math and Applications, [pbboche@sandia.gov](mailto:pbboche@sandia.gov)  
 D. Ridzal, Optimization and Uncertainty Quantification, [dridzal@sandia.gov](mailto:dridzal@sandia.gov)  
 R. Tuminaro, Scalable Algorithms, [rstumin@sandia.gov](mailto:rstumin@sandia.gov)

*The outcome of this research spurred the development of next-generation software tools for the numerical solution of PDEs and enabled remarkable improvements in the performance of the solvers in the ML linear solver package in Trilinos.*

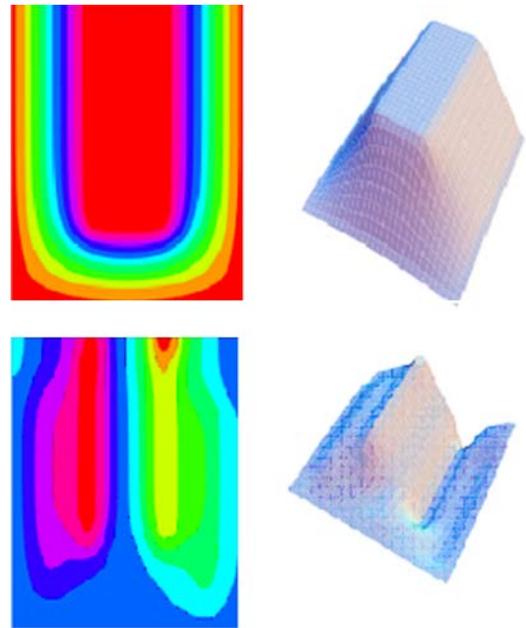
Compatible discretizations that preserve key structural properties of mathematical models, and efficient, discretization-aware solvers for the resulting algebraic equations are at the heart of predictive simulations. The Computation, Computers, Information and Mathematics group has historically been at the forefront of the research in these areas and has led in the development and delivery of cutting edge software tools for high-performance computing, based on the most recent algorithmic advances.

In the past few years researchers have made significant strides in the formulation and analysis of compatible discretizations [1, 2] and new multi-grid methods [3-5] that take into account the special structure of the corresponding discrete equations. The outcome of this research spurred the development of next-generation software tools for the numerical solution of partial differential equations (PDEs) and enabled remarkable improvements in the performance of the solvers in the ML linear solver package in Trilinos.

A unique feature of the research approach is a common framework for compatible, or mimetic, discretizations, based on algebraic topology, which offers mutually consistent notions of discrete derivative and integral, preserves the invariants of the de Rham cohomology groups, and has a discrete Stokes theorem. The framework and all attendant discrete structures are defined by using two basic mappings between differential forms and cochains. Owing to their structure-preserving properties, compatible methods avoid unphysical spurious modes (Figure 1).

This research provides theoretical foundations for the development of an innovative library for compatible discretizations (Intrepid), which is part of the discretization capability area of Trilinos 9.0. Intrepid allows access to finite-element (FEM), -volume (FV) and -difference (FD) methods through a common application programming interface and supports hybrid discretizations on unstructured grids.

Work on compatible discretizations is carried out in close coordination with the development of new multigrid algorithms. Because compatible discretizations preserve



**Figure 1.** Simulation of magnetic diffusion using compatible (top) and standard nodal (bottom) finite elements. The latter develop smooth spurious modes that are difficult to detect and can be detrimental to the accuracy of coupled multiphysics simulations such as magnetohydrodynamics.

the cohomology of the De Rham complex, scalable multi-grid solvers must be designed in a way that respects the structure of the discrete spaces given by a discrete Hodge decomposition. For example, a dramatic decrease in the solution time (Figure 2) for a compatible discretization of the eddy current equations, which leads to a curl-curl matrix with a large null-space, has been achieved by a multigrid algorithm [2] that takes into account the characterization of this null-space as the range of a discrete gradient. Standard iterative algorithms fail due to their inability to handle the null-space and the ensuing near singularity of the solution operator.

Further improvements to this algorithm [1], and its extension to grad-div operators [3] resulted from a new reformulation approach for  $H(\text{curl})$  and  $H(\text{div})$  problems based on the existence of a discrete Hodge decomposition for

compatible approximations of these spaces. The table below illustrates the success of the new approach to handle magnetic diffusion problems with widely varying conductivities. It shows the number of iterations required to reduce the residual by 10 orders of magnitude for a 2D model problem corresponding to two materials [1].

| Grid    | cplx | $\sigma_2$ |           |           |           |           |
|---------|------|------------|-----------|-----------|-----------|-----------|
|         |      | $10^0$     | $10^{-2}$ | $10^{-4}$ | $10^{-6}$ | $10^{-8}$ |
| $9^2$   | 1.07 | 7          | 7         | 7         | 7         | 7         |
| $27^2$  | 1.20 | 12         | 12        | 12        | 12        | 12        |
| $81^2$  | 1.25 | 15         | 16        | 16        | 16        | 16        |
| $243^2$ | 1.27 | 17         | 18        | 18        | 18        | 18        |

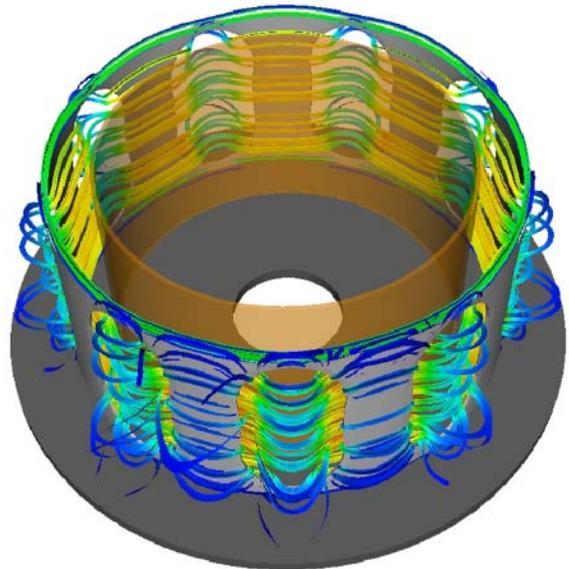
### References

1. Bochev, P., and Hyman, M. (2006). Principles of Mimetic Discretizations, *Volumes in Mathematics and Its Applications*, Vol. 142, Springer, eds. D. Arnold, P. Bochev, R. Lehoucq, R. Nicolaides, and M. Shashkov, pp. 89-121.

2. Bochev, P., and Ridzal, D. (2007). *Rehabilitation of the Lowest-Order Raviart-Thomas Element on Quadrilateral Grids*, SAND 2007-6297J, Sandia National Laboratories, To be published in *SIAM Journal on Numerical Analysis*.
3. Bochev, P., Hu, J., Siefert, C., Tuminaro, R. (2008). An Algebraic Multi-grid Approach Based on a Compatible Gauge Reformulation of Maxwell's Equations, *SIAM: Journal of Scientific Computing*, Vol. 31, No. 1, pp. 557-583.
4. Hu, J., Tuminaro, R., Bochev et al. (2006). Toward an H-Independent Algebraic Multi-Grid Method for Maxwell's Equations, *SIAM: Journal of Scientific Computing*, Vol. 27, No. 5, pp. 1669-1688.
5. Bochev, P. B., Siefert, C. M., Tuminaro, R. S., et al. (2007). *Compatible Gauge Approaches for H(div) Equations*, CSRI Summer Proceedings, SAND2007-7977, Sandia National Laboratories, Albuquerque, NM, pp. 3-14.

*This work has been supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

**Figure 2.** 3D magnetic field generated by electrical current running through an idealized plasma liner and returning through a slotted current return as commonly found in the Z-Machine. The presence of slots perturbs the magnetic field topology, impacting the physics occurring inside. High-resolution versions of this simulation require 60 seconds to solve on 50 processors. Prior to the development of the new multigrid solver in ML this simulation was not feasible due to convergence failures of the more standard linear solvers.



# TIME INTEGRATION ALGORITHMS ENABLE COMPLEX SIMULATIONS

Contacts: Todd Coffey, Applied Math & Applications, [tscoffe@sandia.gov](mailto:tscoffe@sandia.gov)  
 Roscoe Bartlett, Optimization and Uncertainty Quantification, [rabartl@sandia.gov](mailto:rabartl@sandia.gov)  
 John Shadid, Electrical and Microsystem Modeling, [jnshadi@sandia.gov](mailto:jnshadi@sandia.gov)  
 Roger Pawlowski, Applied Math & Applications, [rppawlo@sandia.gov](mailto:rppawlo@sandia.gov)

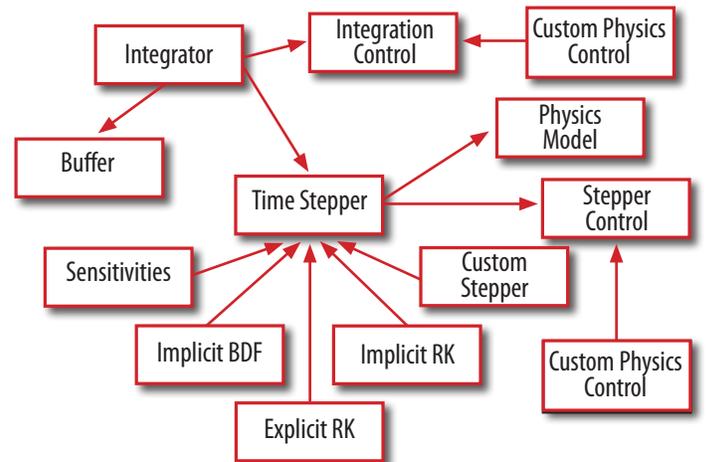
*We have developed Rythmos, a state-of-the-art collection of time integration methods to support Sandia's most demanding simulation needs. ... Rythmos is integrated into several physics packages at Sandia.*

Time integration is the process of moving forward in time to solve a differential equation. This process can be as easy as looking up the solution in a book or as complex as assembling a multitude of available physics models and transforming them together under a numerical theory to solve a huge nonlinear differential equation on thousands of processors of a parallel super computer. We are interested in problems in the latter half of that spectrum. We have developed Rythmos, a state-of-the-art collection of time integration methods to support Sandia's most demanding simulation needs.

The focus of the present work is an advanced software package for time integration of differential equations on parallel computing platforms. Rythmos is a component within the Trilinos object-oriented software framework for numerical algorithms and enabling technologies. The goal of Rythmos is to provide physics applications with a modular implementation of a broad spectrum of time integration algorithms, including the capability to interoperate with solvers and advanced methods, such as sensitivity analysis, through core Trilinos data structures and numerical kernels (Figure 1). Rythmos, together with the larger Trilinos system, is open-source software released under the LGPL and is available from <http://trilinos.sandia.gov>.

Why do we care about time integration? In short, we seek to understand and, ultimately, predict the behavior of complex natural and man-made systems, and at the core of that understanding is the ability to mathematically model and simulate the phenomena inherent to the system. Any phenomenon that changes with time can be described as a transient differential equation and integrated in time for its solution. The application of differential equations is as limitless as the variety of physics seen in the natural world.

Sadly, the mathematics community has yet to find a single time integration algorithm that works for all problems. Every method makes trade-offs between accuracy, stability, speed, and memory. It is often very challenging for a



**Figure 1.** Rythmos has a very flexible design allowing a physics package to modify its behavior at many levels.

physics simulation team to determine an appropriate time integration algorithm. And as projects themselves evolve, the governing physics can change, and different time integration algorithms become better alternatives. Fundamentally, Rythmos is designed to address the problem of finding the right time integration algorithm and keeping it current as a project evolves.

There are many challenges to achieving an accurate and robust time integration capability for a wide class of problems. The first challenge is related to pairing up a good time integration algorithm with the specific physics of the governing physics model. For example, often there are special adjustments made to integration algorithms to account for peculiarities in the physics or to take advantage of special structures or application knowledge. In Rythmos, we have taken the approach of supporting these adjustments by implementing a broad class of integrators in a general way. This means a physics application can migrate to Rythmos and continue to use the same algorithms.

There are additional challenges in implementing new time integration algorithms into existing physics modeling packages. Such applications are often written with a single, deeply embedded, time integration algorithm and require extensive modification to accommodate new integration algorithms. Rythmos has a great deal of flexible code and many innovations in design in order to replicate the functionality of a specific physics package while providing a variety of different algorithms, all under one interface.

Rythmos has been developed with great flexibility in mind and offers the possibility of modifying the behavior of almost all the functions it provides. Figure 1 illustrates an abbreviated diagram of the main classes. Rythmos supports both multistep and one-step time integration algorithms and, implicit and explicit algorithms, and can handle both ordinary differential equations and differential-algebraic equations. More advanced algorithms like forward and adjoint sensitivities are also supported, and research is under way to provide adjoint based global error estimation capabilities too.

Rythmos is integrated into several physics packages at Sandia. The pictures in Figure 2 demonstrate the use of Rythmos in simulating magnetohydrodynamics (MHD)

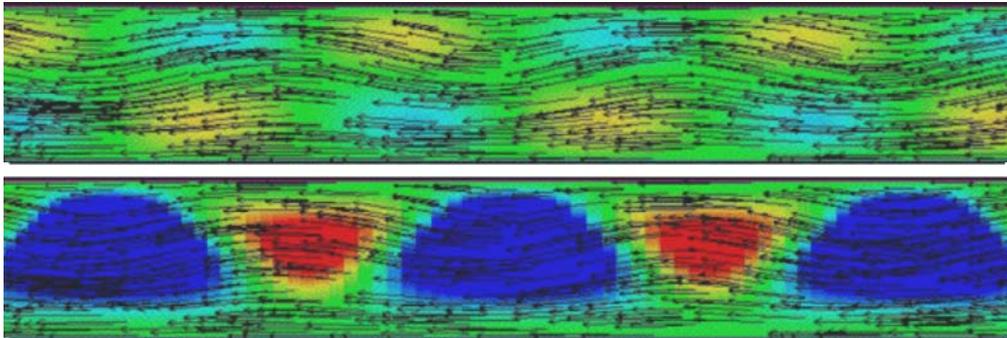
in Sandia's distributed memory parallel code Charon. Rythmos provides both implicit multi-step time integration algorithms and forward sensitivities for Charon. In the near future, Charon will explore the use of all the implicit time integration algorithms in Rythmos to study the trade-offs between accuracy, stability, speed, and memory. This is a unique capability and will greatly enhance Charon's modeling capabilities.

### References

Bartlett, R., Collis, S., Coffey, et al. (2007). *ASC Vertical Integration Milestone*, SAND2007-5839, Sandia National Laboratories, Albuquerque, NM.

Bartlett, R. *Derivation of Forward and Adjoint Sensitivities for ODEs and DAEs*, SAND2007-6699, Sandia National Laboratories Report, Albuquerque, NM.

*This work was partially funded by the Department of Energy Office of Science ASCR Applied Math program, and by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*



**Figure 2.** Employing a fully implicit fully coupled Newton-Krylov solution technology, we have studied stability and nonlinear evolution problems for thermal buoyancy driven flows with interacting magnetic fields (hydro-magnetic Rayleigh-Bernard type instabilities). These coupled mechanisms are, for example, critical components of large-scale geo-dynamo simulations that model the time dependent behavior of the Earth's magnetic field.

# LAGRANGIAN TRACER PARTICLES IN EULERIAN SIMULATIONS HELP MATERIALS MAKE A CLEAN BREAK

Contacts: David M. Hensinger Computational Shock and Multi-physics, [dmhensi@sandia.gov](mailto:dmhensi@sandia.gov)  
O. Erik Strack, Computational Shock and Multi-physics, [oestrac@sandia.gov](mailto:oestrac@sandia.gov)

*The effect of including Lagrangian tracers as a method of preserving advected material properties is easily seen when comparing simulations performed with and without Lagrangian material tracers.*

Introducing tracer particles that carry material state information into Eulerian simulations improves the accuracy of calculations for materials subjected to high strain rates. High-velocity impacts, explosions, and electromagnetic transients are examples of high strain rate producing events. Such events have historically required Eulerian simulation techniques in which a fixed mesh of computational cells is used to model material flowing through the mesh. The calculation of this flow (advection) is carried out based on the assumption that material properties are continuous values (homogenous). Unfortunately, this assumption is frequently invalid because real materials have varying properties (heterogeneous), and this heterogeneity is an important part of the material response (as it often is for fracture and fragmentation).

The phenomenon referred to as “numerical healing” [1] is one consequence of the application of Eulerian advection to the simulation of systems with discontinuously varying material properties. Numerical healing occurs when material moves between analysis cells and the material properties in one cell are mixed with the material properties in a neighboring cell. As an example consider an analysis cell holding failed material with a strength value of zero. When some of this material is advected into an adjacent cell with undamaged material, the resulting material strength is no longer zero. As the material moves further beyond its original

cell, the zero strength value becomes smeared between several cells, and the formerly failed material has now been “numerically healed.”

As a solution to numerical healing, Lagrangian tracer particles are proposed as a mechanism for preserving discontinuous material properties from smearing in Eulerian simulations (Figure 1, right side). Instead of abandoning the Eulerian calculation in favor of a particle formulation [2-6], the traditional Eulerian framework provides the simulation foundation that ensures correct mass and momentum transport, while the tracers preserve the correct material state information. At the beginning of a simulation, the tracers are distributed through the solution domain, and assigned the properties of the material occupying the cells within which they reside. Subsequently, as the solution develops, the tracers are advected through the mesh along with the material. The response of each simulation cell is derived from the tracers within the cell, and the material representation carried by the tracers is updated based on the strain history of the cell within which it resides. If the material within a cell experiences failure, then the tracers within the cell will have failed. Because the properties of a particle are never mixed with the properties of other tracers, failed tracers will remain failed no matter how many times they are advected across cell boundaries.

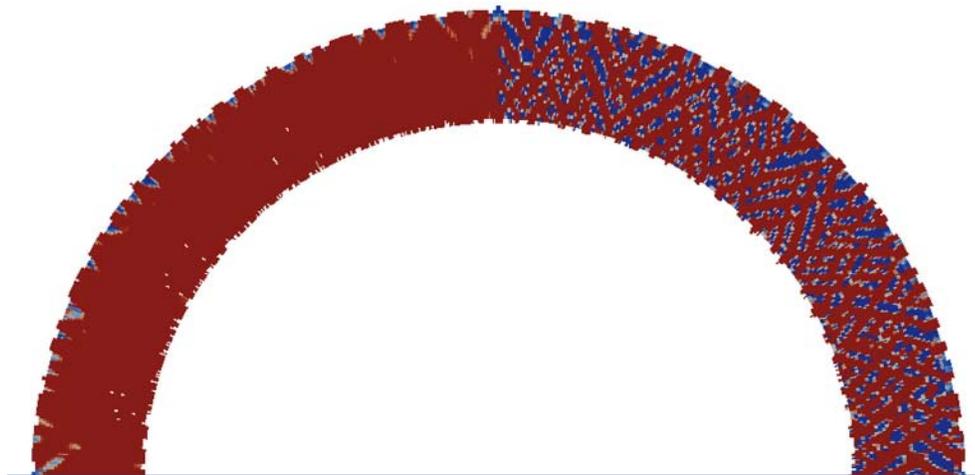


Figure 1. Failure state of a cylindrically imploded ring of heterogeneous material.

Assuring that the traditional Eulerian material formulation remains synchronized with the Lagrangian particle formulation poses many technical challenges. First, the movement of tracers between cells may cause a sudden change in the response of a system, causing numerical noise. Also, expansion of material during a simulation may lead to a dearth of tracers in some regions of a simulation. Finally, inconsistencies between the advection methods for tracers and material interfaces can lead to the migration of tracers across material interfaces. This project involves a substantial effort in verifying that the method preserves the conserved quantities of the underlying Eulerian methods, correctly preserves the material states, and converges as the computational mesh is refined.

Supplementing Eulerian simulations with Lagrangian tracers improves the correspondence between experimental and simulation results. The enhanced Eulerian simulations can capture system response that is driven by heterogeneity. The effect of including Lagrangian tracers as a method of preserving advected material properties is easily seen when comparing simulations performed with and without Lagrangian material tracers (Figure 1). It shows the failure state of a cylindrically imploded ring of heterogeneous material. The left half of the simulation was performed without tracers. The right half of the image included tracers. The colors represent the extent of material failure. On the left hand side, the diffusive nature of the simulation has reduced the heterogeneity of the material and reduced the complexity of the failure mechanisms. On the right hand side, the material properties have been preserved, and the failures retain significant structure.

### References

1. Brannon, R. M. (2007). Validating Theories for Brittle Damage, *Metallurgical and Materials Transactions A*, pp. 2861-2868.
2. Brackbill, J. U. (1988). A Low Dissipation Particle in Cell Method for Fluid Flow, *Computer Physics Communications*, Vol. **48**, No. 1, pp. 25-38.
3. Silling, J. A. A Meshfree Method Based on the Peridynamic Model of Solid Mechanics, *Computers and Structures*, Vol. **83**, Nos. 17-18, pp. 1526-1535.
4. Sukumar, N. (1997). An Element-Free Galerkin Method for Three-Dimensional Fracture Mechanics, *Computational Mechanics*, Vol. **20**, No. 1-2, pp. 170-175.
5. Belytschko, T. (2005). Element-Free Galerkin Methods, *International Journal for Numerical Methods in Engineering*, Vol. **37**, No. 2, pp. 229-256.
6. Moresi, L. (2003). A Lagrangian Integration Point Finite Element Method for Large Deformation Modeling of Viscoelastic Geomaterials, *Journal of Computational Physics*, Vol. **184**, No. 2, pp. 476-497.

***This work was funded by the Project Manager for the Heavy Brigade Combat Team (PM HBCT, Mr. John Rowe).***

# ACHIEVING ACCURATE MATERIAL INTERFACES THROUGH PATTERNED INTERFACE RECONSTRUCTION

Contacts: Stewart Mosso, Computational Shock and Multi-physics, [sjmosso@sandia.gov](mailto:sjmosso@sandia.gov)  
Thomas Voth, Exploratory Simulation Technologies, [tevoth@sandia.gov](mailto:tevoth@sandia.gov)

*The Patterned Interface Reconstruction algorithm will significantly improve the fidelity of multiple material simulations in [the code] ALEGRA and will reduce both the unphysical fragment emission and the unphysical distortion of bodies.*

Computer simulations of challenging mechanics problems, such as explosive-armor interactions, are often performed using operator-split Eulerian hydrodynamics codes. In these codes, multiple materials are moved through the computational mesh via a Lagrangian step followed by a remap step. The position of materials is described by the volume of each material in the computational elements. Interface reconstruction algorithms are used to move (remap) the material volumes conservatively (i.e., the amount of material remains constant) through the mesh, and the shape of the materials changes only through physical interactions with other materials. This goal is more readily achieved with higher order methods. The Patterned Interface Reconstruction algorithm in the ALEGRA code [1, 2] is second order. Briefly, a method's order of accuracy describes the rate at which errors decrease as the element size of the computational mesh decrease. The order of accuracy is the exponent for the rate of reduction.

The challenge of interface reconstruction is to simulate material motion with as little artificial distortion as possible (i.e., distortion not associated with internal or external forces). These distortions can be observed as (i) changing the shape of the body, (ii) inducing a jagged, discontinuous outer boundary, and (iii) emitting unphysical fragments. A second-order interface reconstruction algorithm can exactly reproduce a planar material interface. While there are several

interface reconstruction methods that can represent planar interfaces exactly, the much harder challenge is reconstructing curved material interfaces to second-order accuracy. Patterned Interface Reconstruction achieves this accuracy by modeling the interface with a planar pattern and a curved pattern. Each pattern is used to reproduce the volume fractions in the neighborhood of a target element, and an error is calculated. The error is the difference between the patterns' volume fraction prediction and the actual volume fractions. The pattern that produces the lowest error is locally chosen for further iteration. By choosing patterns locally rather than globally, materials that are composed of planar boundaries joining curved boundaries can be modeled by the appropriate local pattern.

Most interface reconstruction algorithms use material volume information in neighboring elements to reconstruct the interface in the central element. Patterned Interface Reconstruction also uses all the neighboring elements to compute an initial interface. The method then eliminates neighboring elements, which are judged to be part of a different interface of the material. This exclusion greatly improves the interface calculated near material corners and in thin material layers.

A comparison of a second-order and a lower-order interface reconstruction is illustrated in Figure 1.

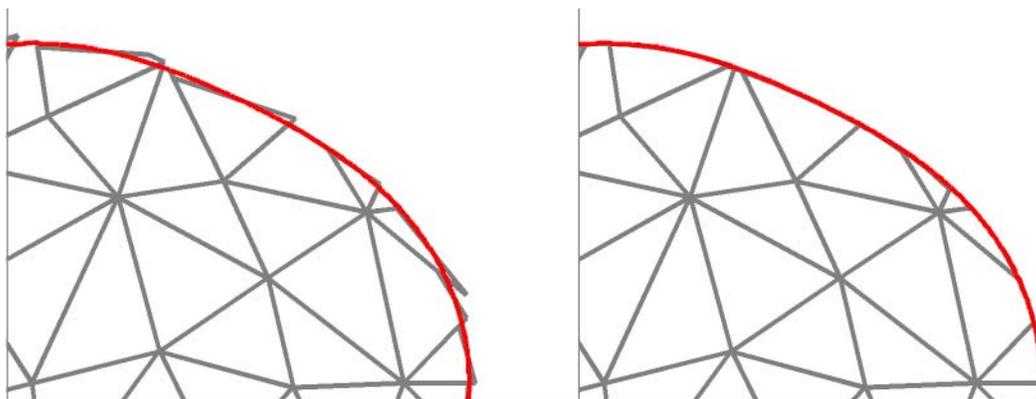


Figure 1. Low-order reconstructed interfaces (on the left) and second-order reconstructed interfaces (on the right) for a Cassini oval. The exact interface is the red line in both reconstructions.

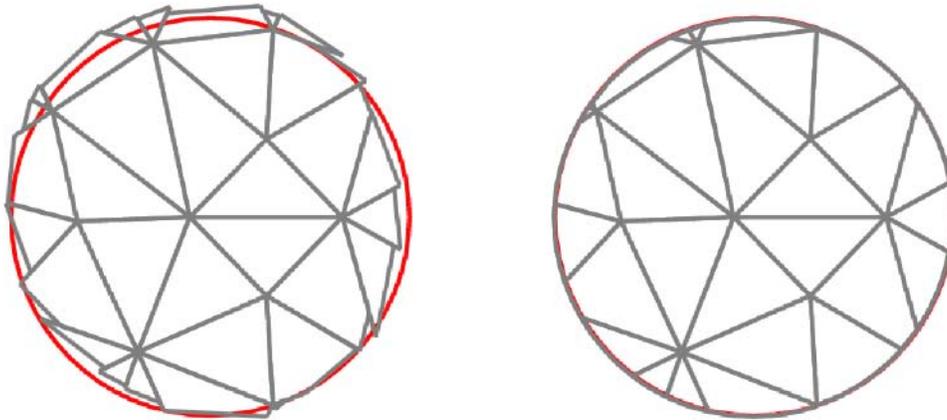
Using planar interfaces for a curved interface will lead to distortion of the body as it moves through the mesh. The order of accuracy of the reconstructed interface also plays a large role. To illustrate this distortion, Figure 2 shows a circle that moves through the mesh using planar, unsmoothed interfaces on the left and circular arc interfaces that are smoothed (Patterned Interface Reconstruction) on the right. The exact answer is included in each result as the red interface. The overall shape of the lower accuracy circular body is distorted. This is in contrast to the higher accuracy reconstruction where the difference between the reconstructed interfaces and the exact answer is barely visible.

The Patterned Interface Reconstruction algorithm will significantly improve the fidelity of multiple material simulations in ALEGRA and will reduce both the unphysical fragment emission and the unphysical distortion of bodies.

### References

1. Dolbow, J., Mosso, S., Robbins, J., et al. (2009). Coupling Volume-of-Fluid Based Interface Reconstruction with the Extended Finite Element Method, *Computer Methods in Applied Mechanics and Engineering*, Vol. **197**, pp 439-447.
2. Mosso, S., Garasi, C., and Drake, R. (2007). A Smoothed Two- and Three-Dimensional Interface Reconstruction Method, *Computing and Visualization in Science*, accepted for publication.

***This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the Army Research Laboratory.***



**Figure 2.** Planar interfaces are used (on the left) versus circular interfaces (on the right) to reconstruct a circle that has moved through a mesh. The exact result is shown as the red interface in both tests.

# NEW FRAMEWORK SUPPORTS ROBUST MULTI-PHYSICS COUPLING

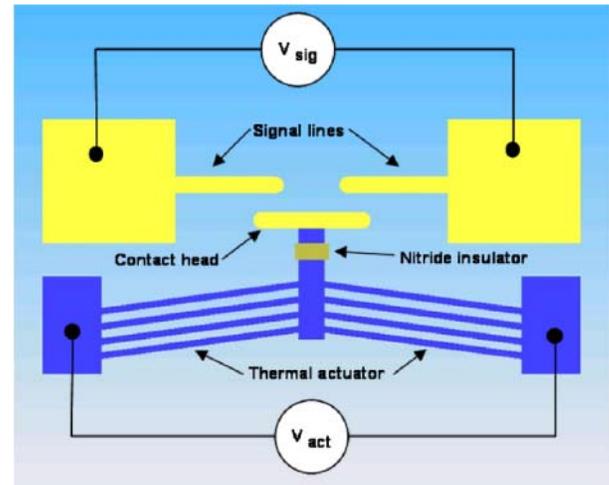
Contacts: Russell Hooper, Applied Math and Applications, [rhooper@sandia.gov](mailto:rhooper@sandia.gov)  
 Roger Pawlowski, Applied Math and Applications, [rppawlo@sandia.gov](mailto:rppawlo@sandia.gov)  
 Noel Belcourt, Computational Thermal and Fluid Mechanics, [kbelco@sandia.gov](mailto:kbelco@sandia.gov)  
 Matt Hopkins, Nanoscale and Reactive Processes, [mmhopki@sandia.gov](mailto:mmhopki@sandia.gov)

*It is now widely recognized that the key to advancing the forefront of modeling and simulation lies in leveraging existing and ongoing simulation capabilities within a multi-physics environment. Our work has focused on making this possible by developing a general-purpose multi-physics driver to enable robust and efficient coupling algorithms.*

Sandia has made, and continues to make, considerable investment in both hardware and software targeting high-fidelity, multi-physics modeling and simulation. This has produced a suite of specialized physics codes and associated algorithms. It is now widely recognized that the key to advancing the forefront of modeling and simulation lies in leveraging existing and ongoing simulation capabilities within a multi-physics environment. Our work has focused on making this possible by developing a general-purpose multi-physics driver to enable robust and efficient coupling algorithms. The multi-physics driver is designed to provide the lowest barrier to entry for existing applications to participate in a richer multi-physics simulation while taking a value-added approach to solution coupling algorithms. The value-added approach preserves any specialized solution algorithms within existing applications and leverages them within the broader coupling solution algorithm.

Our early work explored the feasibility of improving robustness of coupling algorithms using only information obtainable from each application standalone. Using prototype applications from the literature, we demonstrated that limitations inherent in traditional approaches to coupling based on fixed-point iterations can be overcome and robustness dramatically improved using our approach to coupling. We accomplished this by incorporating application data used in fixed-point iterations into a Newton-based algorithm applied to the entire coupled system. This enables solution convergence for coupled problems where previous fixed-point algorithms fail to converge. This approach also achieves a consistent second-order rate of convergence, in contrast to first-order rates for fixed-point iterations, resulting in significantly fewer required iterations to obtain practical solutions.

We designed our multi-physics coupling driver as an extension to the Trilinos solver framework. This provides immediate access to advanced linear solvers,

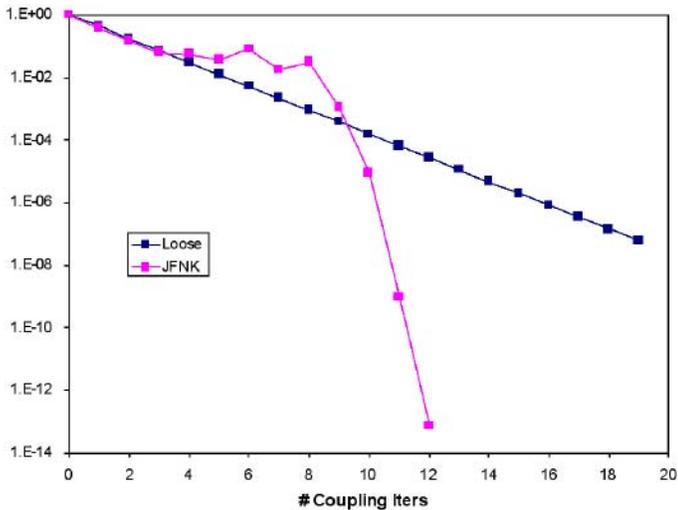


**Figure 1.** A micro-electrical-mechanical system used as a thermal switch. The system is modeled as three applications comprising coupled voltage, temperature, and material displacement fields within the same physical domain.

preconditioners, nonlinear solvers and efficient data management services tuned for performance on massively parallel computing architectures. This design also provides a composite view of the entire coupled system which enables use of more sophisticated algorithms in Trilinos such as optimization, parameter continuation, and higher-order explicit and implicit time integration, all of which would be precluded or severely limited using more traditional coupling algorithms based on fixed-point iterations.

We have adapted our multi-physics driver to work within the SIERRA applications framework to drive real-world coupling scenarios important to Sandia. As an example, Figure 1 depicts a micro-electrical-mechanical system. The device is an electro-thermo-mechanical actuator or switch that utilizes a silicon carbide (SiC) material for fast response. Operation of the switch involves application of a voltage that heats the SiC to produce a thermal expansion, which eventually contacts the shuttle with the plate to close

the switch. The system is modeled as three applications comprising coupled voltage, temperature, and material displacement fields within the same physical domain. A comparison of convergence behavior between a standard fixed-point coupling method and our Newton-based approach is shown in Figure 2. The improved order of convergence characteristic of Newton's method is observed and represents a more robust algorithm. Not shown in the figure is the convergence of our Newton-based algorithm for increased voltage loadings at which the fixed-point algorithm fails to converge.



**Figure 2.** A comparison of convergence using a traditional coupling based on fixed-point (Loose) iterations and our more robust Newton-based (JFNK) approach. Convergence to solution is measured by the reduction in the overall norm of the residual equations for the coupled system.

The multi-physics driver is currently impacting other key efforts at Sandia including an LDRD targeting next-generation burner reactor safety modeling and another LDRD targeting multiphase, multicomponent thermal flow and reactive transport coupled to nonlinear geomechanics in heterogeneous (geologic) porous materials.

The former involves use cases not originally supported in our multi-physics coupling driver but which have since been incorporated. The latter is a recent effort that has identified the need to support various forms of operator splitting for efficient time stepping in the multi-physics driver.

In summary, we have developed an algorithmic driver for multi-physics coupling built upon the Trilinos solver framework. Our design provides the lowest possible barrier for applications to participate in coupled multi-physics simulations while making the best use of any functionality applications can provide. By providing a composite view of coupled systems, our driver also enables more sophisticated algorithms such as optimization, adaptivity and parameter continuation not possible with traditional approaches to coupling.

### Reference

Hooper, R., Hopkins, M., Pawlowski, et al. (2008). *Final Report on LDRD Project: Coupling Strategies for Multi-Physics Applications*, SAND2007-7146, Sandia National Laboratories, Albuquerque, NM.

*This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

# SCALABLE SOLUTION TECHNIQUES DEVELOPED TO ADDRESS DAMAGE TO SEMICONDUCTOR DEVICES

Contacts: Paul T. Lin, Electrical and Microsystems Modeling, [ptlin@sandia.gov](mailto:ptlin@sandia.gov)  
 John N. Shadid, Electrical and Microsystems Modeling, [jnsahdi@sandia.gov](mailto:jnsahdi@sandia.gov)  
 Robert J. Hoekstra, Electrical and Microsystems Modeling, [rjhoeks@sandia.gov](mailto:rjhoeks@sandia.gov)

*The performance of electronics in hostile environments is of critical importance to the weapons program ... researchers are developing scalable solution methods for simulation of radiation damage to semiconductor devices ... these solution techniques will also be of interest to the satellite community due to concerns of damage from cosmic rays and other radiation.*

The performance of electronics in hostile environments is of critical importance to the weapons program. Sandia scientists and computational methods researchers are developing scalable solution methods for simulation of radiation damage to semiconductor devices. These solution techniques are based on robust Newton-Krylov solvers and approximate block factorization and physics-based preconditioners.

A science-based engineering program has been developed to conduct numerical simulation with extensive validation with experiments. The performance of the semiconductor devices in a circuit after irradiation is modeled by a massively parallel finite element fully-implicit Newton-Krylov simulation code. The Newton method solution approach generates large linear systems that are solved by Krylov methods. The choice of the preconditioner is critical to the parallel scaling and reducing the solution time for these linear systems. Transient simulations involve 36 species for a total of 39 unknowns per mesh point. For realistic simulations, these linear systems can be of the order of 100 to 1000 million rows. It is worth noting that these solution techniques will also be of interest to the satellite community due to concerns of damage from cosmic rays and other radiation.

The SNL research team has had extensive experience with developing and evaluating one-level and multi-grid preconditioners for a wide variety of coupled multi-physics applications. These include drift-diffusion equations [1] and general transport/reaction systems [2]. The ML multilevel preconditioner package is part of the Trilinos solver project. Figure 1 shows a weak scaling study on the SNL Red Storm Cray XT3/4 machine for a bipolar junction transistor (BJT) comparing the linear solve time for a one-level incomplete lower/upper (ILU) factorization (red line) with an unsmoothed aggregation ML (green line) and a Petrov-Galerkin smoothed aggregation ML preconditioner

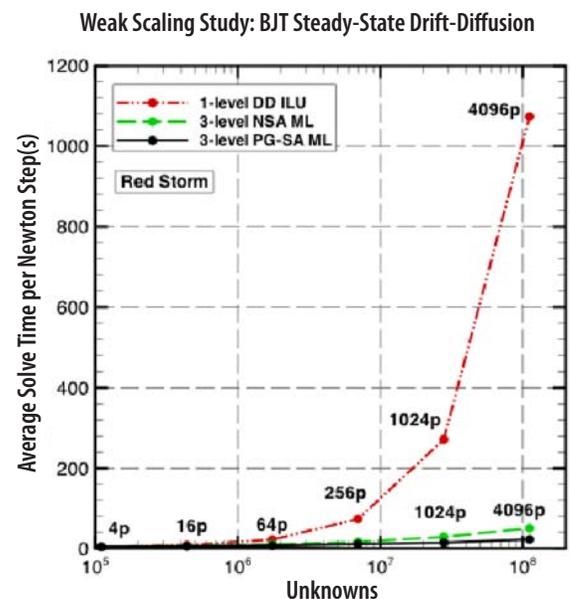


Figure 1. Comparison of one-level and multi-grid preconditioner.

(black line). The steady-state problem is scaled up to 4000 processors and 100 million unknowns, and for the largest problem, the multilevel approach is almost 50 times faster than the one-level preconditioner. The steady-state electric potential for the BJT is presented in Figure 2.

Since it is difficult to develop scalable multilevel preconditioners for mixed parabolic-elliptic systems, this research team has pursued the development of approximate block factorization and physics-based preconditioners. The idea behind these types of preconditioners is to exploit knowledge of the block structure and the physics of the system to produce a more effective preconditioner. Decompositions of the Jacobian matrix are formed to produce sub-block system solves that are more amenable to optimal multilevel methods. This work is complementary

with the multi-grid work mentioned above as the latter is now used as a sub-block iteration. Figure 3 presents a weak scaling study comparing the one-level ILU, three-level multi-grid, and a physics-based preconditioner. The latter introduced physics into the preconditioner by using the relation for the Quasi-Fermi potential. Preliminary results show that the physics-based preconditioner is competitive with the multi-grid preconditioner and has the advantage that less expensive smoothers can be used. More advanced physics-based preconditioners are currently under development, as well as preconditioners to handle the extra defect species due to displacement damage from irradiation.

### References

1. Lin, P. T., Shadid, J. N., Sala, M., et al. (2009). Performance of a Parallel Algebraic Multilevel Preconditioner for Stabilized Finite Element Semiconductor Device Modeling, submitted to *Journal of Computational Physics*.

2. Lin, P. T., Sala, M., Shadid, J. N., et al. (2006). Performance of Fully-Coupled Algebraic Multilevel Domain Decomposition Preconditioners for Incompressible Flow and Transport, *International Journal for Numerical Methods in Engineering*, Vol. 67, No. 2, pp. 208-225.

*This work has been supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

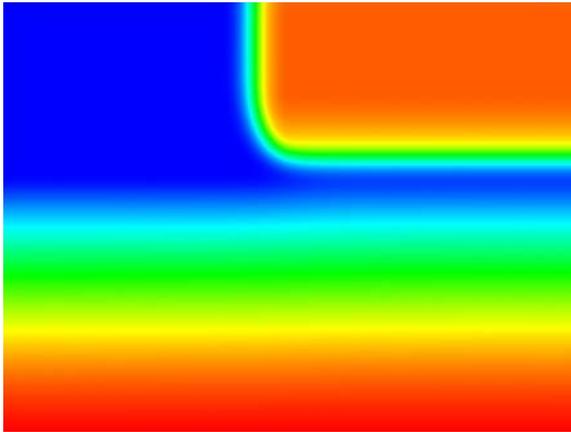


Figure 2. Steady-state electric potential solution for BJT.

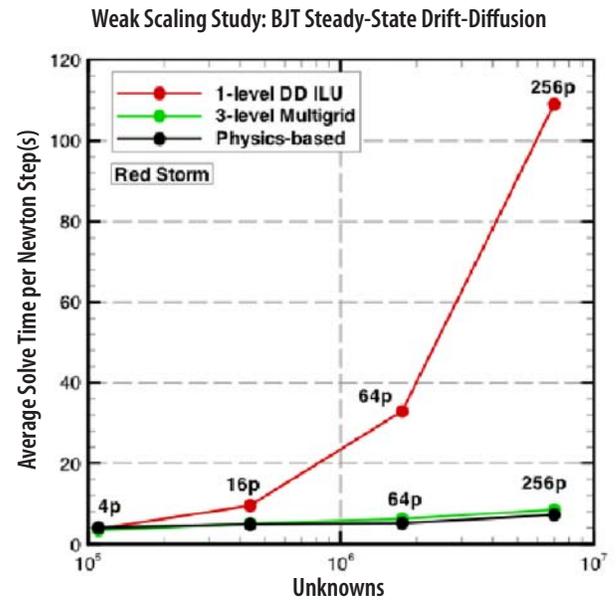


Figure 3. Comparison of one-level, multi-grid, and physics-based preconditioners.

# MODELING MULTI-MATERIAL SYSTEMS WITH THE EXTENDED FINITE ELEMENT METHOD

Contacts: T.E. Voth, Exploratory Simulation Technologies, [tevoth@sandia.gov](mailto:tevoth@sandia.gov)  
 Stewart Mosso, Computational Shock and Multi-physics, [sjmosso@sandia.gov](mailto:sjmosso@sandia.gov)  
 J. Robbins, Multi-Scale Dynamics Materials Modeling, [jrobbin@sandia.gov](mailto:jrobbin@sandia.gov)  
 E. Strack, Computational Shock and Multi-physics, [oestrack@sandia.gov](mailto:oestrack@sandia.gov)

*These algorithms provide increased physics fidelity for a range of Sandia's ALE missions such as the evaluation of armor/anti-armor effectiveness and weapons performance.*

Explicit Eulerian and Arbitrary Lagrangian Eulerian (ALE) codes are often used to model large deformation mechanics where large vortices make a Lagrangian approach problematic. Whereas Lagrangian Finite Element codes can maintain distinct material interfaces at element boundaries, this is not true for ALE simulations. With respect to the discussions that follow, an ALE algorithm advances a solution in a two-step (operator-split) process. First, the standard Lagrangian-step is taken. At this point, the data are moved (remapped) in a mass and energy conserving way to an “improved” mesh. Material interfaces that may have corresponded to element boundaries now lie in their interiors. These intra-element material boundaries are frequently treated via *ad hoc* models [1]. Various models have been used but generally fail as they (i) impose a single velocity field on the mixed element and (ii) model interface physics in a volumetric rather than surface-phenomenological way. In the first case, bad material states may be reached as the materials share a common deformation. In the second case, spurious material bonding may occur. Here we describe advancement in operator-split ALE technology that uses the eXtended Finite Element Method (X-FEM) to i) provide multiple kinematic fields per multi-material element and ii) define surfaces across which the materials can interact. These algorithms provide increased physics fidelity for a range of Sandia's ALE missions such as the evaluation of armor/anti-armor effectiveness and weapons performance.

The X-FEM provides a methodology to enhance standard Finite Element bases while retaining their underlying convergence properties [2]. Here we allow for discontinuities in the displacement fields via Heaviside enrichment functions aligned with material boundaries. In this way, each material in a multi-material element has its own kinematic field. Additionally, a discontinuity surface results through which the materials interact (e.g., via a Coulomb friction model). The Lagrangian portion of the ALE X-FEM algorithm draws from the explicit dynamics X-FEM literature (e.g., [2]). Specifically, it has been shown that a multi-material element may be decomposed into multiple single-material elements,

one per material. This decomposition is demonstrated in Figure 1 for a two-material element. In the illustration, the multi-material X-FEM element is equivalent to two partially-filled single-material elements. These single-material elements are found in operator-split ALE codes. With the proper choice of mass matrix, this formulation has a number of beneficial features, including (i) a stable time-step that depends on the parent element length scale and (ii) use of much of the existing Lagrangian-step framework of the host ALE code. This formulation is inspired by Reference 1 but also differs in several ways. Here, we enforce interface constraints on intra-element discontinuities via Lagrange multipliers. In particular, we provide for intra-element interface constraints using the Forward Increment Lagrange Multiplier approach [3]. A beneficial feature of this technique is that the stable time-step is not modified relative to the unconstrained case. These interfaces are constructed via the second-order accurate Patterned Interface Reconstruction (PIR) algorithm that greatly reduces element-to-element interface segment discontinuities, relative to other schemes. The final ingredient is an unsplit remap that serves to move field data from the end of one X-FEM Lagrange step to the start of a new step.

An example of the method's efficacy is illustrated in Figure 2. Here a leftward-moving cylinder impacts a pair of plates. Materials are inserted into an empty domain (blue) and the X-FEM algorithm generates single-material elements and intra-element surfaces. As the solution advances, searches for intra-element interface interactions are made, and frictionless constraints are enforced. The center figure demonstrates that cylinder and plate rebound after impact, while the right-hand image illustrates spurious bonding typical of a non-X-FEM simulation.

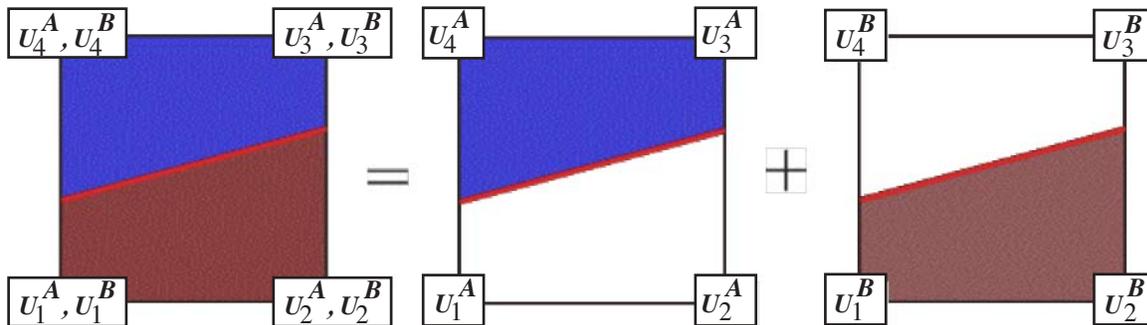
## References

1. Vitali, E., and Benson, D. J. (2008). Contact with Friction in Multi-Material Arbitrary Lagrangian-Eulerian Formulations Using X-FEM, *International Journal of Numerical Methods in Engineering*, Vol. **76**, pp. 893-921.

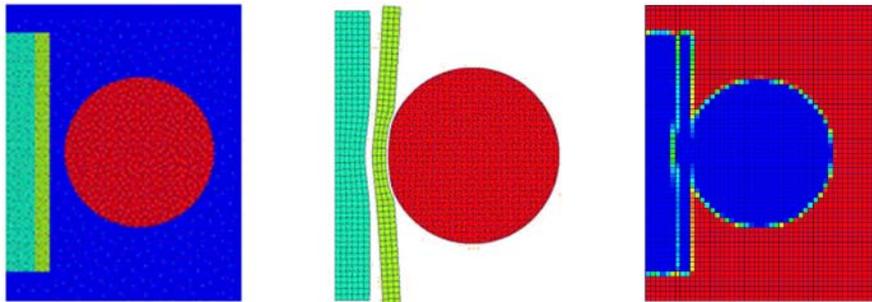
2. Song, J-H, Areias, P. M. A, and Belytschko, T. (2006). A Method for Dynamic Crack and Shear Band Propagation with Phantom Nodes, *International Journal of Numerical Methods in Engineering*, Vol. **67**, pp. 868-893.

*This work is being supported by SNL's Laboratory Directed Research and Development (LDRD) Program, the Computer Science Research Foundation, and the Army Research Laboratory.*

3. Carpenter, N., Taylor, R., and Katona, M. (1991). Lagrange Constraints for Transient Finite Element Surface Contact, *International Journal of Numerical Methods in Engineering*, Vol. **32**, pp. 103-128.



**Figure 1.** A Heaviside eXtended Finite Element enrichment of a multiple material element is identical to the composition of two single material elements. These elements are available in most operator-split ALE codes.



**Figure 2.** Right-hand-side shows two X-FEM plates and cylinder inserted in an empty domain. The cylinder is given a left-ward velocity and the center image shows its rebound. The right-most image demonstrates cylinder-plate bonding for a non-X-FEM simulation.

# TOLERANT GEOMETRY TOOLS MAKE POSSIBLE THE GENERATION OF CONTIGUOUS MESHES FOR SLOPPY ASSEMBLY MODELS

Contacts: Brett Clark, Computational Modeling Science, [bwclark@sandia.gov](mailto:bwclark@sandia.gov)  
Byron Hanks, Computational Modeling Science, [bwhanks@sandia.gov](mailto:bwhanks@sandia.gov)

*Geometric assemblies used as the basis for modeling and simulation, often containing sloppy features, can now be more efficiently meshed using tools.*

Geometric assemblies used as the basis for modeling and simulation, often containing sloppy features, can now be more efficiently meshed using tools developed by the CUBIT Geometry and Meshing team at Sandia. Sandia regularly performs structural and thermal analyses on large nuclear weapon assembly models containing hundreds of components (Figure 1). These analyses require a discretized version of the model called a “mesh.” The Computational Modeling Sciences Department at Sandia focuses on developing software tools that automatically generate these meshes. Generating meshes is not a trivial task and has for many years been one of the main bottlenecks in the computational analysis process. The capabilities described below address a major meshing time sink by greatly reducing the amount of time required to generate meshes for large assemblies containing “slop.”

Slop is defined as significant gaps, overlaps, and misalignments between assembly components. There are several reasons for design slop. Poor modeling practices may introduce slop into a design. A mismatch in modeling tolerances between the original CAD system and the meshing tool reading the input can also create such slop. Sometimes the models are accurate, but details such as press fit tolerances and adhesive layers must be removed for the analyses desired.

Meshes of assembly models are generally required to be contiguous across component-component interfaces. This means that two components that are in contact with one another in an assembly will share a mesh at the surface(s) of contact. Generating such a mesh generally requires the CAD model to share surfaces between contacting parts. CAD models typically do not use this representation (often referred to as a “non-manifold” representation), and so it is necessary for the mesh generation package to generate such a shared topology between parts before meshing.

Automatically recognizing and creating non-manifold connections between neighboring components in an assembly first requires a determination of which components

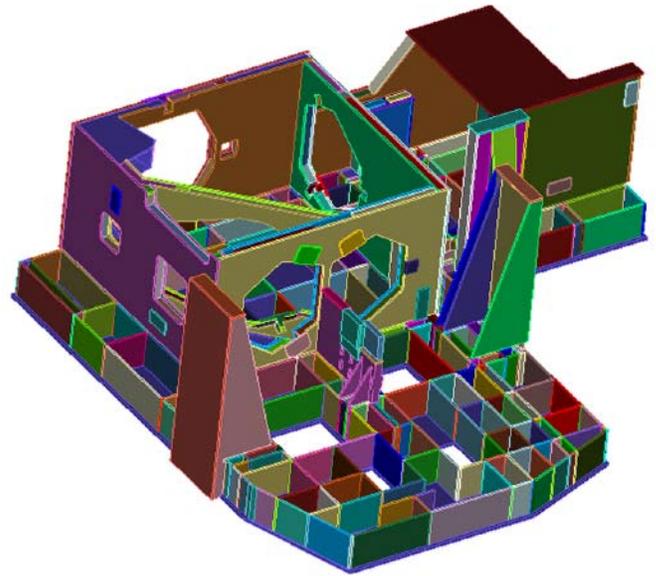


Figure 1. Example of CAD assembly model with hundreds of components.

should be touching using proximity checks. If there are large gaps or misalignments between components it can be difficult to determine if the design intent is for two components to be touching or not. This ambiguity, when trying to define non manifold connections between components generally complicates the process for automation and requires time-consuming manual intervention.

The Computational Modeling Sciences department has developed powerful tools for generating non manifold representations of CAD assembly models. These tools are particularly helpful in the cases where there is significant slop in the CAD model that would normally require hours of manual geometry modification to fix. The improvements include tools to help the user identify the slop in the assembly, algorithms to automatically and tolerantly enforce criteria for creating non manifold connections, and tools to analyze the results and identify remaining issues (Figure 2).

**Reference**

Clark, B. W., Hanks, B. W., Ernst, C. (2008). Conformal Assembly Meshing with Tolerant Imprinting, *Proceedings 17th International Meshing Roundtable*, Part 4, pp. 267-280.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

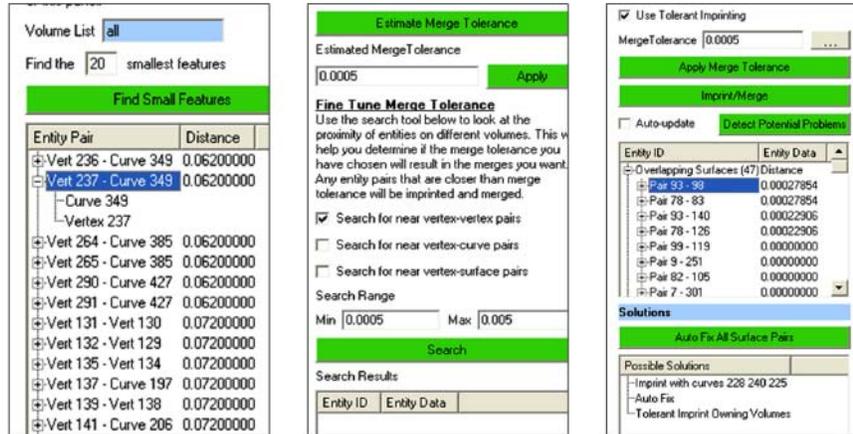


Figure 2. Snapshots of new tools for defining non-manifold connections. The left panel shows close proximities within the model that help determine the smallest allowable feature size. The center panel shows a tool for determining the largest allowable gap in the model. The right panel shows a tool for fixing residual problems after generating nonmanifold connections in the model.

# AUTOMATED TOPOLOGY ADJUSTMENTS FOR SWEEPING ENABLES MORE ACCURATE SIMULATION OF COMPLEX GEOMETRY

Contacts: Brett Clark, Computational Modeling Science, [bwclark@sandia.gov](mailto:bwclark@sandia.gov)  
Byron Hanks, Computational Modeling Science, [bwhanks@sandia.gov](mailto:bwhanks@sandia.gov)

*The automation of CAD topology modification for sweepability greatly reduces the user interaction required in model generation.*

The ability to automatically mesh a complex geometric model with hexahedral elements has been significantly enhanced by new tools developed at Sandia. The Computational Modeling Sciences Department provides tools for automating model generation for Sandia's modeling and simulation needs. Sandia is a global leader in meshing technology, specifically in all-quadrilateral and all-hexahedral meshing. In recent years, efforts have also been centered on automatically modifying CAD models in preparation for meshing.

The CAD model used as input to the mesh generator is generally not designed with meshing in mind. As a result, the original CAD model does not lend itself well to many meshing algorithms. Although an issue for all algorithms, the complexity of generating an all-hexahedral mesh exacerbates the problem. Hexahedral meshing algorithms often require rigid topological constraints on the CAD model being meshed. Consequently, much of the model preparation time is spent modifying the CAD model to prepare it for hexahedral meshing.

This problem is commonly encountered when using the sweeping all-hex meshing algorithm. Sweeping is an algorithm that generates hex elements by sweeping a quadrilateral surface mesh along a trajectory defined by the CAD model. This meshing approach has proven very successful in generating all-hex meshes but relies on the CAD

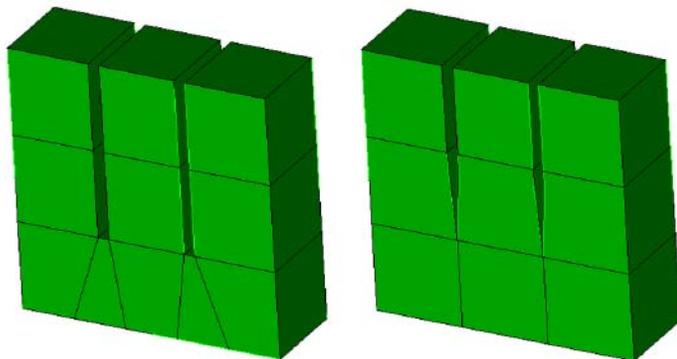


Figure 1. Automatically modify narrow surfaces to improve mesh quality.

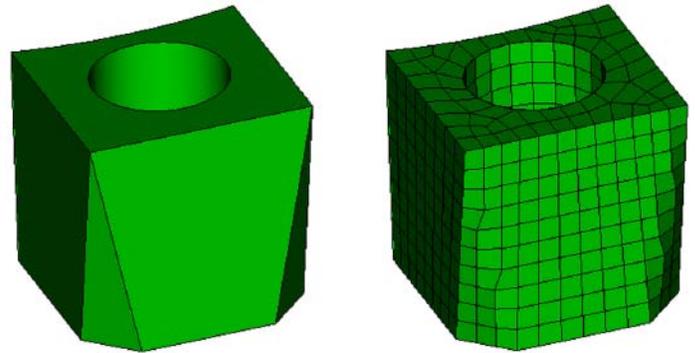


Figure 2. Triangular surfaces are automatically combined with neighbors.

model having a topology that readily admits a high-quality, structured mesh along the sweep direction. The user is often required to significantly modify or decompose the geometry to develop topology on the CAD model that is "sweepable."

Recent efforts have produced tools improving the automation of topology transformations to make a CAD model sweepable and/or improve the resulting mesh quality. An example of the automation is shown in the figures. In Figure 1, narrow surfaces are automatically found and modified to produce a CAD model topology that will result in better element quality.

Figure 2 shows triangular surfaces that prohibit a sweepable topology. The topology is transformed by combining surfaces to admit a structured quad mesh along the sweep direction. The automation of CAD topology modification for sweepability greatly reduces the user interaction required in model generation.

Sandia's geometry and meshing technology is also having an impact at the Goodyear Tire and Rubber Company. Goodyear is focused on improving the traction and treadwear of its products. High-fidelity models of the tire are used in computer simulations (Figure 3) to predict and understand the tire behavior early in the design process. These technologies provide Goodyear a competitive advantage in bringing new products to market in less time.

**Reference**

Owen, S. J., Clark, B. W., Melander, D. J., Brewer, M., Shepherd, J. F., Merkley, K., Ernst, C., and Morris, R. (2007). An Immersive Topology Environment for Meshing, *Proceedings of the 16th International Meshing Roundtable*, Springer-Verlag, pp. 553-577.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program as well as Work for Others agreements, including the Goodyear Tire and Rubber Company.*

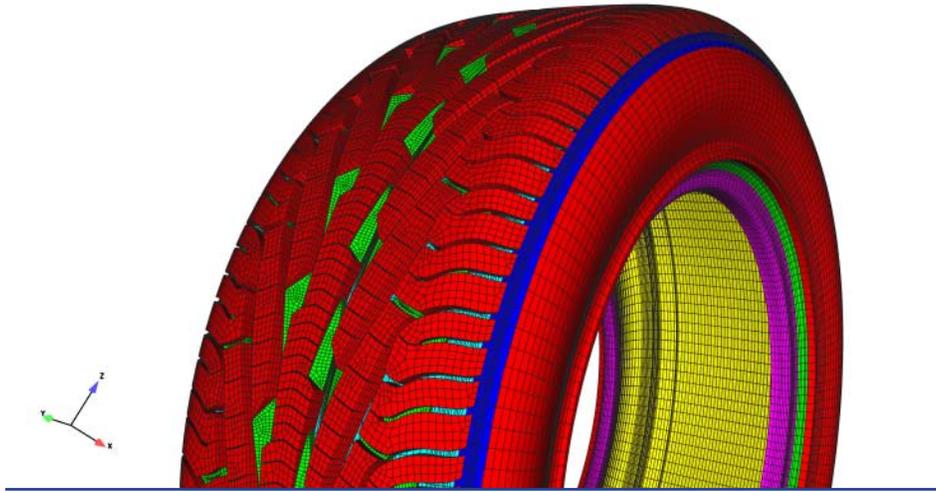


Figure 3. The Goodyear Tire and Rubber Company uses Sandia's geometry and meshing technology to improve the performance of its tires.

# ADVANCES IN MESH OPTIMIZATION CREATE MORE ROBUST AND ACCURATE SIMULATIONS

Contact: Patrick M. Knupp, Applied Mat and Applications, [pknupp@sandia.gov](mailto:pknupp@sandia.gov)

*Among the many results of this research, we discovered a method that smoothes/optimizes unstructured meshes without creating inverted elements.*

Mesh optimization methods improve the accuracy and efficiency of PDE simulations. Among the many results of this research, we discovered a method that smoothes/optimizes unstructured meshes without creating inverted elements. This allows more accurate and efficient partial differential equation simulations, along with a reduction in the time to generate the mesh.

Meshes are used in almost every Modeling and Simulation problem that requires numerical solutions to partial differential equations (PDEs). Researchers perform PDE simulations in studies of energy generation (fusion, combustion), waste disposal and environment (ground-water, climate), basic scientific discovery (electromagnetics, astrophysics), and engineering (fluids, mechanics, accelerator design). The accuracy and efficiency of PDE-based simulations are impacted by properties of the mesh such as element size, shape, orientation, and smoothness. Mesh optimization is a rigorous method for controlling mesh properties via an objective function that defines “mesh quality” in terms of the coordinates of mesh vertices and/or mesh topology. Frequently occurring problems in mesh optimization include (1) the elimination of inverted elements in a mesh so that the simulation can proceed, (2) the improvement of meshes resulting from mesh generation algorithms on complex geometries, (3) the updating of meshes on domains that deform in time or in geometric parameter space (Figure 1), (4) the improvement of Lagrange meshes in a Eulerian-Lagrangian rezone algorithm, and (5) mesh optimization for solution adaptivity.

Ten years ago, very little was known about mesh optimization for unstructured meshes. This has changed dramatically, in part due to funding from ASCR, permitting researchers to study this topic. Before the present research was performed, the main option for improving unstructured mesh quality was Laplace smoothing. The Laplace method often creates inverted elements that generally cannot be used in solving partial differential equations. New barrier metrics prevent the creation of inverted elements to enable faster creation of meshes appropriately tuned to particular PDE simulations. Moreover, the metrics can control shape, size, and orientation of mesh elements and can be related to discretization error estimates—making more accurate simulations possible.

PDE simulations involving time-evolving geometries (shape optimization, metals forging, and/or arbitrary Eulerian-Lagrangian methods) are an important example of our mesh-optimization capabilities. Our new method allows such calculations to proceed much further in time by delaying the onset of mesh inversion. This has enabled, for example, the ALEGRA MHD code to simulate Z-pinch experiments via the ALE operator splitting option, which involves a mesh rezoning step based on mesh optimization.

The new barrier methods for optimization of unstructured mesh quality are available via the MESQUITE mesh quality-improvement toolkit (see <http://www.cs.sandia.gov/optimization/knupp/Mesquite.html>). In general, Mesquite serves as the delivery mechanism for the results of our research on mesh optimization. Mesquite uses

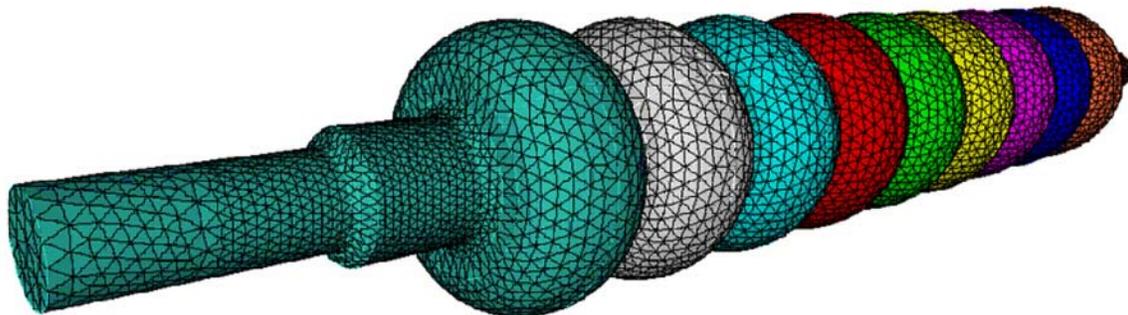


Figure 1. An accelerator shape-optimization study using deforming meshes.

the Interoperable Technologies for Advanced Petascale Simulations (ITAPS) mesh interface and includes capabilities for synchronous mesh optimization on parallel machines.

Important requirements for a successful general-purpose mesh optimization capability include (1) the ability to automatically construct effective weighting functions for local mesh control, (2) the ability to uniformly treat most mesh types (e.g., tetrahedral, hexahedral, surface, volume), (3) numerical optimization with an acceptable computational cost, and (4) the hiding of complexity from users. Researchers are studying all four of these requirements. For example, the first requirement is being addressed via algorithms for automatically generating local descriptions of mesh quality based on application input; the descriptions are used in conjunction with a set of canonical quality metrics to control shape, size, or shape-size-orientation. The second requirement is addressed via direct optimization methods based on mesh edge-lengths, angles, and volumes. Sample points within each element permit uniform treatment of both simplicial and non-simplicial meshes, as well as linear and higher-order elements. Direct optimization is also a promising approach for controlling the mesh on the domain boundary. The use of general-purpose numerical optimization solvers has been shown to result in disappointing performance, so solvers tailored specifically to the mesh optimization problem are being studied.

### **References**

Knupp, P. (2008). Updating Meshes on Deforming Domains, *Comm. Num. Meth. Engr.*, Vol. **24**, pp. 467-476.

Hetmaniuk, U., and Knupp, P. (2008). A Mesh Optimization Algorithm to Decrease the Maximum Error in Finite Element Computations, *Proceedings of the 17th International Meshing Roundtable*, Springer, pp. 533-550.

Shontz, S., and Knupp, P. (2008). Effect of Vertex Reordering on 2D Local Mesh Optimization Efficiency, *Proceedings of the 17th International Meshing Roundtable*, Springer, pp. 107-124.

Freitag L. A., and Knupp, P. (2002). Tetrahedral Mesh Improvement via Optimization of the Element Condition Number, *International Journal for Numerical Methods in Engineering*, Vol. **53**, No. 6, pp. 1377-1391.

***This work has been supported by the Department of Energy's Advanced Scientific Computing Research (ASCR) Program.***

# GEOMETRIC CAPTURE METHODS PROVIDE NEW TOOLS FOR HEXAHEDRAL MESH GENERATION

Contacts: Steven Owen, Computational Modeling Science, [sjowen@sandia.gov](mailto:sjowen@sandia.gov)  
Jason Shepherd, Data Analysis and Visualization, [jfsheph@sandia.gov](mailto:jfsheph@sandia.gov)

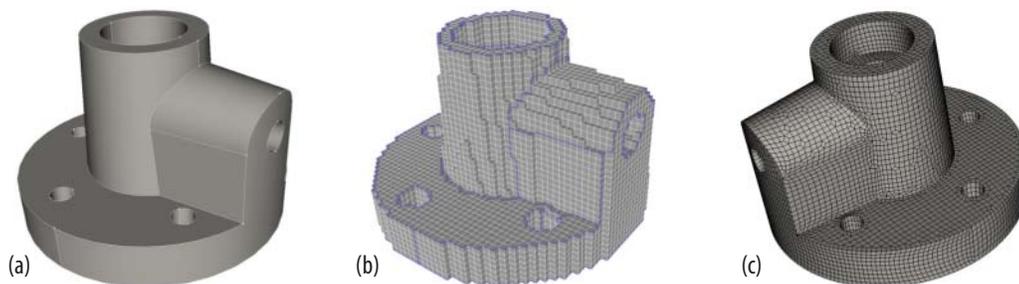
*An automatic general-purpose all-hex meshing procedure would have tremendous impact on modeling and simulation at Sandia and throughout the Nuclear Weapons Complex by enabling larger and more complex simulations in fractions of the time currently required.*

Recent work has contributed both theoretical understanding and practical tools to the all-hex meshing problem [1]. Automatic Mesh Generation is a key enabling component of many numerical methods used in computational simulation. Numerical accuracy requirements for many applications used within Sandia and the Nuclear Weapons Complex drive the need for meshes composed of hexahedral shaped elements versus tetrahedral or mixed element meshes. While theory and practice of tetrahedral mesh generation is very well understood, there remain significant challenges for hexahedral mesh generation.

One approach for hex meshing that is currently being studied is to begin with a nonconforming bounding grid that roughly approximates the geometric domain. Figure 1(a) shows an initial geometry to be meshed and 1(b) shows a non-conforming grid that approximates the boundary. This work has shown that by applying new tools, methods, and theory developed at Sandia, a geometry-conforming mesh can be established. Figure 1(b) also shows how the topology of the solid model has been embedded or *captured* from the non-conforming grid using tools developed in this work. Figure 1(c) shows the final mesh following mesh transformation operations, projection and smoothing procedures. It has been shown that through application of these tools, a hexahedral mesh suitable for numerical analysis can be generated that would be difficult or impossible to generate using traditional hexahedral meshing tools such as pave and sweep.

While similar methods for generating a hexahedral mesh, sometimes known as grid-based or overlay methods, have been studied in the past, they have been primarily based on heuristic algorithms that can often fail, given complex or diverse geometric requirements. While maintaining the objective of a practical implementation, this work has also focused on more theoretical aspects of the problem. For example, the approach taken involves transforming a non-conforming all-hex mesh into a geometry-conforming mesh with quality suitable for analysis. The mathematical proof that such a transformation or set of transformations exists is a fundamental question and the foundation upon which a practical application can be developed. Through joint work with the French Commissariat à l'Énergie Atomique (CEA)/DAM, a theoretical study [2,3] has recently been completed that provides a proof of hexahedral transformation. Essentially, it establishes that there exists a set of transformations that will convert one hexahedral mesh into another through a series of operations called sheet insertions and sheet extractions.

A sheet can be thought of as a continuous layer of hexahedral elements as shown in Figure 2(a). The sheet extraction removes a sheet (layer of hexes) from the mesh with neighboring elements expanding to close the gap to maintain a conformal mesh as shown in Figures 2(b) and 2(c). The insertion operation is an inverse procedure, where a layer of elements is inserted into the mesh. Figure 3 shows one example of sheet insertion using an algorithm called



**Figure 1.** Hex meshing procedure: (a). Original CAD model to be meshed. (b). Non-conforming grid overlays approximate the original CAD model. CAD topology is embedded in the grid to capture vertices and curves. (c) Mesh transformation, projection and smoothing procedures are executed to establish the final geometry-conforming hexahedral mesh.

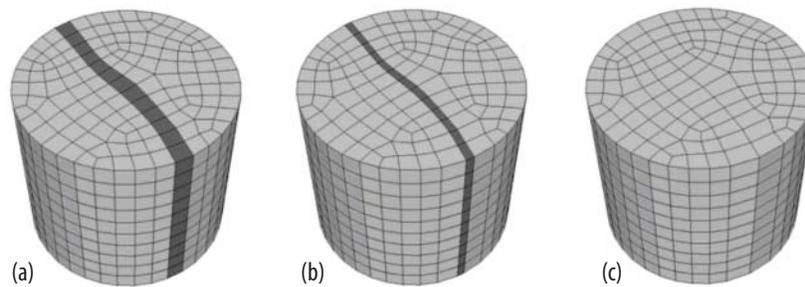
pillowing, where the dark set of hexes in 3(a) are “shrunk” to create a gap from its neighbors and a new layer of hexes are inserted to fill the gap.

Work is ongoing in this area to improve and automate these procedures to develop a practical method for all-hex mesh generation suitable for modeling and simulation. Current methods for hexahedral mesh generation are often expertise intensive and time consuming. An automatic general-purpose all-hex meshing procedure would have tremendous impact on modeling and simulation at Sandia and throughout the Nuclear Weapons Complex by enabling larger and more complex simulations in fractions of the time currently required.

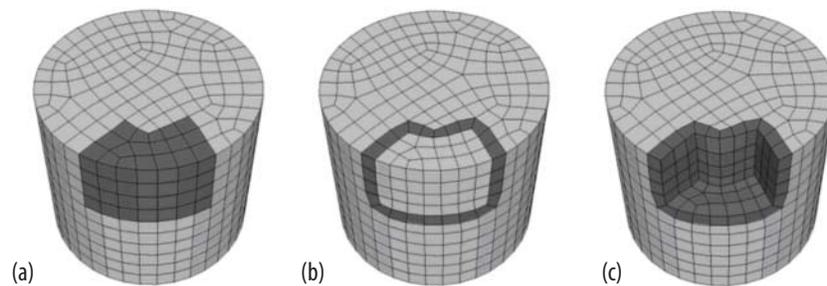
### References

1. Shepherd, J. F. (2007). Topologic and Geometric Constraint-Based Hexahedral Mesh Generation, published Doctoral Dissertation, University of Utah.
2. Ledoux, F., and Shepherd, J. (2008). Topological and Geometrical Properties of Hexahedral Meshes, submitted to *Engineering with Computers*.
3. Ledoux, F., and Shepherd, J. (2008). Topological Modifications of Hexahedral Meshes Via Sheet Operations: A Theoretical Study, submitted to *Engineering with Computers*.

***This work has been supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program.***



**Figure 2.** Sheet Extraction: (a) a sheet (layer of hexahedra) is selected. (b) The edges that define the sheet are collapsed. (c) The sheet is entirely removed from the mesh.



**Figure 3.** Sheet Insertion – *Pillowing*: (a) A “shrink” set is defined. (b) The shrink set is separated from the rest of the mesh and a sheet is inserted to fill the gap. (c) The newly inserted pillow sheet.

# NEW ALGEBRAIC MULTI-GRID CAPABILITIES MODEL STRETCHED MESHES AND ANISOTROPIC PROBLEMS

Contacts: Ray Tuminaro, Scalable Algorithms, [rstumin@sandia.gov](mailto:rstumin@sandia.gov)  
Jonathan Hu, Scalable Algorithms, [jhu@sandia.gov](mailto:jhu@sandia.gov)

*We have developed new algorithms to solve linear systems... Given the often dominant linear solve cost, these algorithm improvements have in some cases reduced overall simulation time by nearly a factor of 5x.*

We have developed new algorithms to solve linear systems arising from discretization of partial differential equations (PDEs) with strong anisotropic behavior. These new capabilities have enabled us to drastically reduce the linear solve time within some applications. Given the often dominant linear solve cost, these algorithm improvements have in some cases reduced overall simulation time by nearly a factor of 5x.

Many realistic engineering problems contain anisotropic aspects. That is, the underlying behavior is significantly different in one direction than in other directions. Some examples include strongly connective flow, boundary layers, and anisotropic materials. It also occurs when isotropic applications are discretized on stretched meshes. Unfortunately, anisotropic problems can be difficult to solve from a linear systems perspective.

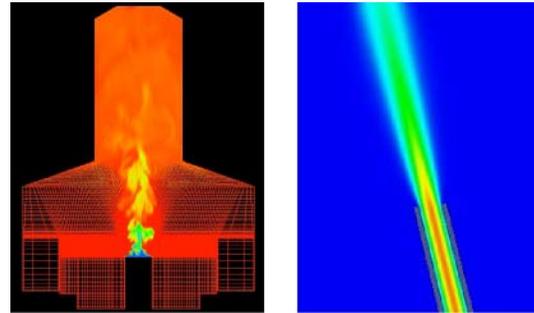
**Table 1.** Iterations/run time for a representative linear solve from a Helium plume calculation.

|            | isotropic coarsening | anisotropic coarsening |
|------------|----------------------|------------------------|
| its        | 161                  | 44                     |
| time (sec) | 17.22                | 4.86                   |

**Table 2.** Iterations/run time for a representative linear solve from an air/sand mixture simulation.

|            | isotropic coarsening | anisotropic coarsening |
|------------|----------------------|------------------------|
| its        | 196                  | 22                     |
| time (sec) | 42.95                | 4.24                   |

Tables 1 and 2 illustrate the benefit of new algorithm improvements (anisotropic coarsening) on two anisotropic FUEGO simulations. The results in Table 1 correspond to a large-scale turbulent helium plume calculation on 16,384 processors while Table 2 is for an Eulerian/Lagrangian coupled air/sand mixture simulation. Run times and iterations are given for the last matrix solution, which is representative of the repeated linear solutions in these simulations. In one



**Figure 1.** FUEGO results for a helium plume (left) and an air/sand mixture (right) simulation.

case, a factor of four improvement is realized while in another case a factor of ten is gained using the new algorithms. Similar improvements have been observed on industrial and academic calculations from external customers. Figure 1 presents corresponding solutions obtained from FUEGO.

Our algebraic multi-grid research and ML software capability (within the Trilinos framework) are focused on scalable solutions of linear systems on massively parallel architectures. Here, scalable means that the time-to-solution scales linearly in the number of unknowns and is constant when the number of processors increases while holding fixed the number of unknowns per processor. Scalability is attained by performing computations on coarse resolutions to accelerate convergence. The key lies in the complementary nature of relaxation (simple iterative schemes such as Gauss-Seidel) and coarse corrections. Relaxation typically damps oscillatory errors while coarse corrections address smooth error.

The main difficulty is that relaxation is sensitive to anisotropic phenomena. Specifically, standard relaxation damps oscillatory errors only in directions of strong coupling. Our new algorithms are based on the notion that automatic coarsening should not be isotropic but instead should only occur in directions where oscillatory errors can be damped. Figure 2 illustrates isotropic algebraic coarsening, while Figure 3 illustrates an anisotropic case. Anisotropic coarsening allows oscillatory errors to be damped on coarse grids as coarsening does not occur in problematic directions.

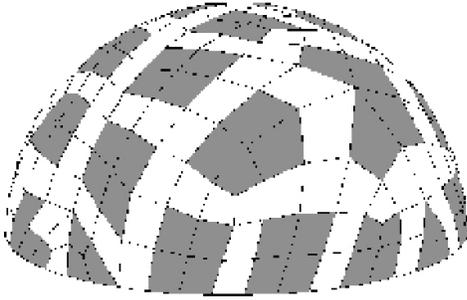


Figure 2. Fine nodes connected by a gray polygon are collapsed to one coarse node.

Two difficulties arise in performing anisotropic coarsening. The first is detecting the phenomena or more specifically detecting directions of strong coupling. The second is in developing an appropriate grid transfer that is sufficiently sparse and accurately transfers functions that are smooth in strong coupling directions.

Multiple approaches have been developed for detection. The most basic drops *small* matrix entries during the coarsening phase. These dropped entries correspond to the removal of edges in the matrix graph to be coarsened. Care must be exercised in defining the precise meaning of *small*. Unfortunately, the size of matrix entries can be misleading. Thus, two capabilities for generating an auxiliary matrix have been developed. A distance Laplacian matrix can be defined when coordinates are available on a stretched mesh. Points that are physically close have large matrix entries, while distant points have small matrix entries. When used in conjunction with dropping, this effectively detects suitable directions of strong coupling where coarsening can occur. Another possibility is based on the propagation of information

in an auxiliary ordinary differential equation (ODE) problem, which is evaluated for a few steps using matrix-vector products.

Modified grid transfer construction algorithms have also been defined for prolongators with arbitrary sparsity patterns. These effectively define coarse basis functions based on an energy minimization principal subject to constraints associated with interpolating certain important modes accurately (e.g., rigid body modes.)

### References

- Gee, M., Hu, J., and Tuminaro, R. (2009). A New Smoothed Aggregation Method for Anisotropic Problems, *Numerical Linear Algebra and Applications*, Vol. **16**, pp.19-37.
- Schroder, J., Olsen, L., and Tuminaro, R. (2007). A Generalized Strength Measure for Algebraic Multigrid, *CSRI Summer Proceedings*, eds. Parks and Collis, SAND2007-7977, Sandia National Laboratories Albuquerque, NM , pp. 14-25.
- Gee, M., Hu, J., Siefert, C., and Tuminaro, R. (2006). *ML 5.0 Smoothed Aggregation User's Guide*, SAND2006-2649, Sandia National Laboratories ,Albuquerque, NM.

***This work has been supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

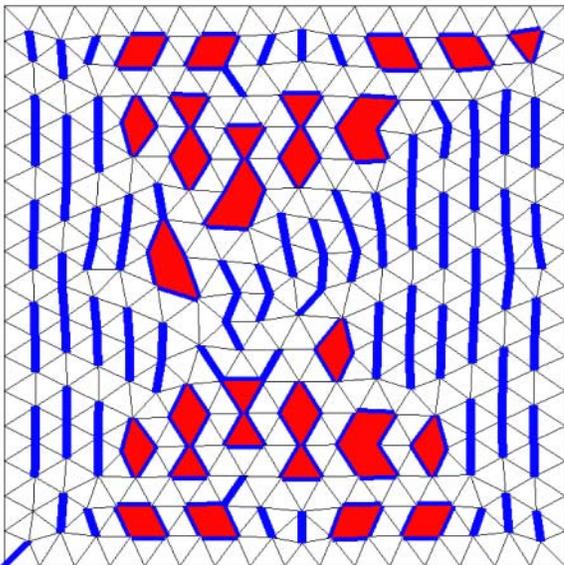


Figure 3.  $\epsilon u_{xx} + u_{yy}$  coarsening. Fine nodes connected by a red polygon and/or blue line are collapsed to one coarse node.

# ALGEBRAIC MULTI-GRID ALGORITHMS SCALE TO THOUSANDS OF PROCESSORS

Contacts: Ray Tuminaro, Scalable Algorithms, [rstumin@sandia.gov](mailto:rstumin@sandia.gov)  
 Jonathan Hu, Scalable Algorithms, [jhu@sandia.gov](mailto:jhu@sandia.gov)  
 Chris Siefert, Computational Shock and Multi-physics, [csiefer@sandia.gov](mailto:csiefer@sandia.gov)  
 Allen Robinson, Computational Shock and Multi-physics, [acrobin@sandia.gov](mailto:acrobin@sandia.gov)  
 Stefan Domino, Computational Thermal and Fluid Mechanics, [spdomin@sandia.gov](mailto:spdomin@sandia.gov)

*Our research focuses on new algorithms that differ significantly from traditional AMG in processing coarse models much more efficiently. This has allowed for unprecedented performance on notoriously difficult applications.*

Highly scalable parallel linear solvers are critical to the success of large-scale scientific and engineering calculations.

This project focuses on developing a new class of novel algebraic multi-grid (AMG) algorithms that scale well to tens of thousand processors. Loss of scalability within traditional AMG is typically associated with a poor communication to computation ratio for coarse resolution processing. This means that during coarse computations, the machine is significantly under-utilized. Our research focuses on new algorithms that differ significantly from traditional AMG in processing coarse models much more efficiently. This has allowed for unprecedented performance on notoriously difficult applications.

We are pursuing two different strategies. The first centers on producing AMG algorithms with significantly different communication characteristics than traditional multi-grid. The second focuses on reducing the number of coarse resolution meshes while also improving the communication to computation ratio on coarse meshes. In both cases, the key to success relies on maintaining optimal numerical convergence properties for the redesigned AMG cycle.

While performing initial benchmarking of existing multi-grid methods on SIERRA/Fuego and SIERRA/ALEGRA simulations on Red Storm, we were fortuitously able to identify and address several subtle issues in the multi-grid setup kernel. The result was a doubling of the efficiency of the overall multi-grid method at 5832 cores of Red Storm on a key SIERRA/ALEGRA magnetics benchmark (Figure 1).

Subsequently, we implemented an aggressive coarsening/domain decomposition capability in Trilinos/ML to explore multi-grid algorithms with significantly different computation and communication characteristics than traditional multi-grid methods. This framework allows us great flexibility in exploring the effect, for example, of using much coarser grids than usual.

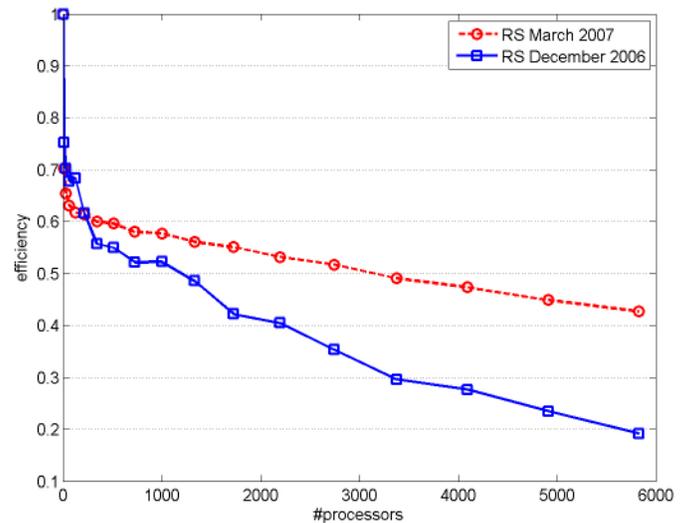


Figure 1. Old versus new AMG parallel efficiencies.

The single most compute intensive kernel in the setup phase of algebraic multi-grid is sparse matrix-matrix multiplication. In anticipation of multi-grid for multi-physics simulations, which have potentially many degrees of freedom per node, we implemented and profiled a block version of the matrix-matrix multiplication kernel. We observed speedups of almost 2.5 for 3-by-3 blocks, over the point kernel for the same input data (Figure 2). In practice, blocks are usually much larger, and we anticipate even larger speedups.

We worked closely with the Zoltan team to test and integrate Zoltan's new hypergraph repartitioning into Trilinos/ML. Zoltan's previous load-balancing capability was critical for achieving good scaling performance during a Level 1 Milestone on Sandia's ASC Red platform. The new hypergraph repartitioning capability allows load-balancing of multi-grid operators without coordinates,

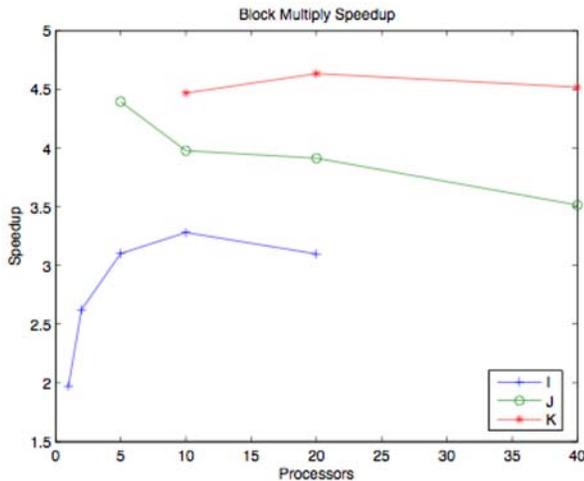


Figure 2. Block matrix-matrix multiply speedups over standard point multiply for various sized linear elasticity systems

which can be particularly advantageous in non-Cartesian domains. Hypergraph repartitioning also permits diffusive data migration, (i.e., starting with an initial decomposition and moving to a better but still similar decomposition). This is particularly important in multi-grid in order to avoid wholesale data transfer between disjoint processor sets when moving between levels.

We have developed, and are currently evaluating, an algebraic multi-grid method that requires communication only at the coarsest level. This is a generalization of a

two-level geometric method published in the literature. The key is that overlapping subdomains are established at the finest level. Multi-grid is applied independently in each subdomain, and information is communicated at the coarsest level. This method has slightly poorer convergence properties than traditional multi-grid, but significantly less communication.

Finally, we are developing AMG methods that are tailored to many core processors. Essentially, these methods will allow for computational subdomains that span multiple processing cores. Within a subdomain, standard MPI-based solvers can be used. Information transfer between compute nodes can be handled by an outer overlapped domain decomposition scheme, which in turn will act as a smoother for an outer multi-grid method.

### References

- Gee, M., Siefert, C., Hu, J., Tuminaro, J., et al. (2006). *ML 5.0 Smoothed Aggregation User's Guide*, SAND2006-2649, Sandia National Laboratories, Albuquerque, NM.
- Karlin, I., and Hu, J. (2007). Implementing and Profiling of a Variable Block Matrix-Matrix Multiply in ML, *CSRI Proceedings*, eds. M.L. Parks and S. S. Collis, pp. 49-56.

***This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) program.***

# NEW ALGORITHMS IMPROVE PERFORMANCE ON NEXT GENERATION MULTI-CORE COMPUTERS

Contacts: Michael Heroux, Scalable Algorithms, [maherou@sandia.gov](mailto:maherou@sandia.gov)

Chris Baker, Scalable Algorithms, [cgbaker@sandia.gov](mailto:cgbaker@sandia.gov)

Ron Brightwell, Scalable System Software, [rbbrigh@sandia.gov](mailto:rbbrigh@sandia.gov)

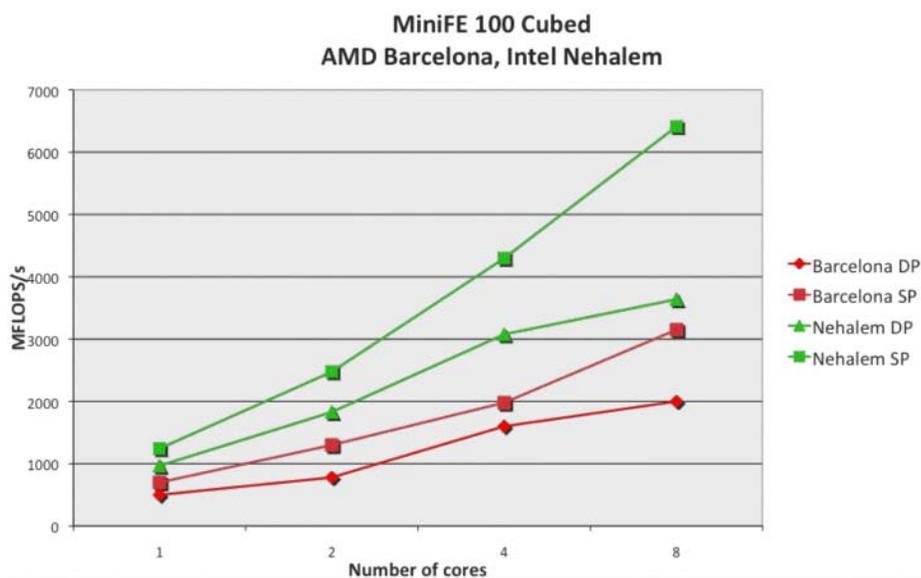
*Our efforts are focused on the development of algorithms that take advantage of lower-precision data (single-precision vs. double-precision floating point data), higher computation-to-bandwidth ratios, and hybrid distributed-shared memory algorithms.*

Multi-core nodes present a new challenge and opportunity for algorithms research especially since memory system performance per core is typically not sufficient for current approaches to effectively utilize all cores. Although we can continue to realize modest performance gains using existing algorithms, novel approaches are needed. Our efforts are focused on the development of algorithms that take advantage of lower-precision data (single-precision vs. double-precision floating point data), higher computation-to-bandwidth ratios, and hybrid distributed-shared memory algorithms. The combination of these efforts allows us to minimize the impact of memory system weaknesses, and exploit shared memory resources on multi-core nodes [1].

**Mixed Precision Algorithms:** Selective use of single-precision (SP) data vs. double-precision (DP) for computations effectively doubles the memory system bandwidth and cache sizes and leads to more effective multi-core use as shown in Figure 1. These results illustrate that effective use of SP data can have a dramatic impact

on performance. At the same time, we must still be able to deliver accurate answers, so our recent efforts are focused on the development of reliable diagnostics and robust iterative methods that can better tolerate the use of lower precision data and still deliver an accurate answer.

**Advanced Linear Solvers:** Scalable linear equation solvers, which are critical to many advanced science and engineering applications, are often the performance determining computation on high-end systems. Standard solvers have low compute to bandwidth ratios, resulting in poor processor utilization. This problem becomes even worse on multicore nodes where bandwidth does not scale with core use. Block iterative methods are a class of solvers that dramatically improve the computation to bandwidth ratio and at the same time improve robustness. Furthermore, block iterative methods are perfectly matched for linear systems that come from advanced simulation needs such as perturbation analysis, uncertainty quantification and advanced optimization efforts. Figure 2 shows the performance of the basic solver



**Figure 1.** These results show performance in megaFLOPS of single-precision computation vs. double precision on two modern multicore nodes. Note that single-precision (green) is faster on a single core and scales well. In contrast, double-precision results on 8 cores are only marginally better than on 4 cores.

kernels on sparse matrix-vector multiplication and sparse matrix-matrix multiplication, illustrating that as block sizes increase single core performance improves, as does utilization of additional cores.

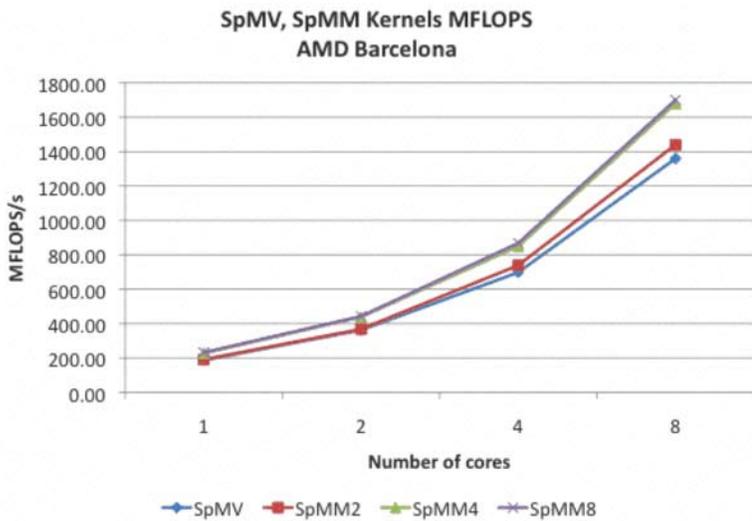
**Hybrid Parallel Environments:** Performance analysis of existing scalable applications shows that in many cases only certain portions of the computation require special treatment for multicore nodes. Figure 3 shows performance results for the application Tramonto. Here the time spent in the application decreases dramatically as core use grows. In contrast, the solver time decreases much more slowly. In other words, use of the traditional distributed memory message passing interface (MPI) approach is sufficient for effective scaling for much of the application's computation. To address this problem, we are developing hybrid runtime environments that allow an application to run unchanged on

multi-core nodes while the solvers take advantage of shared memory via special memory allocation functions. This approach allows solvers to improve performance via shared memory without any impact on the application.

### Reference

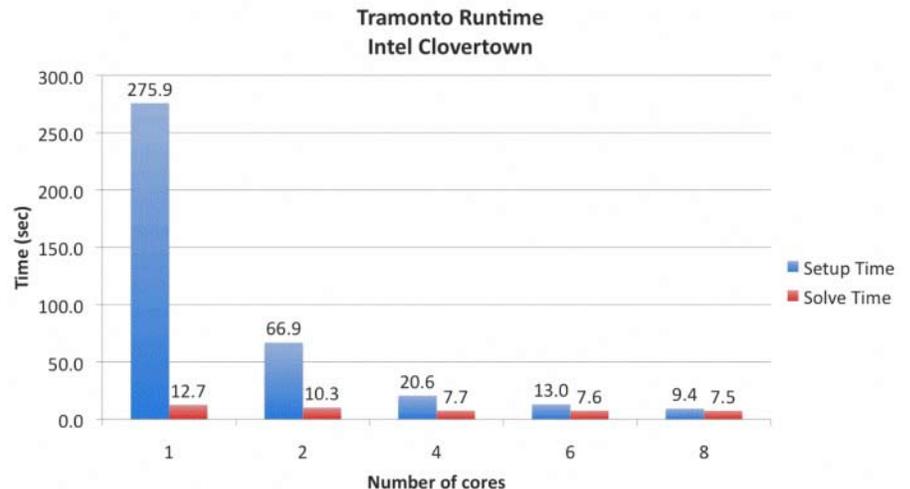
1. Heroux, M. A. (2008). Design Issues for Numerical Libraries on Scalable Multicore Architectures, *Journal of Physics: Conference Series*, p. 012035.

**Acknowledgements:** This work has been supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.



**Figure 2.** These results show the performance in mega FLOP/s for SpMV (block size of 1), SpMM (block sizes of 2, 4, and 8) on 1 to 8 cores of an AMD Barcelona processor. Larger block sizes perform better on a single core and scale better to 8 cores.

**Figure 3.** These results show time spent in the application (setup time) and time spent in the solver. As core count increases, time spent in the application decreases dramatically, while the solver time decreases much more slowly.



# SCALABILITY-CUBED: ALGORITHMS, CAPABILITY DEVELOPMENT, AND INTEROPERABILITY VIA THE TRILINOS FRAMEWORK

Contact: Andrew Salinger, Applied Math and Algorithms, [agsalin@sandia.gov](mailto:agsalin@sandia.gov)

*The Trilinos success in delivering scalable algorithms, capability development, and interoperability is distinguishing Sandia in its ability to develop and deliver world-class predictive simulation capabilities.*

The Trilinos Framework is a development environment and delivery mechanism for a wide range of computational science algorithms developed at Sandia. Most of Sandia's flagship algorithmic efforts for high performance computers are part of Trilinos, including libraries for linear and nonlinear solvers, as well as preconditioning (Aztec, ML, NOX). Trilinos is in a period of rapid growth in its capabilities, and impact. The success that Trilinos is experiencing can be attributed to two sources: the software principles behind the Trilinos framework and the excellent computational science capabilities delivered through it.

Trilinos has two identities. The *Trilinos framework* is a set of software tools that provide an excellent environment for the development and delivery of scientific software: e.g. version control, an automated regression test capability, a common build system, an established release process. *Trilinos* also refers to the aggregate of the computational capabilities that are delivered through the framework, and consist of semi-autonomous "packages." with small development teams focusing on a specific area of expertise.

The two-level design is very powerful. Small, focused development teams can develop packages around specific areas of expertise, while leveraging the framework's mature set of tools and processes. The benefit is even greater to the *users* of the Trilinos capabilities, most notably the Advanced Simulation & Computing (ASC) Program's SIERRA and RAMSES code suites, since all of Trilinos' packages can be downloaded and installed as one application.

The Trilinos project is delivering excellent — and striving for optimal — scalability in three distinct dimensions. (Scalability refers to the desirable behavior that an improvement grows linearly with an increase in devoted resources. Complexity, synchronization, and communication are all enemies of scalability.)

1. *Algorithmic Scalability:* Trilinos has developed its reputation on providing world-class scalable solution algorithms for massively parallel computers. Iterative linear solvers and preconditioners often form the

computational kernels of large-scale scientific computations and have the highest burden in terms of achieving algorithmic scalability. Active projects in this area include matrix-reordering schemes for electronic-circuit simulations and block-Krylov methods applied to linear stability analysis calculations.

Another key aspect of achieving scalability on parallel computers is load balancing, an area where Sandia's Zoltan project (now delivered via Trilinos) has made excellent progress in the development and application of hypergraph partitioning techniques.

2. *Capability Development Scalability:* The two-level Trilinos design, with the top-level framework and numerous semi-autonomous packages underneath, has created a very efficient environment for scientific software developers. The scalability of the Trilinos model for Capability Development has led to an explosion in the number of algorithmic capabilities that are leveraging the common Trilinos framework as a development environment and deployment vehicle. In September 2008, Trilinos version 9.0 was released with 41 packages. These capabilities are spreading beyond the core of linear and nonlinear solution methods to automatic differentiation tools, a compatible discretization library, a mesh database, and new low-level algorithms tailored to multicore architectures. Through this rapid growth, the framework design is proving to be robust and highly scalable.
3. *Scalability of Capability Interoperability:* The third dimension of scalability being actively addressed in Trilinos is the interoperability of the packages. One of the strategic goals of Trilinos is to make separate packages interoperable wherever it makes sense algorithmically. For example, any Trilinos load-balancing algorithm should be able to operate on any Trilinos mesh database. To manage the complexity inherent in the growth of capabilities, a concerted effort is ongoing to introduce interoperability layers around groups of related packages. Figure 1 shows the domain model for a computer code with six such abstract layers.

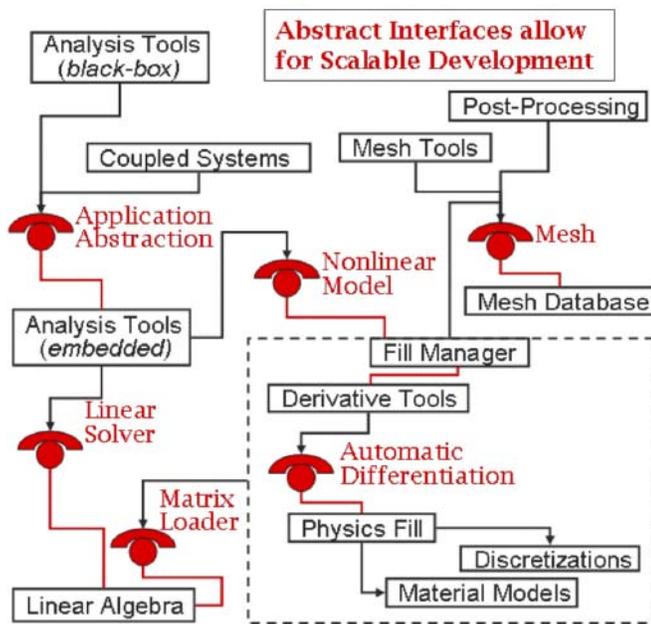
The recently completed “Vertical Integration” milestone for the ASC program culminated in some demonstrations of how Trilinos’ three dimensions of scalability can be used to attack a mission-critical problem: in this case, simulating the effects of neutron damage on the performance of an electronic device. Algorithmic scalability has been shown up to  $O(108)$  unknowns and  $O(104)$  computational cores. Scalable capability development was demonstrated with over a dozen distinct capabilities being brought to bear on this application, ranging from multipoint optimization to automatic differentiation. Interoperability was showcased on a transient sensitivity analysis calculation, as detailed in Figure 2, where 10 Trilinos packages worked in coordination to produce this crucial capability.

The Trilinos success in delivering scalable algorithms, capability development, and interoperability is distinguishing Sandia in its ability to develop and deliver world-class predictive simulation capabilities.

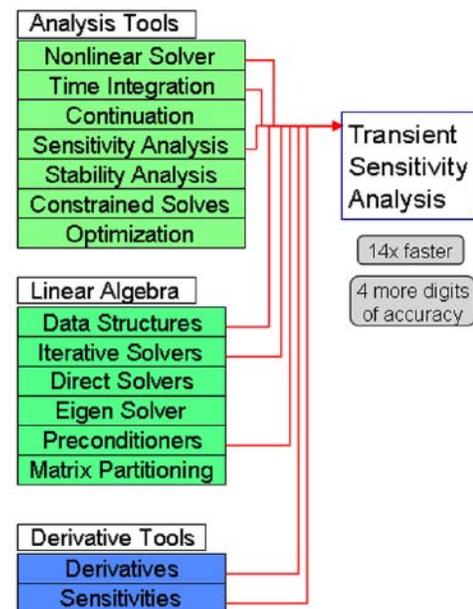
### Reference

Heroux, M., et al. (2005). An Overview of Trilinos. (2005). *ACM Transactions on Mathematical Software*, Vol. 31, No. 3, pp. 397-423.

*The Trilinos Framework has been supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation & Computing (ASC) Program, while Trilinos packages have been supported by ASC, DOE ASCR, SciDAC, and LDRD Programs.*



**Figure 1.** The domain model for a PDE code built with interoperable Trilinos components shows six abstract interfaces, in red. These allow for scalable access to predictive science capabilities.



**Figure 2.** A schematic of the Vertical Integration Milestone illustrates how several independent, interoperable algorithmic pieces combine to achieve a transformational capability: the transient sensitivity analysis of neutron damage effects on an electronic device, with 14x more efficient and 10<sup>4</sup>x more accurate results than the previous finite differencing approach.

# TENSOR DECOMPOSITIONS PROVIDE NEW METHODS FOR ANALYZING CRITICAL DATA MINING APPLICATIONS

Contact: Brett W. Bader, Computer Science and Informatics, [bwbader@sandia.gov](mailto:bwbader@sandia.gov)

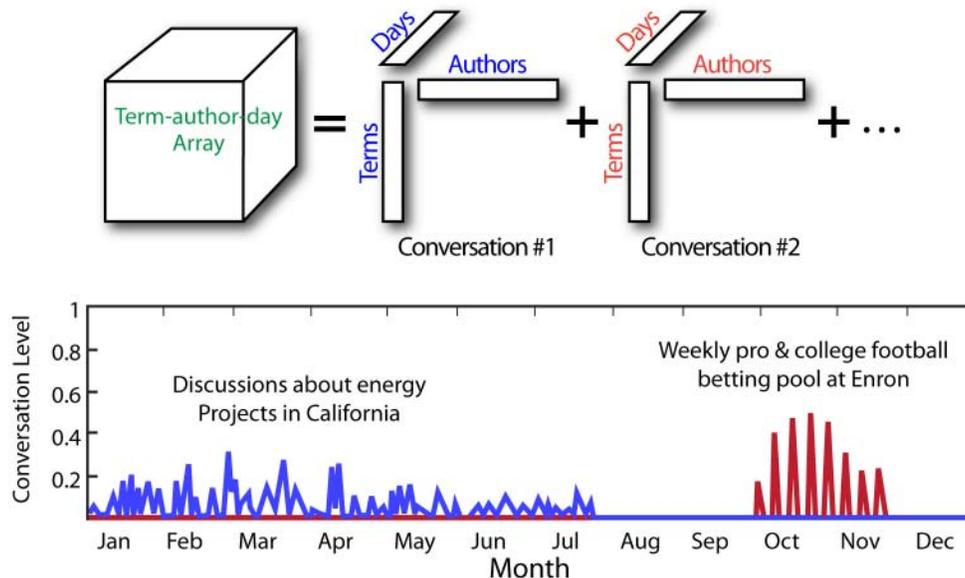
*Tensor decomposition techniques have been applied at Sandia to a wide range of problems related to national security, including social network analysis, email surveillance, cross-language information retrieval, and web link analysis.*

For many years, data mining has benefited from linear algebraic techniques and matrix decompositions, such as the singular value decomposition (SVD). These techniques have proven successful in challenging problems, from web search to social network analysis. Recently, research at Sandia with funding from the Lab Directed Research and Development (LDRD) program has taken data mining to another dimension, quite literally. Sandia researchers have moved beyond matrices and have formulated data mining problems using multidimensional arrays, which are analyzed subsequently with tensor decomposition [1]. By incorporating metadata in the problem as an extra dimension, these novel techniques extract latent information and subtle relationships in the multivariate data that are missed by matrix techniques. Tensor decompositions techniques have been applied at Sandia to a wide range of problems related to national security, including social network analysis, email surveillance, cross-language information retrieval, and web link analysis.

In joint research with Prof. Michael Berry, of the University of Tennessee, a new approach was developed for automatic conversation detection in email over time where a term-by-author-by-time array encodes an email collection [2]. Tensor

analysis of over fifty thousand email messages released in the Enron investigation identified dominant conversations based on the theoretical information gain of terms used in the body of the emails and their co-occurrence information over time and by authors (Figure 1). Strong correlations among terms, authors, and time were identified automatically and, when taken together, represented a conversation. These conversations helped to illuminate activities that occurred just prior to the downfall of Enron.

Another application of tensor decompositions was a data mining validation demonstration that showed how one or two analysts could extract preprogrammed terrorist activities from a large synthetic data set [3]. In this study, we developed a strategy based on algebraic models to extract and sequence important activities and specific events from data sources such as news articles and blogs. The ability to automatically reconstruct a plot or confirm involvement in a questionable activity is greatly facilitated by our approach. We applied our algorithms to the terrorism-based scenarios of the VAST 2007 Contest data set to demonstrate how term-by-entity associations can be used for scenario/plot discovery and evaluation.



**Figure 1.** Tensor analysis of Enron emails identified conversations (topic, participants, and temporal profile), ranging from routine business topics to a weekly betting pool involving two dozen employees.

For this study, we set out to automatically detect associations of entities (specific people, locations, organizations, etc.) with terms that appear within the same document and, concurrently, have the same co-occurrence pattern in time. We used the nonnegative tensor factorization (NTF), which is based on the well-known PARAFAC model for multidimensional data, to extract useful term-entity-time associations. These associations provide context for intelligence analysts and provide concise summaries for further investigation. Using the VAST 2007 Contest data set, we have successfully applied our analysis tools to a large, synthetic data set with a “buried” illegal activity and terror scenario.

Our algorithms were implemented in the Tensor Toolbox for MATLAB (<http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/>), which was created in a previous LDRD project to facilitate research with tensors and now has more than one thousand registered users worldwide. An ongoing effort is to achieve scalable solutions for larger data mining problems using Sandia’s expertise in high-performance computing. With current work in our Networks Grand Challenge LDRD, we are developing a tensor decomposition library in C++ for fast algorithms on serial and parallel

computers. This advance will enable us to analyze even larger problems, whether from text corpora for content analysis or from network traffic in cyber security. In addition, we are developing different types of tensor decompositions for pattern analysis and clustering in, for example, directed graphs.

### **References**

1. Kolda, T. G., and Bader, B. W. (2009). Tensor Decompositions and Applications, to appear in *SIAM Review*.
2. Bader, B. W., Berry, M., and Browne, M. (2008). Discussion Tracking in Enron email using PARAFAC Part III, Email Surveillance and Filtering, p. 147 in *Survey of Text Mining II*, Springer.
3. Bader, B. W., Purotskiy, A. A., and Berry, M. W. (2008). Scenario Discovery Using Nonnegative Tensor Factorization, *Lecture Notes in Computer Science*, Vol. **5197**, pp. 791-805.

***This work has been supported by SNL’s Laboratory Directed Research and Development (LDRD) Program.***

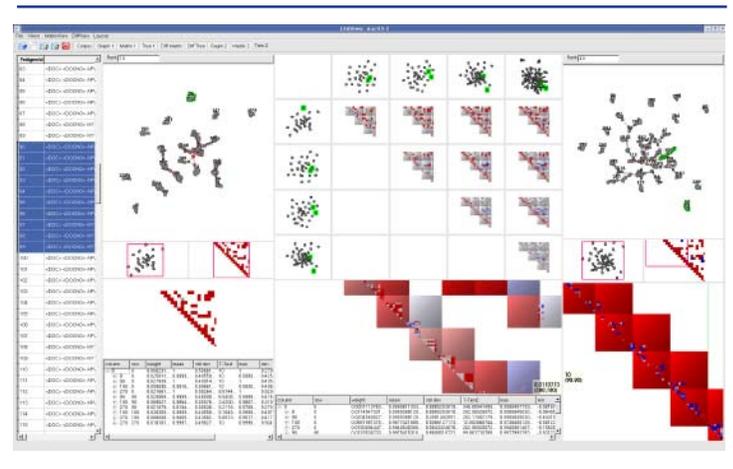
# ADVANCED METHODS ALLOW PROCESSING AND SEARCHING OF VERY LARGE DOCUMENT COLLECTIONS

Contacts: Daniel Dunlavy, Computer Science & Informatics, [dmdunla@sandia.gov](mailto:dmdunla@sandia.gov)  
 Tim Shead, Data Analysis and Visualization, [tshead@sandia.gov](mailto:tshead@sandia.gov)  
 Patricia Crossno, Data Analysis and Visualization, [pcross@sandia.gov](mailto:pcross@sandia.gov)

*...we have developed a flexible suite of scalable text analysis software components and applications, visual algorithm analysis tools, and a software system for creating and evaluating ensemble models for general machine learning tasks. These tools are currently being used to solve problems in the areas of nonproliferation, technology surprise, nuclear materials attribution and computer simulation data analysis.*

Intelligence analysts have a big data problem. They answer questions of national security under extreme time pressure, but current tools do not scale to the volume of data they must consider. No end-to-end scalable visual text analysis capabilities exist today, and this prevents analysts from exploring, analyzing, and annotating existing petascale document collections. For this project, we have developed a flexible suite of scalable text analysis software components and applications, visual algorithm analysis tools, and a software system for creating and evaluating ensemble models for general machine learning tasks. These tools are currently being used to solve problems in the areas of nonproliferation, technology surprise, nuclear materials attribution and computer simulation data analysis.

The main contributions to date of this project include the design and development of the ParaText™ Server (Figure 1), Algebraic Engine, LSAView (Figure 2), and HEMLOCK data modeling and analysis software frameworks and applications. The ParaText™ Server is both a platform for development

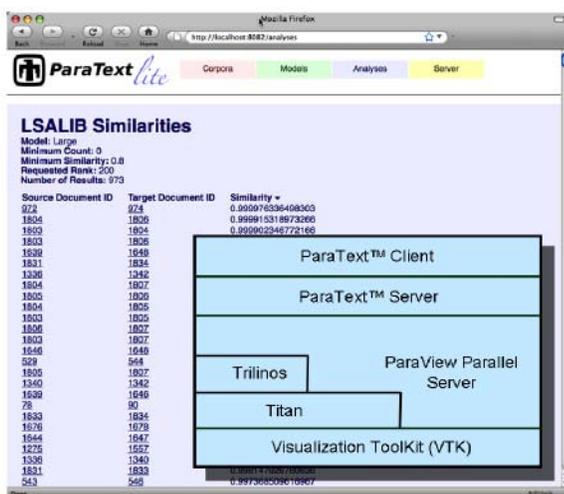


**Figure 2.** The LSAView application is being used to visually analyze text analysis algorithms and their impact on end-user interpretation of results.

and experimentation of large-scale text analysis algorithms. It provides a web services API and includes components for text processing and latent semantic analysis (i.e., statistical language modeling, conceptual information retrieval, and similarity analysis based on term usage). Figure 1 presents the application stack and an HTML example of the ParaText™ Server.

The Algebraic Engine generalizes the capabilities of the ParaText Server to provide an architecture for the storage and processing of algebraic data analysis models. It consists of a set of Visualization Toolkit (VTK) components for creating and processing algebraic models (e.g., matrix SVD and tensor PARAFAC models), analyzing data using such models (e.g., similarities between data objects), and interacting with SQL databases for model caching. The Algebraic Engine is currently being used in the following applications:

- ParaText™: scalable text analysis
- OverView: open source information visualization framework
- TreeView: temporal information visualization for nuclear materials forensic attribution



**Figure 1.** The ParaText™ Server is built on existing large-scale analysis capabilities developed at Sandia (inset). Through its web services API, ParaText™ Server can deliver text analysis capabilities to any desktop with a web browser.

- LSAView: text algorithm analysis
- LDRDView: funding portfolio analysis
- ParaSpace: simulation data analysis

LSAView is a visualization application for analyzing and exploring the impact of informatics algorithms on end-user visual analysis of data. LSAView aids developers in the process of algorithm design and aids users in the discovery of optimal parameter settings for specific data analysis problems. Figure 2 illustrates the current features:

- Side-by-side data comparison views for two sets of algorithm parameters
- Small multiple view for analyzing three or more parameter sets simultaneously
- Linked document, graph, matrix, and tree data views
- Interactive, zoomable, hierarchical sparse matrix and difference views
- Statistical inference tests used in highlighting novel parameter impact

For this project, we have also been investigating machine learning methods for incorporating relevance feedback into models and automatically tuning data analyses to align with end-user requirements. To this end, we have developed the Heterogeneous Ensemble Machine Learning Open Classification Kit (HEMLOCK) [1] for creating and evaluating ensemble models aimed at classification problems. More specifically, HEMLOCK can generate metamodels of classifier models of different types (e.g., probabilistic naïve Bayes, support vector machine, and decision tree models). Several methods for combining models (e.g., majority voting, sum rule, stacking, etc.) and creating diverse models (e.g., bagging, random forests, random subspace method, etc.) are currently available and have been investigated in solving

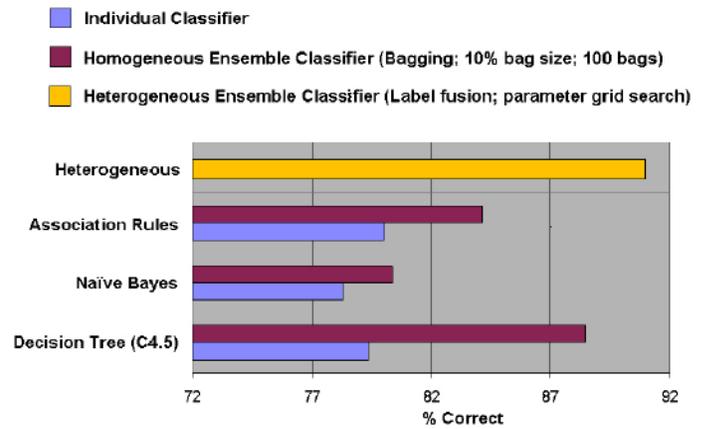


Figure 3. Results using HEMLOCK ensembles (yellow) indicate improved performance over existing individual and homogeneous ensemble models on a standard image classification task.

classification problems using images. Figure 3 illustrates the performance of HEMLOCK ensemble classifiers over individual and homogeneous ensemble classifiers. By combining diverse individual models to form ensembles of models, HEMLOCK models classify new data more effectively.

#### Reference

1. Gilpin S. A., and Dunlavy, D. M. (2008). Heterogeneous Ensemble Classification, *Computer Science Research Institute Summer Proceedings*, SAND 2008-8257P, Sandia National Laboratories, Albuquerque, NM.

*This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.*

# NEW TECHNIQUES SUPPORT THE CALIBRATION OF MODEL PARAMETERS UNDER UNCERTAINTY

Contacts: Laura P. Swiler, Optimization and UQ, [lpswire@sandia.gov](mailto:lpswire@sandia.gov)  
 Brian M. Adams, Optimization and UQ, [briadam@sandia.gov](mailto:briadam@sandia.gov)  
 Michael S. Eldred, Optimization and UQ, [mseldre@sandia.gov](mailto:mseldre@sandia.gov)

*We developed an approach for estimating model parameters which results in the “best distribution fit” between experimental data and simulation output.*

Scientists rely heavily on mathematical models to understand and predict the behavior of various engineered and physical systems. Before using such models for prediction, they must be validated to ensure that the model is appropriate and accurate enough for the intended use. Part of the validation process usually involves a calibration step, where a model is “calibrated” by choosing model parameters,  $\theta$ , so that model simulation results agree with experimental data to some degree (Figure 1). Classical statistical methods such as least-squares approaches are available for performing this calibration. However, there are some cases where we do not want to match simulation output to experimental data at each data point, but we may want to match some overall characteristics such as moments of the data (matching mean or variance). Often experimental data is of poor quality, has very few data points, or may not be directly relevant to the particular simulation model under investigation. Another case is where modelers just want to obtain a ballpark estimate based on bounds matching: i.e., analysts want the model upper and lower bound estimates (accounting for model uncertainty) to match the upper and lower bounds on the data. One might have several sets of experiments and want to match the mean or moments of each to a simulation model. Finally, the model parameters may have some uncertainty that needs to be incorporated into the calibration process.

We developed an approach for estimating model parameters that results in the “best distribution fit” between

experimental data and simulation output. Best distribution fit means matching moments of experimental data to those of a simulation (and possibly matching a full probability distribution). This approach extends typical nonlinear least squares methods that identify parameters maximizing agreement between experimental points and simulation results. The idea of matching percentiles of distributions between experimental data and simulation results is presented on the left side of Figure 2, where one wants to minimize the distance between the simulation percentiles shown in red and the empirical percentiles of the experimental data distribution shown in blue, with some results from a real problem shown on the right side of Figure 2.

We based our formulation (shown in Figure 3) on optimization under uncertainty (OUU) methods, focusing on nested approaches consisting of an outer loop nonlinear least squares optimization method (such as Gauss-Newton) and an inner loop uncertainty quantification (UQ) method that iterates on the simulation and provides statistical information (such as mean, variance, or percentile values) to the outer loop for incorporation in the NLLS objective function. Testing various approaches demonstrated that analytic reliability methods or stochastic expansion methods are more effective than sampling in the propagation of uncertainty in the inner loop [1]. This is partly due to the fact that reliability and stochastic expansion methods tend to be more accurate and efficient, and partly due to the availability

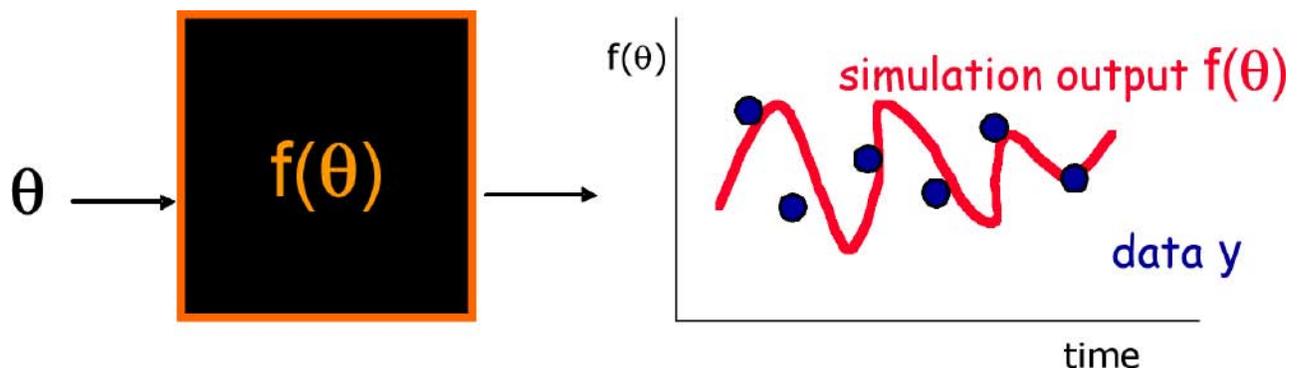


Figure 1. Example of model calibration.

of analytic derivatives of the statistics for these methods. We are performing additional research incorporating other uncertainty aspects into calibration, such as epistemic (“lack of knowledge”) uncertainty. These approaches that explicitly account for uncertainty in the calibration process offer significant possibilities for improving our parameter estimation capability.

### Reference

1. Swiler, L. P., Adams, B. M., and Eldred, M. S. (2008). Model Calibration Under Uncertainty: Matching Distribution Information, *Proceedings of the 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, AIAA-2008-5944. Victoria, British Columbia, Canada.

*This work has been supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program.*

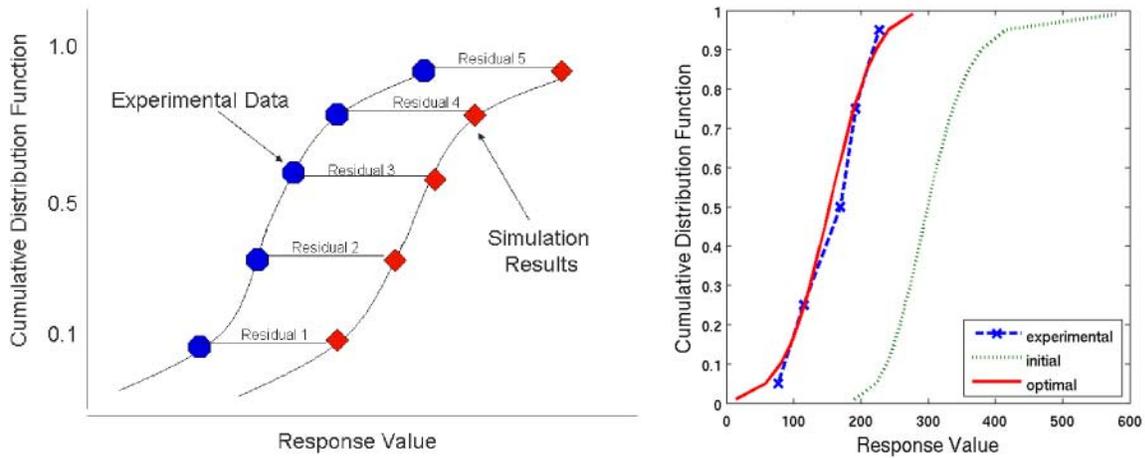


Figure 2. Matching distribution information: Left figure shows experimental data and simulation before optimization. Right figure shows simulation results after being calibrated to match experimental percentiles.

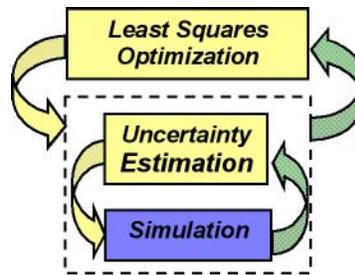


Figure 3. Solution approach.

# NEW COMPUTATIONAL METHODS ENABLE NOVEL NANOPARTICLE SIMULATIONS

Contacts: W. Michael Brown, Discrete Math & Complex Systems, [wmbrown@sandia.gov](mailto:wmbrown@sandia.gov)  
 Matt K. Petersen, Surface and Interface Sciences, [mkpeter@sandia.gov](mailto:mkpeter@sandia.gov)  
 Steven J. Plimpton, Scalable Algorithms, [sjplimp@sandia.gov](mailto:sjplimp@sandia.gov)  
 Gary S. Grest, Surface and Interface Sciences, [gsgrest@sandia.gov](mailto:gsgrest@sandia.gov)

*...we can now perform analysis and simulation of aspherical particles at time-scales that were not previously accessible with molecular dynamics software...This work will enable future studies with the potential for improving the processing and manufacturability of nanoparticle coatings and thin films.*

Nanoparticles are now more than ever being used to tailor materials function and performance in differentiating technologies because of their profound effect on thermo-physical, mechanical, and optical properties. The most feasible way to disperse particles in a bulk material or control their packing at a substrate is through fluidization in a carrier, followed by solidification. Unfortunately, the processing of particles as concentrated, fluidized suspensions into useful products remains an art. This is largely because the effect of particle shape and volume fraction on fluidic properties and suspension stability remains unexplored in a regime where particle-particle interaction mechanics is prevalent. For example, particles in nature and many manufacturing processes often have highly irregular shapes due to the way they were created (via aggregation, phase changes, milling, etc.). Although particle shape can have a large effect on the structure and mechanical properties of the composite material, the majority of computational particle mechanics simulators (Newtonian dynamics, Monte Carlo solvers, fluid/particle Stokesian dynamics) treat only spherical particles. This is due to the considerable complexity that asphericity introduces.

As part of a large effort aimed towards ultimately improving the processing and manufacturability of nanoparticle coatings and films [1, 2], we have incorporated a framework for treating aspherical particles into the massively parallel molecular dynamics package LAMMPS [3]. Currently, this includes treatment of ellipsoidal particles with anisotropic force fields (RE-squared and Gay-Berne) and integrators and thermostats that incorporate the additional degrees of freedom imposed by particle torque. Together with tools we have developed for triangulation and visualization of aspherical particles, we can now perform analysis and simulation of aspherical particles at time-scales that were not previously accessible with molecular dynamics software. Currently, we are utilizing these tools to perform studies on the aggregation, rotational diffusivity, and shear rheology of aspherical nanoparticles.

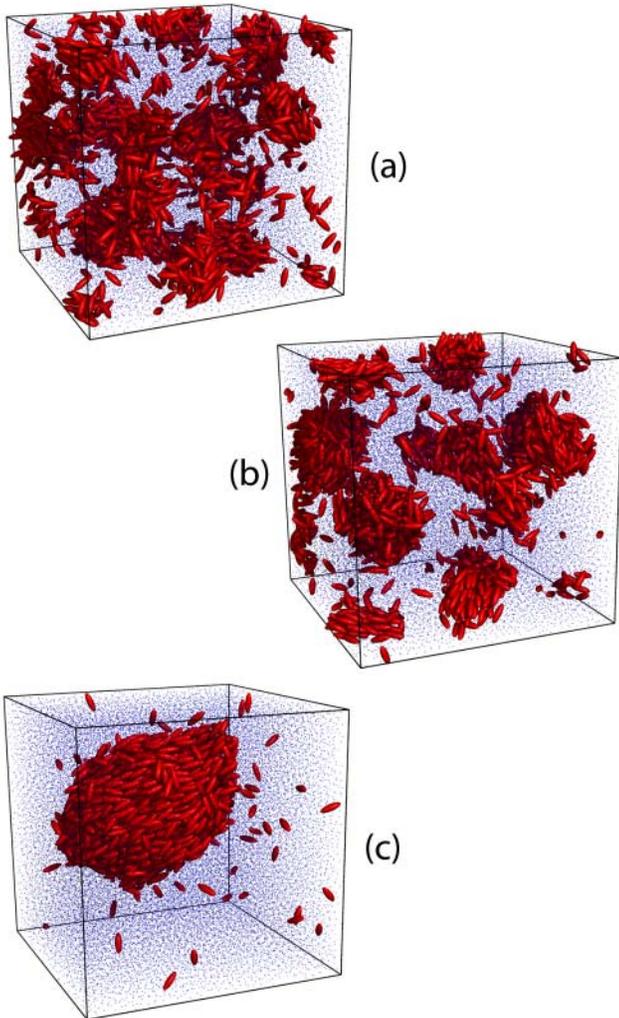
In one such study [4], we have analyzed the aggregation of liquid crystals into nematic emulsions. These liquid crystal nanodroplets have potential applications in electro-optical devices with applications including optical switches, light shutters, and reflective displays. Because the physical properties of dispersed liquid crystal droplets depend on the phase and structural characteristics of the droplets, understanding the crystallization/self-assembly process is important for targeted design of new materials. To obtain a better understanding of this process, we have performed parallel molecular dynamics simulations of droplet aggregation at long time scales. We have shown that nanodroplets formed from liquid crystal aggregation are only metastable (Figure 1) and will coalesce into a single large droplet. Cross-linking or cooling the solution to form a glassy matrix is required for stable electro-optical materials. The shape of the droplets is temperature dependent, with spherical droplets forming at higher temperatures and elongated structures forming at lower temperatures (Figure 2). The orientational order of the individual liquid crystal particles is rapidly equilibrated during droplet formation into nematic or smectic mesophases at low temperatures. When low shear rates are imposed, we find little impact on droplet structure. Interestingly, at high shear rates, the long axis of a droplet reorients to align with shear flow, but there is little change in the orientation of individual particles.

Currently, our efforts are focused on developing accurate anisotropic potentials to model the interactions of very large nanoparticles with solvent along with efficient strategies for modeling the large number of particles required in such simulations. This work will enable future studies with the potential for improving the processing and manufacturability of nanoparticle coatings and thin films.

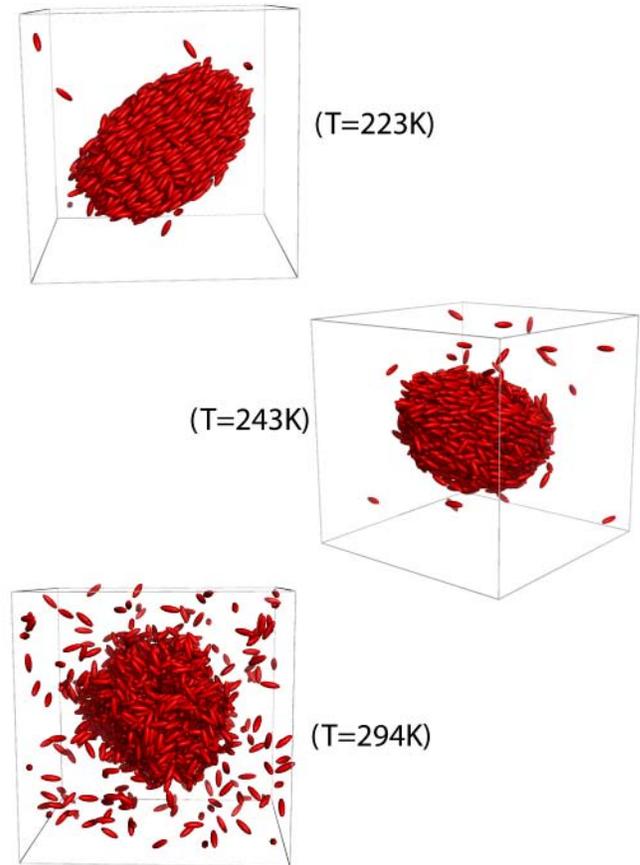
### References

1. German, J., Sandia Leads Three-Year Alliance with Companies Interested in Nanoparticle Coatings, Films, (2008). *Sandia Lab News*, Sandia National Laboratories, Albuquerque, NM.
2. Grest, G. S., Brown, W. M., Lechman, J. B., et al. (2008). *Nanoparticle Flow, Ordering, and Self-Assembly*, SAND2008-6516. Sandia National Laboratories, Albuquerque, NM, pp. 1-29.
3. Plimpton, S. (2008). *LAMMPS Molecular Dynamics Simulator*, <http://lammps.sandia.gov/>.
4. Brown, W. M., Petersen, M. K., Plimpton, S. J., and Grest, G. S. (2009). Liquid Crystal Nanodroplets in Solution, *Journal of Chemical Physics*, Vol. **130**, p. 044901.

*This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by Nanoparticle Flow Consortium CRADA in collaboration with Corning, BASF, 3M, Proctor & Gamble, and National Starch and Chemical.*



**Figure 1.** Snapshots of droplet aggregation at simulation times of 0.132\_s (a), 0.264\_s (b), and 2.112\_s (c).



**Figure 2.** Droplet shape at various simulation temperatures.

# ADVANCED ALGORITHMS ENABLE UNIQUE MODELING OF NANOPARTICLES IN SOLUTION

Contacts: Steve Plimpton, Scalable Algorithms, [sjplimp@sandia.gov](mailto:sjplimp@sandia.gov)  
 Randy Schunk, Nanoscale and Reactive Processes, [prschun@sandia.gov](mailto:prschun@sandia.gov)  
 Jeremy Lechman, Nanoscale and Reactive Processes, [jblechm@sandia.gov](mailto:jblechm@sandia.gov)  
 Gary Grest, Surface and Interface Sciences, [gsgrest@sandia.gov](mailto:gsgrest@sandia.gov)

*We have been developing computational methods to simulate how nanoparticulate fluids will behave during manufacturing processes... [The goal is] to develop a predictive computational capability that can guide the design of manufacturing processes involving nanoparticulate fluids.*

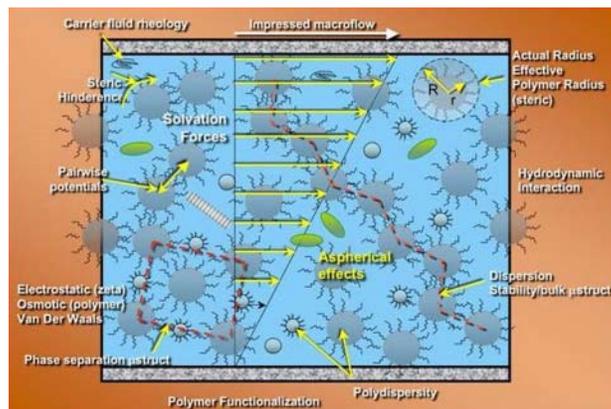
Compare the juice of a peach and honey: both are aqueous solutions containing sugar. Yet one is more viscous (thicker) and slowly flowing; the other is thinner. This illustrates how the flow (rheologic) properties of fluids can change markedly even when the same components are present in different proportions. More generally, the physics of flow can become far more complicated for fluids containing suspended or dissolved particles of different shapes and sizes with varied chemical interactions, as in Figure 1.

We have been developing computational methods to simulate how nanoparticulate fluids will behave during manufacturing processes. Multimillion particle simulations are possible, and the challenge is to develop methods that can integrate particle dynamics across a broad range of scales — from molecular dynamical interactions occurring at the nanosecond (billionth of a second) timescale through engineering dynamics at the multisecond scale.

This work has two sponsors: a Cooperative Research and Development Agreement (CRADA) between Sandia and several industrial partners, as well as the National Institute for Nano-Engineering (NINE), funded by the ASC Program within DOE/NNSA. The goal is the same for both customers: to develop a predictive computational capability that can guide the design of manufacturing processes involving nanoparticulate fluids. Consider, for example, the extrusion and drying of thin coatings. In this context, the fluid is a carrier medium for the nanoparticles. It can move them to desired locations and keep them dispersed until the fluid is removed and the particles then self-assemble into structures of interest.

Our work to date has focused on five aspects of this problem. All code development has been in our widely used open-source LAMMPS molecular dynamics package [1] to extend its functionality for nanoparticle modeling.

**1. Particle-particle interactions:** Since nano or colloidal particles can contain millions or billions of atoms, they



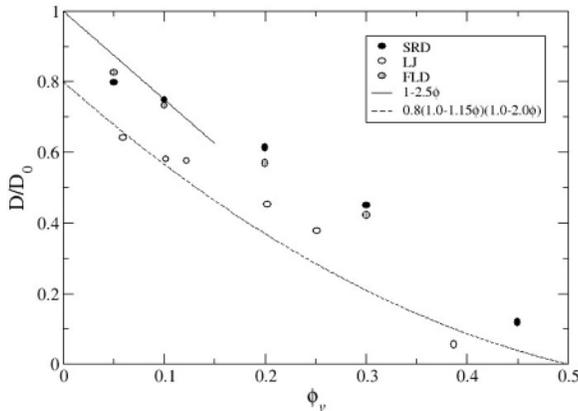
**Figure 1.** Schematic of a simulation box containing nanoparticles of various shapes, sizes, and chemical functionality in a background fluid.

must be coarse-grained to model them efficiently. We have implemented integrated forms of atomic-level potentials, based on van der Waals interactions that allow the pairwise interaction of two nanoparticles (or a nanoparticle with a solvent particle) to be computed cheaply via an analytic formula [2]. By appropriate choice of parameters, bare or coated nanoparticles can be modeled in this way.

## 2. Efficient algorithms for nanoparticles in explicit solvent:

For small nanoparticles in solution, the solvent must be modeled explicitly to capture layering and excluded volume effects. However, the two (or more) size scales present due to big coarse-grain nanoparticles interacting with small solvent particles pose computational challenges. We have developed parallel algorithms for finding neighbors, communicating ghost atoms to nearby processors, and treating the widely different interaction cutoffs (big/big, big/small, small/small) efficiently. For problems of interest, these methods are from 10 to 100x faster than traditional molecular dynamics algorithms, enabling much larger scale simulations [2].

**3. Stochastic rotation dynamics:** For larger nanoparticles, the solvent must also be coarse-grained, else the cost of



**Figure 2.** Diffusivity of spherical nanoparticles of size 20 (relative to solvent particles) as a function of volume fraction (packing density). Results from several competing computational models are shown: explicit solvent (LJ), coarse-grained (SRD) solvent, and implicit solvent (FLD), as well as predictions from theory (2 lines).

computing solvent-solvent interactions dominates. The recently formulated stochastic rotation dynamics (SRD) model is a possible solution to this problem. Fluid/nanoparticle interactions are mediated by collisions between SRD particles and nanoparticles. We have implemented an SRD model in LAMMPS; preliminary comparisons to explicit solvent models are quite favorable (Figure 2). For these problems, the SRD model is 600x faster than an explicit solvent model due to two effects: interactions between SRD particles can be ignored (20x savings), and a 30x larger timestep can be used at the time scale of nanoparticle/nanoparticle interactions. Our methodology for implementing SRD in a parallel MD code is detailed in reference [3].

**4. Implicit solvent models via Stokesian Dynamics:** At larger length and time scales, the background fluid must be treated implicitly. We have collaborated with Prof. Jon Higdon's group at the University of Illinois (Urbana-Champaign) to implement his fast-lubrication dynamics (FLD) algorithm in LAMMPS. This is a simplified form of Stokesian Dynamics that approximates long-range

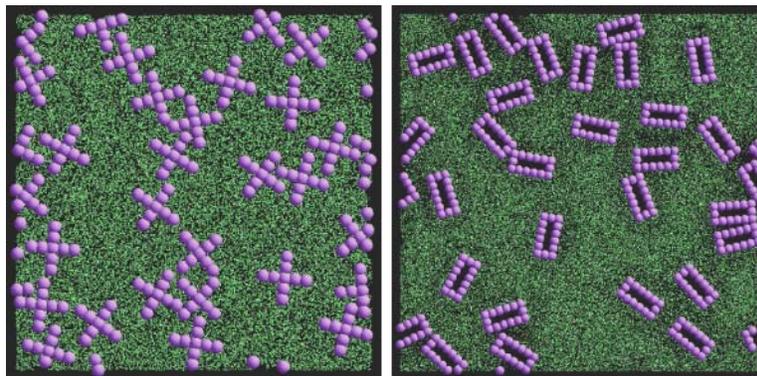
interactions by fitting procedures and thus allows cheaper, more parallel computations. Again, our preliminary comparisons of FLD to particle-based solvent models are favorable (Figure 2).

**5. Aspherical nanoparticles:** All of the preceding development was for large spherical nanoparticles in solution. We can treat aspherical particles in two ways, either as single ellipsoidal particles, or as composite particles, consisting of multiple spheres and ellipsoids glued together. Extending the ideas described above to these kinds of particles is ongoing work, though we have made considerable progress. Figure 3 shows composite nanoparticles embedded in an SRD fluid that is being sheared to measure the viscosity of the mixture. The FLD methods can be extended to mildly aspherical particles, and we are also developing coupled particle-fluid models where the fluid is treated via finite-element methods (using the Sandia SIERRA-ARIA code) on a grid conformal to the (aspherical) particle surfaces. This will enable modeling of general aspherical particles in an implicit fluid.

#### References

1. Plimpton, S. J., <http://lammps.sandia.gov>.
2. Veld, P. J., Plimpton, S. J., Grest, G. S. (2008). Accurate and Efficient Methods for Modeling Colloidal Mixtures in an Explicit Solvent Using Molecular Dynamics, *Comp Phys Comm*, Vol. **179**, pp. 320-329.
3. Letersen, M. K., Lechman, J. B., Veld, P. J., et al. (2008). Meso-scale Hydrodynamics via Stochastic Rotation Dynamics: Comparison with Lennard-Jones Fluid, submitted to *Journal of Chemical Physics*.

*This work has been supported by a Cooperative Research and Development Agreement (CRADA) between Sandia and several industrial partners, as well as the National Institute for Nano-Engineering (NINE), funded by the the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*



**Figure 3.** Simulation snapshots of aspherical nanoparticles (star-shaped, box-shaped) in a background fluid modeled as stochastic-rotation dynamics (SRD) particles.

# SANDIA OPERATIONS RESEARCH REDUCES DRINKING WATER CONTAMINATION RISKS

Contact: William E. Hart, Discrete Math and Concrete Systems, [wehart@sandia.gov](mailto:wehart@sandia.gov)

*Sandia researchers have developed mathematical algorithms and software to address several CWS [contaminant water systems] design challenges. A key issue is the placement of sensors ... Sandia has developed the TEVA-SPOT Toolkit to automatically optimize sensor locations.*

Sandia is partnering with the Environmental Protection Agency's National Homeland Security Research Center (EPA NHRSC) to develop contaminant warning systems (CWSs) that help protect drinking water systems. A CWS uses water sensors to monitor water quality and provide early detection of chemical or biological contaminants. Sandia researchers have developed mathematical algorithms and software to address several CWS design challenges. A key issue is the placement of sensors to protect the largest possible population from a broad set of contamination incidents.

Sandia has developed the TEVA-SPOT Toolkit to automatically optimize sensor locations. Sandia and the EPA have collaborated with the American Water Works Association (AWWA) to design CWSs for nine US water utilities, including very large distribution networks. The sensor placements developed in these studies could significantly reduce contamination risks. An economic assessment predicts that these CWS designs would reduce economic impacts due to fatalities by billions of dollars.

The 9/11 attacks have led to a new U.S. government focus on reducing risks to critical infrastructure, including water distribution systems. A terrorist attack on our drinking water infrastructure could severely impact the public health and economic vitality of our communities. An abundant supply of safe, high-quality drinking water is critical to the American way of life: at home, we count on safe water for drinking, cooking, washing clothes, bathing, and watering lawns; at work, we rely on high-quality water to operate restaurants, hospitals, and manufacturing plants; and in our communities, we depend on an abundant supply of water for fighting fires.

The distributed physical layout of drinking water systems makes them inherently vulnerable to a variety of incidents, including contamination with deadly agents that could result in large numbers of illnesses and fatalities. However, significant challenges remain in detecting contamination incidents early enough to mitigate both public health and economic impacts. Current monitoring practices at water



**Figure 1.** Illustration of how different sensor locations detect potential contamination incidents that start at different locations in a water distribution network.

utilities rely on infrequent sampling to meet regulatory requirements. Thus, the first evidence of a contamination incident is likely to be in an emergency room.

CWSs offer a promising strategy to minimize contamination risks. A CWS integrates monitoring and surveillance data from multiple detection streams to provide early detection of contamination in drinking water distribution systems. Sandia has partnered with the US EPA, the AWWA and academic and national laboratory researchers to develop and apply TEVA-SPOT, a tool for automatically selecting the locations of water quality sensors in a CWS.

Given a limited number of sensors, a desirable sensor placement minimizes the potential economic and public health impacts of a contaminant incident (Figure 1). The TEVA-SPOT software was developed to support a flexible decision framework for sensor placement that involves two major steps: a modeling process and a decision-making

process. The modeling process includes (1) describing sensor characteristics, (2) an ensemble of contamination incidents that a CWS should be designed to protect against, (3) selecting impact measures for the CWS, (4) planning utility response to sensor detection, and (5) identifying feasible sensor locations.

TEVA-SPOT was designed to model a wide range of sensor placement problems. For example, TEVA-SPOT supports a number of impact measures, including the number of people exposed to dangerous levels of a contaminant, the volume of contaminated water used by customers, the number of feet of contaminated pipe, and the time to detection. Response delays can also be specified to account for the time a water utility would need to verify a contamination incident before notifying the public. Finally, the user can specify the feasible locations for sensors and fix sensor locations during optimization. This flexibility allows a user to evaluate how different factors impact the CWS performance and to iteratively refine a CWS design.

Sandia has developed several state-of-the-art optimization strategies that can quickly find near-optimal sensor placements. These optimizers can be applied to a variety of sensor placement formulations that have been developed by Sandia. The goal of the standard TEVA-SPOT optimization model is to minimize the expected impact of contamination incidents. This model makes the common assumption that sensors never make detection errors. Other models have also been developed to optimize with data uncertainties and sensor detection failures. Finally, the TEVA-SPOT optimizers have been customized to perform limited-memory optimization. This capability is critical to enable application of TEVA-SPOT to real-world applications with 10,000s of junctions on commonly available workstations.

Sandia has partnered with the EPA and the American Water Works Association to use the TEVA-SPOT tool to design CWSs for nine of the largest US water systems, which serve

more than 6.5 million customers. For those utilities, estimated fatalities from high consequence terrorist attacks on drinking water are decreased by a median of 48%, with a range of 4-87%. Further, the estimated value of lives lost due to high consequence terrorist attacks on drinking water is reduced by a median of \$19 billion dollars, with a range of \$3-172 billion. Four of these utilities have already installed sensors for contamination warning systems, and the remaining five utilities are waiting for the funds needed to purchase sensors and other equipment.

Sandia's operations research expertise was critical to this success. Previously, optimization methods could only be effectively applied to small distribution networks; the largest network studied had fewer than 100 junctions. Thus, the only options available to analyze real-world distribution networks (with thousands of junctions) were expert judgment or ranking methods that use location features (e.g., distance to schools or pipe size). Sandia developed optimization methods that could quickly analyze these large systems, and in many cases, these optimizers have been shown to generate optimal solutions.

### **References**

- Berry, J., Hart, W. E., Phillips, C. E., et al. (2006). Sensor Placements in Municipal Water Networks with Temporal Integer programming Modules, *Journal of Water Resources Planning and Management*, Vol. **132**, No. 4, pp. 218-224.
- Murray, R., Janke, R., Hart, W. E., et al. (2008). Sensor Network Design of Contamination Warning Systems: A Decision Framework, *Journal American Water Works Association*, Vol. **100**, No. 11.

***This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the EPA NHSRC.***

# NEW ALGORITHMS IMPROVE WATER QUALITY WARNINGS

Contacts: Shawn Martin, Computer Science and Informatics, [smartin@sandia.gov](mailto:smartin@sandia.gov)  
 David Hart, National Security Applications, [dbhart@sandia.gov](mailto:dbhart@sandia.gov)  
 William Hart, Discrete Mathematics and Complex Systems, [wehart@sandia.gov](mailto:wehart@sandia.gov)  
 Sean McKenna, National Security Applications, [samcken@sandia.gov](mailto:samcken@sandia.gov)

*CANARY has been tested and successfully detects water quality changes on real-world datasets. ... many real-world events produce water quality changes ... detected by CANARY, resulting in what we would consider to be false positive identifications. In this project, we have developed alternative algorithms for the elimination of these false positives.*

CANARY is a Matlab ([www.mathworks.com](http://www.mathworks.com)) toolkit for detecting changes in water quality. It was developed for the U.S. Environmental Protection Agency (EPA) National Homeland Security Research Center Threat Ensemble Vulnerability Assessment program in conjunction with Sandia's Water Sentinel program.

CANARY has been tested and successfully detects water quality changes on real-world datasets. However, many real-world events produce water quality changes that are to be expected, such as addition of CL2 or the merging of different water sources throughout the day. (An example is shown in Figure 1.) These events are of course detected by CANARY, resulting in what we would consider to be false positive identifications. In this project, we have developed alternative algorithms for the elimination of these false positives.

In order to accurately assess our ability to identify false positives, we first generated simulated data using the EDDIES program with the help of Katherine Umberg at the EPA. The EDDIES program was able to generate multiple test data sets by combining data from actual locations in Cincinnati, Ohio.

False positive events were added to the data (by EDDIES) in order to create test data for our algorithms.

According to past results [1], the most reliable change detection algorithms in CANARY include the linear filter (LPC) and a multivariate nearest neighbor (MVNN) method. In this work, we implemented the ability to combine these (and other) methods in serial or parallel in order to filter false positives. The serial and parallel modes for LPC and MVNN are illustrated in Figure 2.

We have performed extensive tests using various combinations of algorithms in both serial and parallel mode, using the EDDIES simulation data to quantify our success.

Our results indicate that the serial mode with LPC and MVNN works best for eliminating false positive water quality changes. However, this mode is often too restrictive and may eliminate actual events as well. The general conclusion is that expert knowledge is needed to accurately identify an event as a false positive. This knowledge may be in the form of an event profile that occurs periodically at certain times of the day, or the knowledge that an increase in CL2 will produce a

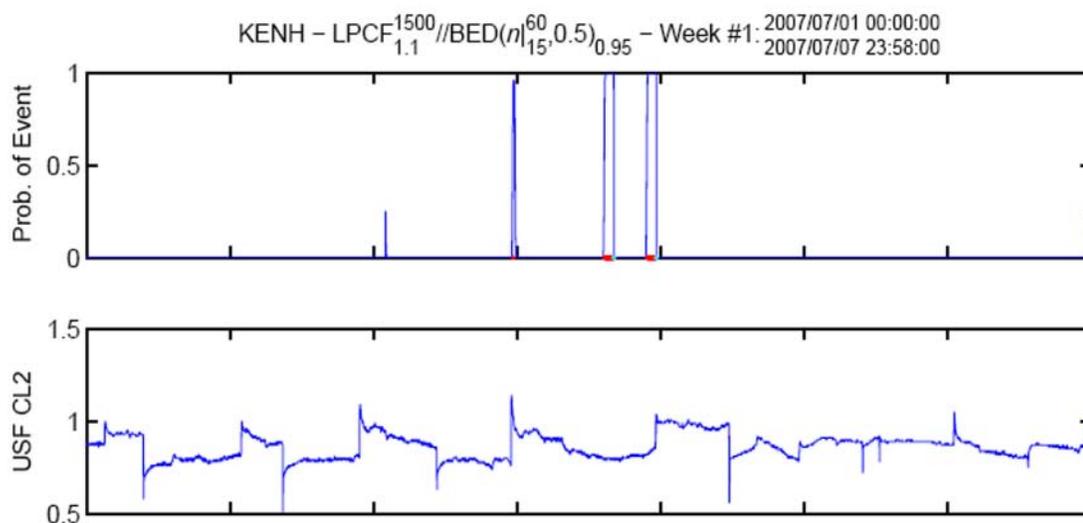


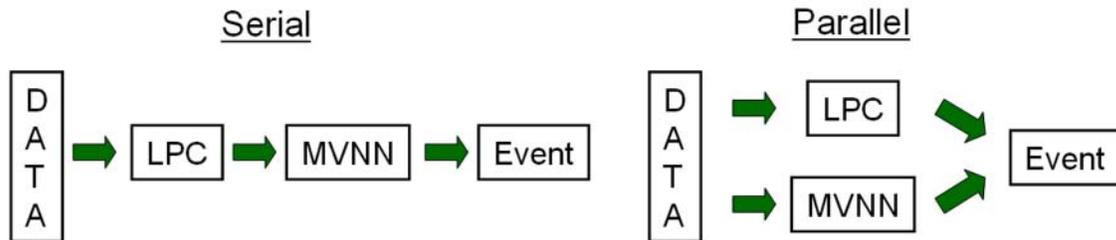
Figure 1. False positive event detected by CANARY. The first large spike in the upper plot is a false positive caused by the addition of CL2, as can be seen in the lower plot at the same time step.

known event. Future versions of CANARY will incorporate the ability to encode this type of expert knowledge.

*This work has been supported by the U.S. Environmental Protection Agency.*

### Reference

1. McKenna, S., Wilson, M., and Klise, K. (2008). Detecting Changes in Water Quality Data, *Journal of the American Water Works Association* 100(1):74-85.



**Figure 2.** Left: Example of serial mode operation of Canary, using the LPC and MVNN algorithms. Right: Example of parallel mode operation of Canary using the LPC and MVNN algorithms.

# CORESIM ENABLES LARGER LOGISTICS AND SYSTEM-OF-SYSTEMS SIMULATIONS

Contact: Scott A. Mitchell, Computer Science & Informatics, [samitch@sandia.gov](mailto:samitch@sandia.gov)

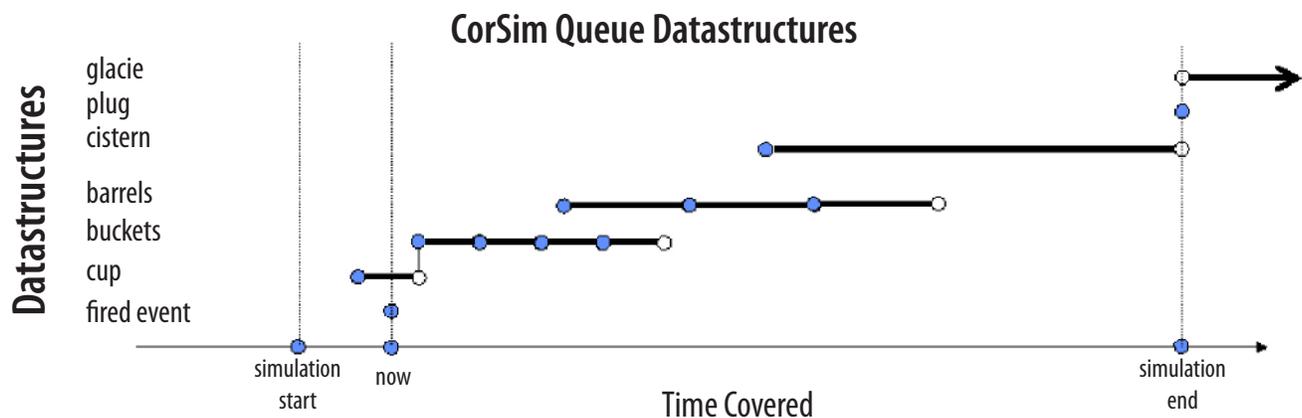
*CoreSim achieves a speedup of 3 to 5 orders of magnitude... A key simulation from a Lockheed Martin Joint Strike Fighter (JSF) project went from batch mode to interactive: its run-time decreased from over one day to a few seconds.*

The project's goal is to design and develop a discrete-event [1] logistics simulator. The intent of the "CoreSim" software engine is to replace older logistic simulation codes and become the core logistics simulator for the System of Systems Analysis Toolset (SoSAT). CoreSim is also an R&D proof-of-concept project for finding the right balance between generality and specialization, between a generic discrete event simulator and a one-application logistics simulator. Our contribution is to design and develop a new general simulation capability and optimize code performance, including designing efficient algorithms with complexity theory. Partners in Sandia's Defense System and Assessments area lead the parts of the project dealing with customer interactions, requirements gathering, and simulation summary statistics capabilities. Ultimate customers of SoSAT [2] include military and defense contractors, notably the US Army's Program Manager for Future Combat Systems (PM FCS) and the Program Executive Office for Ground Combat Systems (PEO GCS).

The primary challenge was to find the right balance between the generality and specialization of features and implementation. A logistics simulator that uses a completely generic discrete event simulation engine typically runs slowly. At the other extreme, the external community has developed some very specialized logistics simulators that can only be

used for one application. The old piece of logistics simulation code that CoreSim replaces has some fundamental design choices that, while providing generality and extensibility, severely impact code performance. CoreSim achieves a 3 to 5 order of magnitude speedup, through specialization of certain types of events, careful algorithm design and complexity analysis, as well as traditional code tuning.

CoreSim's main discrete event loop (Figure 1) achieves near-constant time enqueue and dequeue operations despite irregular event times by using a novel hierarchical variant of calendar queue [3]; see the figure for more detail. As another algorithmic speed up example, consider work-shifts. Their periodic start and end times could be handled by explicitly placing them as events on the event queue. CoreSim, however, is able to represent these events implicitly, because they occur in a predetermined manner. Another example of CoreSim's efficiency is related to scheduling, specifically, starting queued work-tasks. When a work-task completes, the resources (equipment, personnel) it was using are released, which might enable other waiting work-tasks to start. CoreSim uses sophisticated data structures to facilitate an efficient determination of which queued work-tasks, if any, should start next. (Choosing which task to start depends both on the resources and user-selected prioritization rules.)



**Figure 1.** CoreSim's main discrete event loop employs a bucketing heuristic for queued events that dynamically exploits the frequency distribution of events. This gives added efficiency by avoiding maintaining a heap of large numbers of events, and instead performing sorts over small numbers of events just before they occur.

CoreSim achieves a speedup of 3 to 5 orders of magnitude, while maintaining approximately the same degree of flexibility in the types of scenarios it can simulate, over the older logistics component it replaces. A key simulation from a Lockheed Martin Joint Strike Fighter (JSF) project went from batch mode to interactive: its run-time decreased from over one day to a few seconds. CoreSim's software was carefully designed to be extensible, meaning new features can be added in response to what scenarios and trade studies customers need. The design team is also attempting to anticipate customer needs by making new features as general-purpose as practical within the specialization and efficiency requirements. Our quick turn-around efforts have been critical to this customer-focused project's success.

### **References**

1. Schriber, T. J., and Brunner, D. T. (1997). Inside Discrete-Event Simulation Software: How it Works and Why it Matters, *Proceedings of the 1997 Winter Simulation Conference*, eds. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, (<http://www.informs-sim.org/wsc97papers/0014.PDF>).

2. For more information about SoSAT and Sandia's role in the Army's Future Combat System, contact Bruce Thompson, Manager Systems Readiness & Sustainment Technologies Department, [bmthomp@sandia.gov](mailto:bmthomp@sandia.gov).

3. Siangsukone, T., Aswakul, C., Wuttisittikulij, L. (2003). Study of Optimised Bucket Widths in Calendar Queue for Discrete Event Simulator, *Proceedings of Thailand's Electrical Engineering Conference (EECON-26)*, 6-7 November 2003, Phetchaburi. (<http://pioneer.netserv.chula.ac.th/~achaodit/paper5.pdf>).

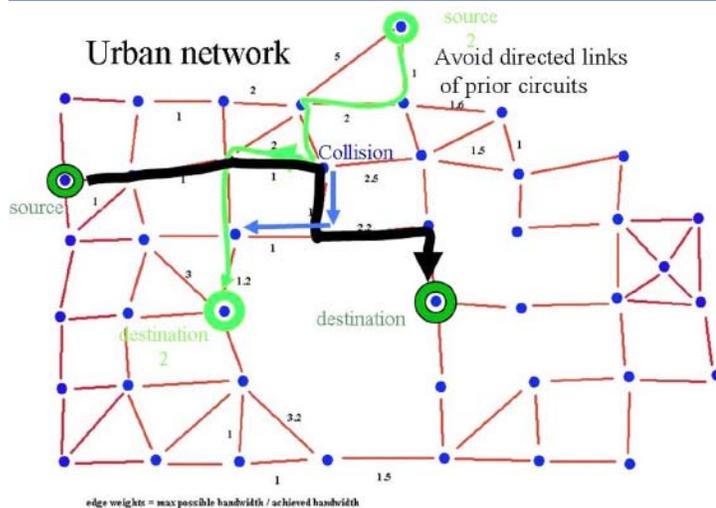
***This work has been supported by Defense Systems & Assessments (DS&A).***

# BIOLOGICAL INSPIRATION MOTIVATES NEW NETWORK PROTOCOLS

Contact: Scott A. Mitchell, Computer Science & Informatics, [samitch@sandia.gov](mailto:samitch@sandia.gov)

*The project provided a new bio-inspired protocol that met the needs of the peculiar scenario... a small military surveillance system installed on-the-fly over the public spaces of an urban environment.*

We researched and developed a new protocol for radio network communication for certain military applications. It was prototyped as a standalone program and in OPNET Modeler [1]. A conference paper was submitted [2]. Our protocol, named RatNest, was bio-inspired [3] in that it mimicked certain aspects of the structured society of naked mole rats. Our scenario of interest was nontraditional for *ad hoc* networks: a small military surveillance system installed on the fly over the public spaces of an urban environment. The dynamics of the network were limited: nodes were nonmobile, but locations were unknown, and nodes fail and recover occasionally. Communication was established in an ad-hoc fashion, meaning source and destination were unknowable in advance. But, once established, large amounts of video, audio, or other sensor data were streamed. Nodes were sparsely arranged, say at street corners, and communicated through line of sight, say, along the streets. Nodes had dedicated radio links to each of its (few) neighbors. Link quality varied somewhat.



**Figure 1.** A black circuit from source to destination is established which minimizes the maximum weight of an edge, which maximizes the throughput of the circuit. At a later time, a second circuit from source2 to destination2 is requested. Source2 has outdated information of the network state, in that it doesn't know about the first circuit, and attempts to establish a circuit that uses a directed edge of the first circuit, the blue vertical edge below "Collision." However, when this is detected at node "Collision," the protocol re-routes the second circuit using local information, using an unused backwards-edge of the first circuit, establishing the green circuit.

The protocol we designed found good routes quickly; adapted routes when local information was outdated; was robust (despite the nonrobust nodes) by repairing routes; and had low overhead (a few bits for the protocol and the rest of the bandwidth dedicated to the streamed data) (Figure 1). The protocol can find either a short-latency or a high-bandwidth route for each circuit request.

The scenario of interest was nontraditional. Many of the challenges of this project appeared in the literature and in a variety of other scenarios, but this particular combination of challenges was unique. No other work considered the scenario of *ad hoc* circuit establishment without global positioning system information, with nodes with these capabilities and limitations, and with this hierarchy of protocol goals:

1. Low routing overhead especially when a data message is in hand.
  - a. When establishing a circuit, few bits are used for routing. When sending a message, few bits are used for routing overhead and maintenance.
2. Low memory and computational requirements for each node.
  - a. The protocol should require very little nonbuffer memory for storing routing tables, network graphs, etc.
  - b. A simpler and faster routing engine is preferred.
  - c. Messages are not queued. If a circuit cannot be established (or was broken and cannot be repaired), then packets are dropped.
3. High transmission data-rates (good circuits).
  - a. For the scenario's network, this translates into low message latency and route redundancy. If packets are only occasionally sent, (e.g., sensor chirps), then low message latency is equivalent to the traditional weighted path length metric. When streaming packets in a bent-pipe mode, (e.g., streaming video where the data are large compared to the bandwidth of a link), then the bottleneck metric is more appropriate.

#### 4. Route repair or “self-healing.”

- a. In the event of a node or link failure, the protocol finds alternate routes quickly. Since packets are dropped while the route is repaired (2c), repairing the route quickly and locally is more important than finding the best-metric route.

The project provided a new bio-inspired protocol that met the needs of the peculiar scenario.

The protocol takes a unique technical approach to the structure of data at a node. Since nodes are nonmobile, each node maintains a graph model of the network in its local memory. This includes the graph of all nodes of the network, and the maximum capability of each link of that node. This “persistent” graph is annotated with transient information about which nodes are currently failed, and which links are currently dedicated to circuits. This computer science graph approach contrasts with most other protocols, which store specific route tables. The transient information is gathered proactively and also reactively. The proactive strategy mimics “soldier rats,” a fixed pool of agents traversing the network searching for quiet issues; the reactive strategy mimics “tunneling rats,” which are created by route events and quickly get information disseminated for rerouting. When requesting a circuit, the graph model at the source node is used to compute a preferred route. This route is then vetted on the actual network; when the route is not feasible,

a node close to the problem has the ability to reroute around the problem (Figure 1) or, in extreme cases, to back up and allow a prior node to compute a route. The route is pruned of these dead-ends and loops during route acknowledgement. In a similar way, if a node goes down in an already-established circuit, a neighboring node will detect this and can reroute around the problem, using local information to quickly re-establish the circuit.

We delivered a description of our research to the customer. We built a small C++ program that implements many aspects of the protocol. This protocol was wrapped and demonstrated in OPNET, the standard network simulator protocol.

#### **References**

1. OPNET Modeler, [http://www.opnet.com/solutions/network\\_rd/index.html](http://www.opnet.com/solutions/network_rd/index.html)
2. Mitchell, Scott A. (2009). The RatNest Routing Protocol for Ad-Hoc Circuits Over Fixed Radio Networks, submitted to GlobeCom.
3. Wedde, H. F., Farooq, M. (2006). A Comprehensive Review of Nature Inspired Routing Algorithms for Fixed Telecommunication Networks, *Journal of Systems Architectures*, Vol. **52**(8-9), pp. 461–484.

***This work has been supported by Defense Systems & Assessments (DS&A).***

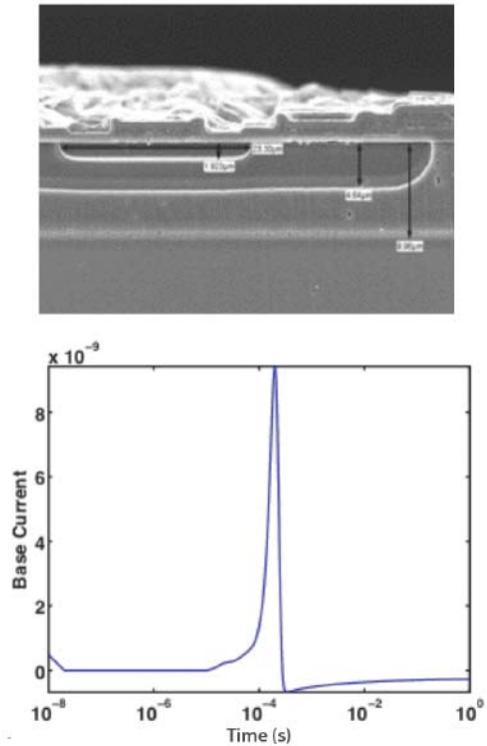
# EMBEDDED SENSITIVITY ANALYSIS ENABLES HIGH IMPACT PREDICTIVE SIMULATION

Contacts: Roger Pawlowski, Applied Math and Applications, [rppawlo@sandia.gov](mailto:rppawlo@sandia.gov)  
Eric Phipps, Optimization and Uncertainty Quantification, [etphipp@sandia.gov](mailto:etphipp@sandia.gov)

*Embedded sensitivity analysis is rapidly being incorporated into Sandia's next-generation simulation software.*

Computational simulation plays a critical role in the design, risk assessment, and qualification of complex engineering systems relevant to Sandia's missions. Examples include the design of re-entry vehicles and stable fusion reactors, the licensing of nuclear reactors and waste repositories, and the qualification of electronics in radiation environments. Predictive simulation entails significantly more analysis than merely high-fidelity simulation. Rather, it additionally requires verification of simulation code correctness, validation of the simulation's effectiveness at modeling the system, measuring the sensitivity of simulation results with respect to input data, and quantifying the effects of uncertainty in the data. Accordingly, simulation and analysis tools must provide these capabilities with reasonable execution times for predictive simulation to be feasible. Sandia researchers have developed new capabilities in "embedded" sensitivity analysis, which are advancing the state of the art in predictive simulation.

Computing derivatives provides the foundation for many analysis algorithms supporting predictive simulation. Derivatives, also called sensitivities, determine how a computation's results vary with its inputs, and are useful in (1) estimating how errors and uncertainties in those inputs affect simulation outcomes, (2) highlighting which physics or subsystems are most important in a simulation, and (3) calibrating simulations against empirical data. Thus, estimating derivatives quickly and accurately is critical for predictive simulation. Traditional sensitivity approaches for complex, large-scale simulations are based on running repeated simulations while varying inputs. While this approach is simple and convenient, it disregards much of the knowledge of the underlying system that could be leveraged for more efficient computations. Recently, Sandia developed revolutionary technologies for "embedded" sensitivity analysis based on the ideas of automatic differentiation (see section 2, "Automatic Differentiation Enables Accurate Derivative Evaluation for Predictive Simulation" for more information). Where simulations normally compute solution values, this technology re-uses the same simulation code-base but automatically produces derivatives along with the solution. These derivatives are then coupled with implicit



**Figure 1.** Top: Experimentally obtained micrograph of a stockpile bipolar junction transistor. Bottom: Simulation of the electric current flowing through the device during radiation damage.

time integration, nonlinear solvers, and linear solver technologies to create a state-of-the-art embedded sensitivity analysis capability that can be applied to even the most complex science and engineering problems.

This technology has recently been showcased [1, 2] in the simulation of radiation damage of an electrical circuit as part of the Qualification Alternatives to the Sandia Pulse Reactor (QASPR) project. Figure 1 shows a stockpile bipolar junction transistor along with computational simulation of the device under radiation. Using the embedded sensitivity technique, Sandia researchers were able to determine the parameters in the radiation damage mechanism that the operation of the device was most sensitive to (Figure 2). This improved understanding of the damage mechanisms, indicating

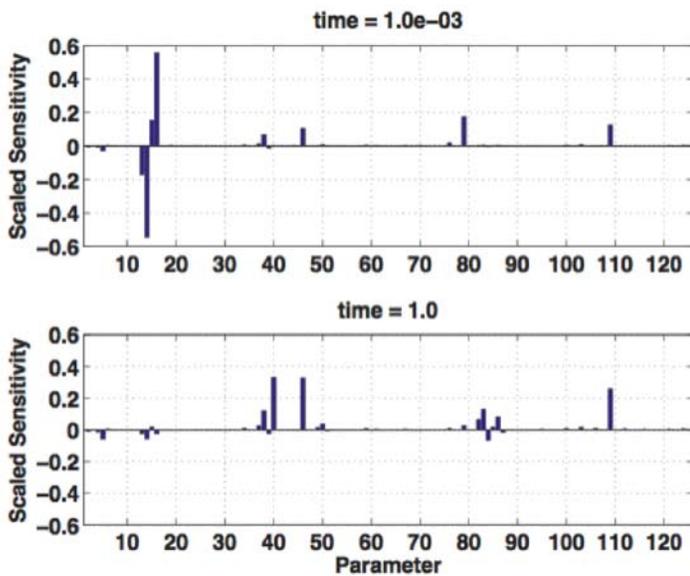


Figure 2. Plot of the scaled sensitivity of the current from Figure 1 with respect to 126 parameters in the radiation damage mechanism, at early (top) and late (bottom) times during the radiation damage. These results help focus future radiation damage mechanism research.

where further modeling efforts are required to increase simulation fidelity. This technique was demonstrated to be approximately 14 times faster than traditional techniques, and significantly more accurate and robust as well.

Embedded sensitivity analysis is rapidly being incorporated into Sandia's next generation simulation software. It has impacted a variety of internal projects including security and strategic evaluations of the national natural gas network, rail gun design, circuit network design, and fundamental magneto-hydrodynamics research. Externally, this work has fostered a strong collaboration in modeling advanced tactical lasers for the U.S. Air Force as well as potential impact to oil and gas exploration.

The embedded sensitivity analysis capabilities described here were made possible through the vertical integration of numerous embedded simulation and analysis capabilities, all of which are available through the Trilinos Software Framework [3].

### References

1. Phipps, E., Bartlett, R., Gay, D., et al. (2008). Large-scale Transient Sensitivity Analysis of a Radiation-damaged Bipolar Junction Transistor via AD, *Advances in Automatic Differentiation*, Vol. 64 of *Lecture Notes in Computational Science and Engineering*, pp. 351–362.
2. Bartlett, R., Collis, T., Coffey, D., et al. (2007). *ASC Vertical Integration Milestone*, SAND2007-5839. Sandia National Laboratories, Albuquerque, NM.
3. See <http://trilinos.sandia.gov> for more information on the Trilinos software.

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

# PERIDYNAMICS UNIFIES THE MECHANICS OF CONTINUA, FRACTURE, AND STATISTICAL MECHANICS

Contacts: Richard B. Lehoucq, Applied Mathematics and Applications, [rblehou@sandia.gov](mailto:rblehou@sandia.gov)  
Stewart A. Silling, Dynamic Multiscale Material Modeling, [sasilli@sandia.gov](mailto:sasilli@sandia.gov)

*Although peridynamics was originally developed as an engineering tool, we are now exploring the underlying mathematical theory. This exploration is both clarifying the relationship between it and the traditional theory and providing a sound basis for new applications.*

Many of the engineering problems of central importance to Sandia's national security missions for NNSA and to its external customers involve the deformation of materials and structures under extreme loading conditions. The traditional methods of continuum mechanics and their software implementations encounter limitations when applied to discontinuities such as cracks, which are of inherent importance to simulating large deformation behavior. Similar limitations arise when trying to adapt the standard version of continuum mechanics to systems of discrete particles, as occur in molecular dynamics.

Sandia researchers have developed and analyzed Peridynamics, a more general framework for modeling mechanics that allows for discontinuities. The peridynamic model, pioneered by Silling, has been under development at Sandia and with external collaborators for ten years. Peridynamics provides a mathematical framework for modeling the evolution of defects within a continuous body. This is accomplished by representing the internal force density by an integral equation rather than a partial differential equation. Mechanically, the integral equations represent direct interactions between points in a body within a predefined distance of each other, called the **horizon**. This nonlocality of force is a consequence of kinematics that employs the true

deformation instead of an approximation (e.g., strain) as in the classical theory. Figure 1 shows an example of quantitatively accurate results for unguided crack nucleation and growth for a peridynamic simulation of dynamic fracture. Similarly, the utility and appeal of peridynamics has been established through successful application of the method to a wide variety of impact mechanics problems on heterogeneous brittle materials. See the recent review paper [1] for further details and references to various papers on peridynamics.

Although peridynamics was originally developed as an engineering tool, we are now exploring the underlying mathematical theory. This exploration is both clarifying the relationship between it and the traditional theory and providing a sound basis for new applications. Some of these new insights are as follows.

1. We proved that in the limit as the horizon approaches zero, the peridynamic equations converge to the traditional differential equations [2] for smooth deformations. This result is important because it shows that, in this sense, the traditional theory is a subset of the peridynamic theory.
2. We showed that a set of discrete classical particles, such as atoms interacting through a multibody potential, can be modeled exactly as a peridynamic body. This exact

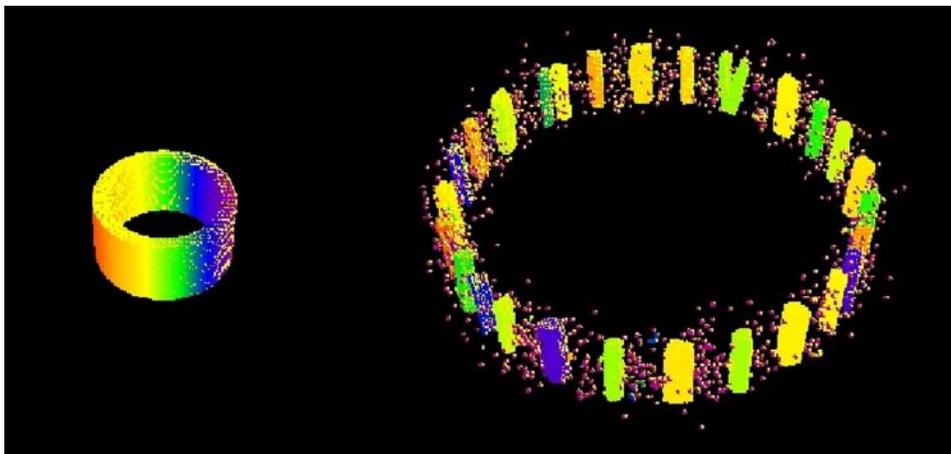


Figure 1. Peridynamic simulation of the fragmentation of an expanding cylinder.

representation is not possible within the standard theory. This result is important because it may enable us to use the peridynamic equations as a rigorous method to connect molecular dynamics with a continuum.

3. We discovered a stress tensor within the peridynamic model that has most of the same properties as the stress tensor within the traditional theory [3]. For example, the peridynamic stress tensor associates force per unit area with surfaces within a region. This result provides an additional connection between the standard and the peridynamic models.
4. Extending the seminal results of Irving & Kirkwood [4], we showed that the integral equations of peridynamics can be obtained from a statistical treatment of the distribution of a set of  $N$  particles within  $6N$ -dimensional phase space. By considering the evolution of the probability density for the system according to its Hamiltonian dynamics, the peridynamic equations emerge in the limit of a large number of particles. The results establish an important link with statistical physics.
5. We showed that any material model in the traditional theory — for example a model borrowed from a finite element code — can be adapted to the peridynamic framework [5]. This result should provide a direct way to incorporate fracture into traditional material models, (e.g., [6]).
6. We derived a peridynamic energy equation (first law of thermodynamics) that includes a more general notion of work conjugacy than in the classical theory. As with the momentum balance (see item 1 above), when the deformation is smooth, the peridynamic energy equation converges to the classical energy equation as the horizon approaches zero [7].

These recent advances may enable us to use the peridynamic theory as a mathematically consistent framework in which to coarse-grain atomistics and to couple molecular dynamics with a continuum model.

### References

1. Askari, E., Bobaru, F., Lehoucq, R. B., et al. (2008). Peridynamics for Multiscale Materials Modeling, *Journal of Physics: Conference Series*, Vol. **125**, 012078.
2. Silling, S. A., and Lehoucq, R. B. (2008). Convergence of Peridynamics to Classical Elasticity Theory, *Journal of Elasticity*, Vol. **93**, pp. 13-17.
3. Lehoucq, R. B., and Silling, S. A. (2008). Force Flux and the Peridynamic Stress Tensor, *Journal of the Mechanics and Physics of Solids*, Vol. **56**, pp. 1566-1577.
4. Irving, J. H., and Kirkwood, J. G. (1950). The Statistical Mechanical Theory of Transport Processes. IV. The Equations of Hydrodynamics, *Journal of Chemical Physics*, Vol. **18**, pp. 817-829.
5. Silling, S. A., Epton, M., Weckner, O., et al. (2007). Peridynamic States and Constitutive Modeling, *Journal of Elasticity*, Vol. **88**, pp. 151-184.
6. Warren, T. L., Silling, S. A., and Askari, E. (2009). A Non-Ordinary State-Based Peridynamic Method to Model Solid Material Deformation and Fracture, *International Journal of Solids and Structures*, Vol. **46**, pp. 1186-1195.
7. Lehoucq, R. B., Sears, M. P., Silling, S. A., Peridynamic Balance Laws, in preparation.

***This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

# REDEFINING CHEMICAL COMPOUND SPACE FOR MORE EFFICIENT MATERIALS DESIGN ALGORITHMS

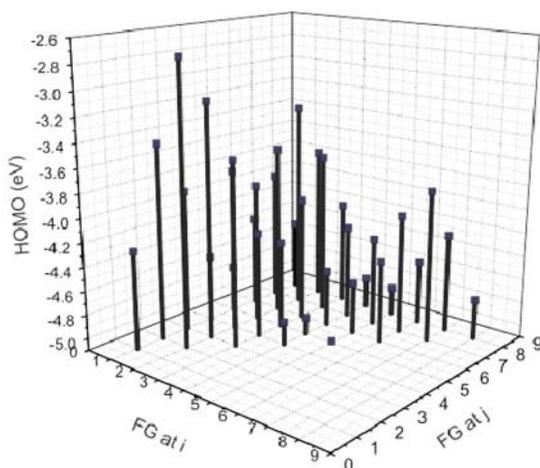
Contacts: Anatole von Lilienfeld, Multiscale Dynamics Materials Modeling, [oavonli@sandia.gov](mailto:oavonli@sandia.gov)

*It appears that most of matter's property diversity originates not in the diversity of the chemical building blocks...but rather in their specific combination and spatial arrangement.*

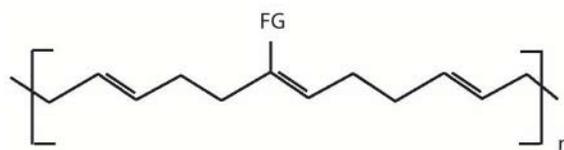
It is a holy grail in computational materials design research to engineer the chemical composition of compounds such that they exhibit desired physical properties. It appears that most of matter's property diversity originates not in the diversity of the chemical building blocks, dubbed functional groups (FGs) — the smallest FG being a single atom — but rather in their specific combination and spatial arrangement.

The vast number of possible compounds has led to the usage of the term chemical compound space (CCS) as the set comprising all stable compounds. Albeit rarely defined rigorously, CCS is routinely being 'explored' when performing materials or drug design. Due to its gigantic cardinal number; however, the simple enumeration of CCS, let alone screening, for the purpose of identifying compounds with desired properties is beyond the means of any computational power. A modest illustration of CCS for only two "sites", being decorated with only eight possible FGs, is given in Figure 1 for an electronic property of a polymer.

This has motivated the proposition of more "rational" design methodologies such as discussed in [1, 2]. Commonly, these approaches attempt to recast the exploration of CCS



**Figure 1.** Total CCS of a property of interest. Here, the highest molecular orbital (HOMO) eigen-value for cis (i) and trans (j) sites in poly-butadiene (see Figure 2) being decorated with FGs from Table 1.



**Figure 2.** Example of a poly-butadiene unit cell with three monomers. Here, FG replaced a hydrogen atom at one of the  $sp^2$ -hybridized carbon atoms.

in terms of a property optimization problem, using standard optimization algorithms, based on gradients or evolution. While the vastness of CCS certainly poses convergence problems for stochastic algorithms, all gradient based algorithms presented so far suffer (a) from the fact that a measure of distance in CCS is not known *a priori*, and (b) that the first order derivatives can be insufficiently accurate to correctly predict mutations of FGs.

We have resolved these two challenges by redefining CCS and derivatives therein: (a) We work directly with pairs of FGs, thereby circumventing the need to define proximity in CCS *a priori*; and (b) we identify and tabulate transforming paths, coupling any two FGs, that are constrained to linearize the property of interest (and not the Hamiltonian, as is conventionally done), thereby enhancing the accuracy of first order derivatives as estimators of the property's response due to mutation.

**Table 1.** Set of functional groups and corresponding index.

| Index | 1               | 2               | 3  | 4 | 5                | 6               | 7  | 8  |
|-------|-----------------|-----------------|----|---|------------------|-----------------|----|----|
| FG    | CH <sub>3</sub> | NH <sub>2</sub> | OH | F | SiH <sub>3</sub> | PH <sub>2</sub> | SH | Cl |

We have demonstrated our scheme for the highest occupied molecular orbital (HOMO) eigen-value (molecular Fermi level), which is very relevant to photo-excitation, and electrochemical properties. Within conceptual density functional theory, the HOMO eigen-value can be obtained as the derivative of the potential energy with respect to number of electrons via finite difference. As a system, we chose poly-butadiene (Figure 2) with variable FG sites, featured in Table 1, replacing H atoms at  $sp^2$ -hybridized carbons.

We have linearized the paths between all pairs  $\{i,j\}$  of FGs in Table 1. by minimization of the deviation from linearity through variational optimization of parameter  $a$  in our coupling quantum mechanical Hamiltonian,  $H(\lambda) = [H_{FG_i} + \lambda \times (H_{FG_j} - \lambda H_{FG_i})] \times f(\lambda, a)$ . Here,  $f$  has been assumed to be of quadratic order in  $\lambda$ , the other two of its three coefficients  $a$ ,  $b$ , and  $c$  being defined through the boundary condition that  $f(\lambda = 0) = f(\lambda = 1) = 1$ .

In Figure 3, the iterative optimization of  $a$  is illustrated for transforming  $FG_2$  ( $NH_2$ ) into  $FG_3$  ( $OH$ ). After successive linearization of the path, the deviation from perfect prediction of the change in HOMO eigen-value due to mutation could be decreased from 2.0 eV to 0.7 eV, with further improvements in the linearization techniques ahead of us. Figure 4 displays corresponding predictions using linearized paths between all FGs. In Figure 4, the mean average deviation of predicted HOMO eigen-value from the actual eigen-value is decreased from 1.99 to 0.69 eV when using the derivatives with respect to the linearized path.

We have presented numerical evidence how to drastically improve the accuracy of gradients in CCS, enabling accurate predictions of changes in properties due to mutation. Leveraging these advances for the implementation of gradient based property optimization algorithms within materials design software will thereby allow reduction from the formal

scaling of a full screening,  $N_f^{N_s}$ , not only to  $N_s \times N_f \times N_{fi}$  (as in evolutionary property optimizations with  $N_f$  generations), but even down to  $N_f' \cdot N_f \sim N_f'$ , and  $N_f$  and  $N_s$  being the number of FGs and the number of sites, respectively.

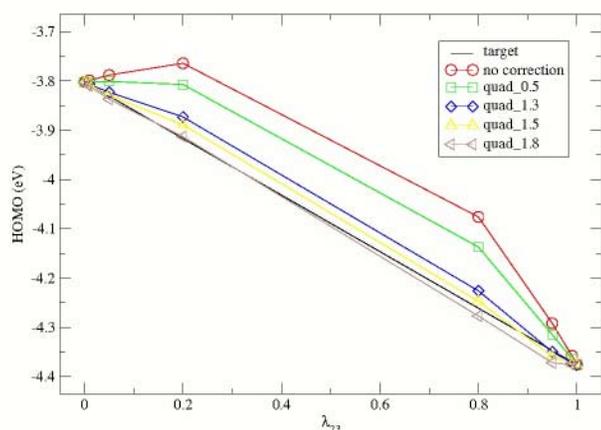
Since CCS is so mind-bogglingly large and diverse, any improvement of efficiency in the design software promises to greatly enhance high-performance computing capabilities to routinely present experimentalists with new and promising materials candidates for properties of interest.

A patent application was filed for this work. We are currently finalizing this project and plan to submit a manuscript for publication in the peer-reviewed literature.

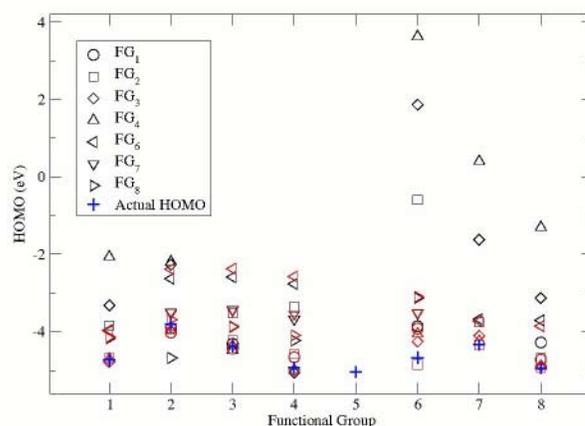
### References

1. Franceschetti, A., and Zunger, A. (1999). The Inverse Band-Structure Problem of Finding an Atomic Configuration with Given Electronic Properties, *Nature*, Vol. **402**, 60.
2. von Lilienfeld, O. A., Lins, R., Rothlisberger, U. (2005). Variational Particle Number Approach for Rational Compound Design, *Physical Review Letters*, Vol. **95**, 153002.

***This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.***



**Figure 3.** Iterative linearization of the path connecting the HOMO eigen-values of  $FG_2$  ( $NH_2$ ) with  $FG_3$  ( $OH$ ). For  $a = 0$ , the linearizing function is turned off (red), and then increasingly turned on,  $a = 0.5$  (green),  $a = 1.3$  (blue),  $a = 1.5$  (yellow),  $a = 1.8$  (brown) until convergence to the linear target line.



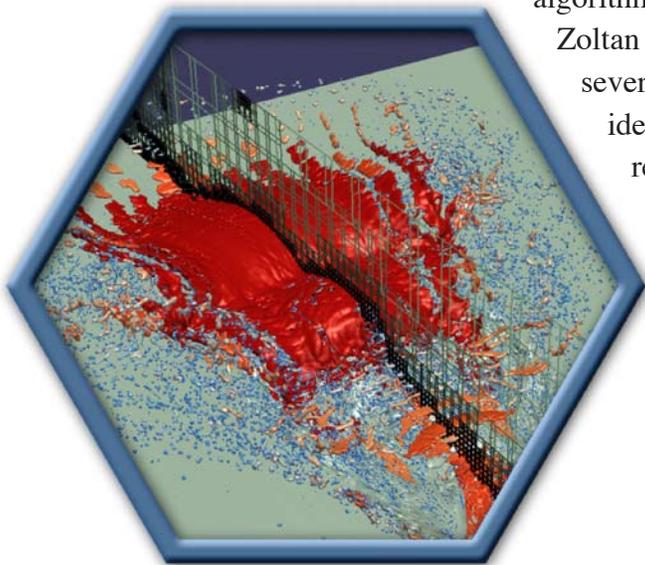
**Figure 4.** Actual (blue cross) and predicted HOMO eigen-values before (black) and after (red) linearization for poly-butadiene. The symbol indicates from which FG the prediction has been made. Predictions of HOMO eigen-values associated to  $FG_5$  ( $SiH_3$ ) were omitted due to technical issues.

*End of Section I*



# *ENABLING TOOLS FOR PREDICTIVE SIMULATION*

Tools are specific implementations of specific algorithms or classes of algorithms that support broad usage of these algorithms for a variety of purposes. This includes further algorithm and applications development, and aiding architecture research. While not necessary, the tool implementation of an algorithm increases focus on its production quality and on the usability of computational technology within the implemented algorithm(s). The implementation and availability of tools represent a relatively delicate decision process within the Computation, Computers, Information and Mathematics (CCIM) Center 1400, that is, balancing the needs of broad usage, the cost of support, and forecasting the future evolution of significant algorithm research and development within CCIM). Tools also represent major intellectual capital within CCIM and an enabling partnership of people with distinct skill sets. Major tools sets that have been created, supported, and evolved by CCIM algorithm and mathematics R&D tend to reflect the broad algorithm classes that we have seen in the first section. These include Trilinos (parallel linear algebra and uncertainty quantification algorithms, which is in wide use within Sandia and out); DAKOTA (a concentration point for continuous optimization and uncertainty quantification algorithms, which is a unique capability within NNSA); Zoltan (dynamic load balancing); and Paraview (one of several graphics tools). Finally, implementations that we identify as tools capture an element of the stewardship responsibilities that CCIM brings to the enablement of high-performance computing for the nation.



# DAKOTA TOOLKIT ENABLES LARGE-SCALE OPTIMIZATION AND UNCERTAINTY QUANTIFICATION FOR SANDIA'S MISSIONS

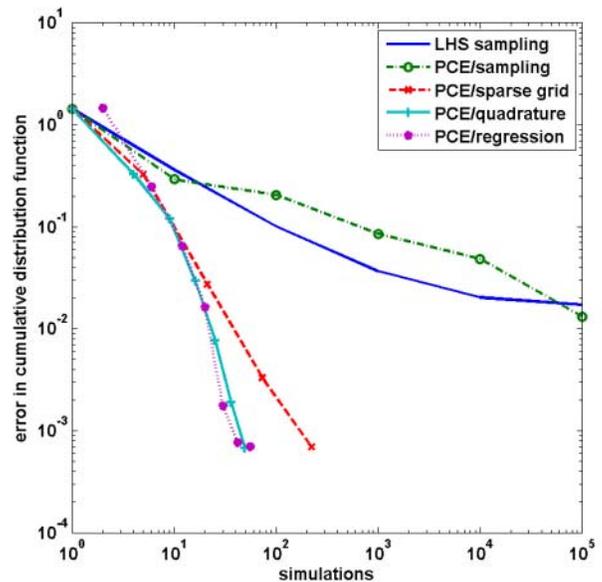
Contact: Brian Adams, Optimization and Uncertainty Quantification, [briadam@sandia.gov](mailto:briadam@sandia.gov)

*DAKOTA methods are central to Sandia's verification and validation (V&V) and QMU strategies, and will be used with full-system models in support of nuclear weapon safety and surety assessments.*

Solutions to fundamental science and engineering problems increasingly rely on iterative analysis with computational models (simulations) for the phenomena of interest. When using simulations, decision makers need answers to the following questions: What are the crucial factors/parameters and how do they affect key metrics (sensitivity)? How safe, reliable, robust, or variable is my system (quantification of margins and uncertainty, QMU)? What is the best performing design or control (optimization)? What models/parameters best match experimental data (calibration)? The Design and Analysis toolKit for Optimization and Terascale Applications (DAKOTA) [1] is a framework with diverse algorithms designed for such iterative analysis challenges in support of safety, suitability, or optimality assessments.

DAKOTA unifies time-tested and cutting-edge algorithms for sensitivity analysis, optimization, and uncertainty quantification (UQ) in a single package. It enables best-in-class mixed deterministic and probabilistic analysis with nearly any computational model. DAKOTA strategies combine methods in novel approaches, including multi-level, hybrid, and hierarchical optimization or nested uncertainty analysis for second-order probability, which treats epistemic and aleatory uncertainties encountered in QMU. Strategies also enable optimization/model calibration under uncertainty and uncertainty of optimal designs, both crucial to risk-informed decision making. Such DAKOTA methods are central to Sandia's verification and validation (V&V) and QMU strategies, and they will be used with full-system models in support of nuclear weapon safety and surety assessments.

Users can rapidly couple DAKOTA to any simulation through a universal file system-based interface. When performance or integration is paramount, DAKOTA can be directly embedded into other software (e.g., the Trilinos algorithms framework or the Xyce electrical circuit simulator) (Figure 1). Both coupling modes can leverage large-scale parallelism, where DAKOTA, the application, or both run on multiple processors. This unique multilevel



**Figure 1.** Traditional sampling-based probability calculations cannot compete with the efficiency of advanced stochastic expansion methods. For the “log ratio” benchmark problem, the plot depicts exponential convergence for polynomial chaos expansions (PCE) constructed with advanced multidimensional integration techniques, compared with the drastically slower convergence typical of LHS sampling approaches.

parallelism addresses simulation cost through distributed, simultaneous model evaluations. For particularly noisy or computationally expensive simulations, DAKOTA's extensive collection of surrogate models facilitates efficient optimization or UQ with comparably few model runs.

DAKOTA has long been a research platform and delivery vehicle for advanced optimization algorithms, including gradient-based, derivative-free, Pareto, and surrogate-based methods. These address challenges inherent in model-based engineering analysis and design, including mixed-variable models and nonsmooth, discontinuous, or multimodal outputs. Recent research efforts have focused on complementary efficient, robust methods for uncertainty quantification. Monte Carlo methods such as Latin hypercube sampling (LHS) are robust and straightforward, but require prohibitively large sample sizes to resolve tail

probabilities of interest in QMU. Structural reliability methods require comparatively few model evaluations, but can be misleading or inaccurate.

New efficient global reliability analysis and stochastic expansion methods, developed in DAKOTA, bridge this gap with a mix of efficiency and accuracy. Stochastic expansion methods can resolve response probability distributions considerably more efficiently than can traditional sampling-based approaches [2], as demonstrated in Figure 1. These advances enable uncertainty-quantified predictive simulation and efficient, robust analysis and design of critical systems in the presence of uncertainty. DAKOTA UQ algorithms are a crucial component of qualification alternatives to the Sandia pulse reactor (QASPR). They help quantify experimental, configuration, parametric, and model uncertainty, so computational models can supplant infeasible physical radiation tests.

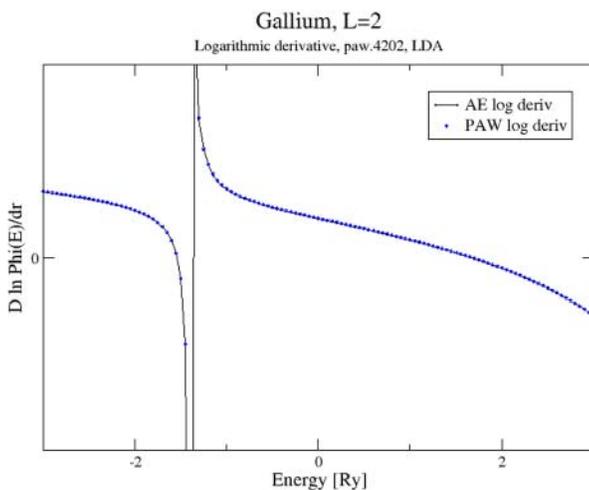


Figure 2. Optimal PAW function for gallium determined by coupling DAKOTA to the SOCORRO density functional theory (DFT) code.

DAKOTA continually impacts basic science and discovery. Researchers are using its genetic algorithms for novel calibration of quantum mechanics models, enabling first-principles simulation of new materials and semiconductors. Most Sandia quantum mechanics calculations utilize a formulation of density-functional theory wherein the atomic electrons that participate in bonding are treated explicitly, while the nucleus and the remaining electrons are represented with a set of functions that reproduce the electron-scattering features of this ionic configuration. While the norm-conserving pseudo-potential form of these functions can be used, Sandia research and development programs prefer projector-augmented wave (PAW) functions. These permit a smaller basis set to be used for most atoms in the periodic table, more accurate electron scattering features, and explicit treatment of more atoms.

The Sandia-developed SOCORRO code can use the PAW method, but its use has been limited by lack of publicly available PAW functions, which are typically determined through hand calibration.

In response, DAKOTA is being used to guide the optimization of PAW functions, in (to the researchers' knowledge), a first-ever attempt to automate this process [3]. They use DAKOTA in conjunction with the Wake Forest University ATOMPAW code and SOCORRO to assess optimality based on electron scattering properties of the PAW functions, computed properties of various materials, and the efficiency of the SOCORRO calculations used to obtain the material properties. DAKOTA algorithms have explored regions of parameter space unlikely to have been examined by hand calibration, yielding optimal PAW functions (as shown in Figure 2). While this development is still at an early stage, these functions are being used to study gallium arsenide defects for the QASPR program and eventually group-III nitride semiconductor defects for a Basic Energy Sciences program.

Version 4.2 of the DAKOTA software toolkit was released and deployed in November 2008 and is freely available under the GNU GPL. Over 4000 download registrations have been received from DOE laboratories, as well as other government, academic, and corporate institutions.

### References

1. Eldred, M. S., et al. *The DAKOTA Project: Large-Scale Engineering Optimization and Uncertainty Analysis*, Sandia National Laboratories, Albuquerque, NM. <http://www.cs.sandia.gov/dakota/>.
2. Eldred, M. S., and Burkardt, J. (2009). Comparison of Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Quantification, *Proceedings of the 47th AIAA Aerospace Sciences Meeting*.
3. Snow, R. J., Wright, A. F., and Fong, C. Y. (2009). Projector Augmented Wave Database with Automatic Parameter Optimization, Abstract X13.00008, *MAR09 Meeting of the American Physical Society*.

***This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

# AUTOMATIC DIFFERENTIATION ENABLES ACCURATE DERIVATIVE EVALUATIONS FOR PREDICTIVE SIMULATION

Contacts: Eric Phipps, Optimization and Uncertainty Quantification, [etphipp@sandia.gov](mailto:etphipp@sandia.gov)  
David Gay Optimization and Uncertainty Quantification, [dmgay@sandia.gov](mailto:dmgay@sandia.gov)

*Through a new package called Sacado, we have developed AD tools that are specifically designed for [large engineering] codes. ... Sacado eliminates the need for QASPR developers to hand-code and verify derivatives, directly contributing to their ability to develop complex physics simulations in a short period of time*

A crucial, yet often overlooked, component of computational simulation and analysis is computing derivatives. Derivatives are often needed in time integration, equation solving, optimization, sensitivity analysis, stability analysis, and uncertainty quantification, all of which lie at the heart of predictive simulation. The efficiency and robustness of these algorithms depend on how quickly and accurately derivative quantities can be estimated, which in turn strongly affects the fidelity and complexity of systems that can be simulated. Computing the derivatives that these algorithms need is nontrivial. For example, the staple of many simulation algorithms is Newton's method, which requires evaluating the system's Jacobian matrix of first partial derivatives. It is up to the computational scientist to derive and implement the Jacobian's evaluation.

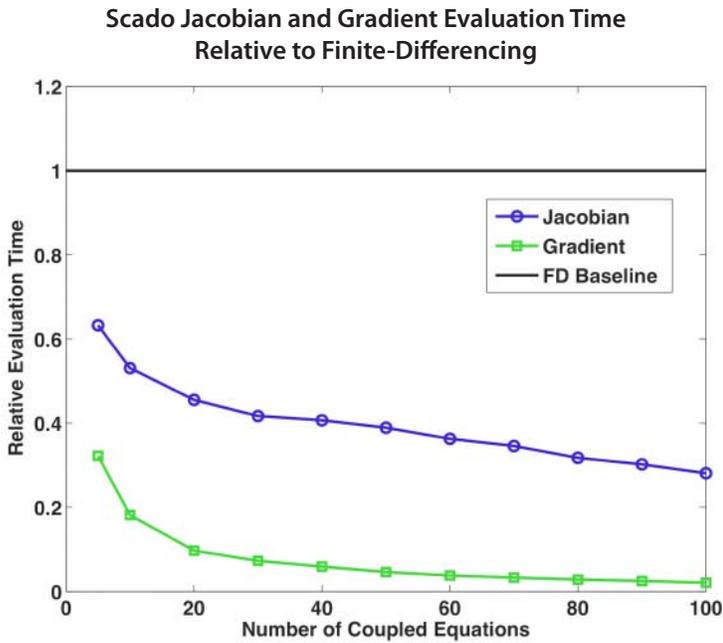
Unfortunately, this process is often laborious and error-prone, and it distracts the code development team from implementing complicated physics. This slows the pace of simulation code development and makes verifying the code's correctness difficult. Therefore, scientists often resort to numerical, finite differencing. This eliminates the need to derive, implement, and verify derivatives at the expense of reduced accuracy resulting from inherent truncation and round-off errors. Moreover, this approach is not efficient when computing derivatives with respect to many quantities, such as in optimization of complex physics. Sandia's Sacado tool is an important step forward in automating this process and simplifying the utilization of advanced simulation capabilities.

Automatic differentiation (AD) is an alternative technology that provides accurate first and higher derivatives without requiring manual hand-coding. It works by recognizing that any differentiable computation implemented in a simulation code must be composed of a basic set of elementary operations with simple derivative formulas. For each statement in a computer program, AD uses these rules to generate derivative statements that are pieced together using the Chain Rule, resulting in a modified

program that also computes derivatives. The AD derivative calculation is just as accurate as if it was hand-coded, and there are no differencing errors. This technique has been known for many years, and there are several tools available for implementing it [1]. However, until recently, AD has had little impact on Sandia's computational simulation efforts because it was widely believed that Sandia's large engineering codes were too complex for AD to be effective.

Through a new package called Sacado, we have developed AD tools that are specifically designed for such codes. Sacado is part of the larger Trilinos [2] suite of scientific computing tools. The new tools exploit features of the modern C++ programming language to easily integrate AD into simulation codes in a way that scales to complex physics code development. They work by replacing the code's basic floating-point data type with special AD data types that call overloaded versions of the basic elementary operations to produce derivatives as a side effect. Sacado leverages the "templating" features of C++ to automate type changes, and they use such techniques as "expression templates" to make the resulting derivative code very efficient. Moreover, the Sacado tools can exploit code structure for even greater efficiency. For example, one often needs to compute gradients with respect to a very large number of quantities in complex optimization problems. In these cases, Sacado uses the technique of "reverse accumulation" to compute the derivatives in reverse order, making gradient evaluations much cheaper than otherwise would be possible.

The Sacado tools have impacted several of Sandia's important predictive simulation projects, particularly the Qualification Alternatives to the Sandia Pulsed Reactor (QASPR) electronic device qualification project. Sacado eliminates the need for QASPR developers to hand-code and verify derivatives, directly contributing to their ability to develop complex physics simulations in a short period of time [3]. The appeal of just programming the governing equations has made Sacado attractive for new development



**Figure 1.** Comparison of Sacado Jacobian and gradient calculations to traditional finite-differencing as the number of coupled equations is varied in a simulation.

efforts as well, including plasma transport and reaction simulations for semiconductor processing, simulations for Z-pinch modeling, and reacting-flow simulations for chemical laser applications. Additionally, the Sacado derivative calculations are much more efficient than numerical finite differencing (Figure 1) and significantly more accurate. Moreover, having access to fast and accurate derivative capabilities has opened new doors for advanced analysis algorithms research in support of predictive simulation. Researchers recently completed an important milestone demonstrating a 14x speedup of embedded transient sensitivity analysis over traditional numerical differencing on a QASPR simulation (see “Embedded Sensitivity Analysis Enables Predictive Simulation” p. I-54 for more information). Along with other Trilinos software, this approach exploits the fast and accurate parameter derivatives provided by Sacado. Researchers are now building on this work to provide even more advanced analysis capabilities such as adjoint sensitivity analysis, optimization, and uncertainty quantification. The Sacado AD tools provide scalable embedded derivative machinery to simulation codes, enabling predictive simulation research and development efforts that have never been possible at Sandia before.

### References

1. See <http://www.autodiff.org> for information on automatic differentiation tools, techniques, and applications.
2. The Trilinos tools, including Sacado, are available at <http://trilinos.sandia.gov>.
3. Bartlett, R., Gay, D., and Phipps, E. (2006). Automatic Differentiation of C++ Codes for Large Scale Scientific Computing, *Computational Science – ICCS 2006, volume 3994 of Lecture Notes in Computer Science*, eds. V. N. Alexandrov, G. D. van Albada, P. M. A. Sloot, and J. Dongarra, 525–532. Springer.

*This work has been supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program.*

# MESHING AND GEOMETRY FOR MODELING AND SIMULATION WITH CUBIT

Contacts: Steven Owen, Computational Modeling Science, [sjowen@sandia.gov](mailto:sjowen@sandia.gov)  
 Brett Clark, Computational Modeling Science, [bwclark@sandia.gov](mailto:bwclark@sandia.gov)  
 Byron Hanks, Computational Modeling Science, [bwhanks@sandia.gov](mailto:bwhanks@sandia.gov)  
 Roshan Quadros, Computational Modeling Science, [wrquad@sandia.gov](mailto:wrquad@sandia.gov)  
 Ted Blacker, Computational Modeling Science, [tdblack@sandia.gov](mailto:tdblack@sandia.gov)

*CUBIT is a vital tool used by engineers within the Nuclear Weapons Complex (NWC) for the design, verification and validation of components within the stockpile. It is used on a daily basis by NWC scientists and engineers in computational simulation to ensure the safety and longevity of the U.S. nuclear stockpile.*

The CUBIT Geometry and Meshing Toolkit, developed at Sandia National Laboratories in the early 1990s, has had a tremendous impact on how computational modeling and simulation is done within the nuclear weapons complex (NWC). The problem of reducing cost and improving efficiency of the engineering design cycle through modeling and simulation is often hampered by the bottleneck of preparing and meshing the geometry. Geometry preparation and mesh generation remain one of the most time consuming, expertise intensive, and problem specific tasks of simulation for users of these technologies. Estimates [1] have placed geometry preparation and mesh generation at seventy percent of the time needed to perform analysis. CUBIT addresses these issues by developing tools that make the process of geometry preparation and mesh generation more efficient.

In addition to being a highly usable tool, used by hundreds, CUBIT serves as a common research and development platform for new meshing and geometry technology. Beginning with the award winning paving algorithm for unstructured quadrilateral meshing, it has grown into a comprehensive environment for geometry preparation and mesh generation. Including state-of-the-art algorithms and methods for all-hexahedral mesh generation, CUBIT also has a list of meshing and geometry capabilities that rival and, in many cases, exceed any capability currently available in industry.

CUBIT is a vital tool used by engineers within the nuclear weapons complex for the design, verification, and validation of components within the stockpile. It is used daily by NWC scientists and engineers in computational simulation to ensure the safety and longevity of the U.S. nuclear stockpile. Figure 1 shows example hexahedral meshes of components that have recently been modeled using CUBIT's geometry and meshing capabilities that have directly contributed to stockpile stewardship.

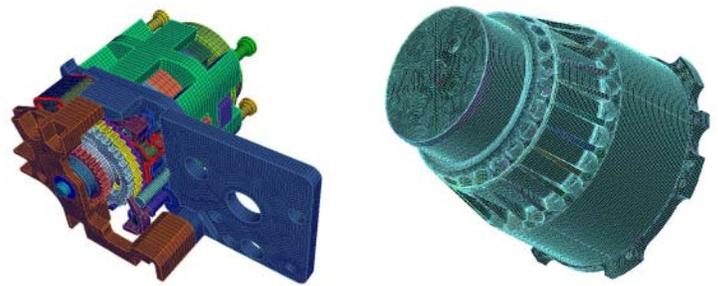


Figure 1. All-hexahedral meshes generated with CUBIT that have been used for stockpile stewardship programs.

In addition to the Department of Energy's Advanced Simulation and Computing (ASC) program, Corporate Research and Development Agreements (CRADAs) and other partnerships have fueled many advances within the CUBIT code. One example is the long-term CRADA relationship with Goodyear in which CUBIT's meshing technology has become an integral part of Goodyear's corporate solution for modeling and simulation. Recently, CUBIT has been tightly integrated with CATIA, a commercial computer-aided design software application used in the engineering design process at Goodyear. Using CUBIT with CATIA enables Goodyear to perform engineering simulations and "what-if" studies on a broad range of tire engineering problems. Other partnerships include Caterpillar and a new initiative with the Department of Defense to provide geometry and meshing capability for the design of U.S. Navy ships, air vehicles, and radio-frequency devices.

Figure 2 illustrates the impact of CUBIT by showing measured usage of the program both internally at Sandia and externally. These figures track the monthly repeat users over a six-year period. To be included in these metrics, a user must execute CUBIT at least twice in any given month. While internal Sandia usage has nearly doubled over the past few years, the external usage has had a dramatic increase.

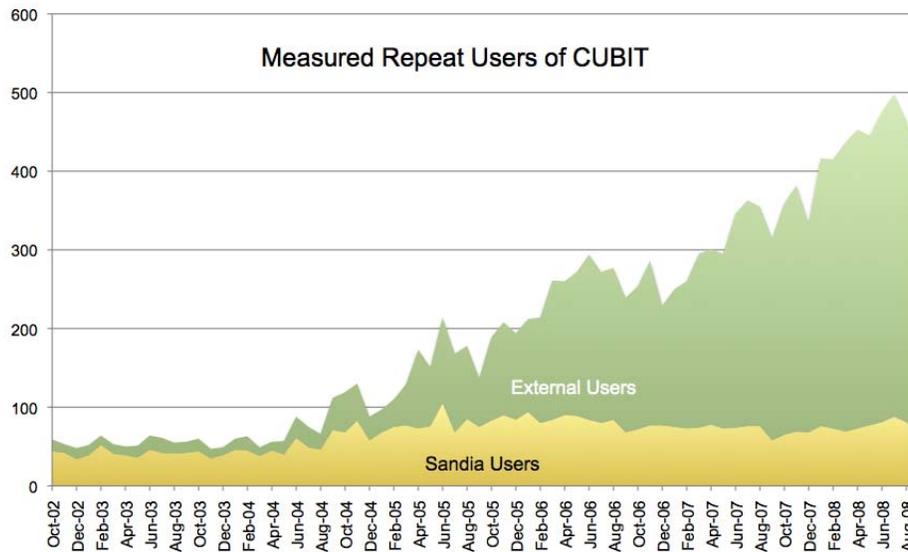


Figure 2. Measured repeat users of CUBIT over a six year period starting in 2002.

Monthly repeat users are now approaching 500. Although these numbers reflect only government and academic usage made available via a minimal distribution fee, CUBIT has also been made available for commercial licensing through an external nonexclusive distribution agreement. Recent commercial sales figures have shown a growing interest in CUBIT's geometry and meshing capabilities within industry.

The CUBIT development team recently released version 11.1, which included a variety of new state-of-the-art geometry and meshing capabilities. Some of the principal new capabilities recently released or in development include the following:

- a. The Immersive Topology Environment for Meshing [2]: A wizard like environment for guiding the user through the geometry cleanup and meshing process for all-hexahedral meshes.
- b. Geometry Tolerant Mesh Generation: The ability to build an unstructured mesh on a less-than-ideal geometry representation without the need for geometry cleanup or defeaturing.
- c. Tolerant Imprinting [3]: The ability to tolerantly match adjoining assembly parts to ensure continuous mesh connectivity between components.
- d. CUBIT Adaptive Meshing Algorithm Library (CAMAL): meshing algorithm components usable in third-party applications via a convenient API.
- e. Many-to-Many Sweeping: Auto-decomposition of some classes of geometry to permit an all-hexahedral mesh.
- f. Ongoing All-Hexahedral Meshing Research [4]: Including techniques in unconstrained plastering, conformal mesh

interface matching, hexahedral mesh manipulation to improve quality, grid-based methods, and others.

See <http://cubit.sandia.gov> for more information on CUBIT, including complete online documentation.

#### References

1. Hardwick, Mike (2005). *DART System Analysis, Presentation to Simulation Sciences Seminar*, Sandia National Laboratories, Albuquerque, NM,
2. Owen, Steven J., Clark, Brett W., Melander, Darryl J., et al. (2007). *An Immersive Topology Environment for Meshing, Proceedings, 16th International Meshing Roundtable*, pp. 553-577.
3. Clark, Brett W., Hanks, Byron W., and Ernst, Corey D. (2008). *Conformal Assembly Meshing with Tolerant Imprinting, Proceedings, 17th International Meshing Roundtable*, pp. 267-280.
4. Ledoux, Franck, and Shepherd, Jason (2008). *Topological Modifications of Hexahedral Meshes via Sheet Operations: A Theoretical Study*, submitted to *Engineering with Computers*.

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program as well as Work for Others agreements, including Goodyear Tire and Rubber Company.***

# STRUCTURAL SIMULATION TOOLKIT ALLOWS EARLY INSIGHT INTO PERFORMANCE OF MICROSCALE SUPERCOMPUTER SIMULATION

Contacts: Arun F. Rodrigues, Scalable Computer Architecture, [afrodri@sandia.gov](mailto:afrodri@sandia.gov)  
 Richard C. Murphy, Scalable Computer Architecture, [rcmurph@sandia.gov](mailto:rcmurph@sandia.gov)  
 K. Scott Hemmert, Scalable Computer Architecture, [kshemme@sandia.gov](mailto:kshemme@sandia.gov)  
 Brian W. Barrett, Scalable Computer Architecture, [bwbarre@sandia.gov](mailto:bwbarre@sandia.gov)

*The SST enables researchers to explore performance of design hardware running complex simulations virtually, all before funds and resources are committed. It has already had dramatic impact on the success of supercomputer deployments.*

Cost and complexity in supercomputer design are high, innovation and uncertainty are extreme, simulation codes to be run are diverse—potential unknowns are everywhere. Sandia has responded to this high risk with a modeling toolkit (Structural Simulation Toolkit or SST). The SST enables researchers to explore performance of design hardware running complex simulations virtually, all before funds and resources are committed. It has already had dramatic impact on the success of supercomputer deployments. It has also identified significant issues in scaling multicore hardware and pinpointed differences between current performance tests and real-world applications.

The performance characteristics of high-performance computers are continuously evolving. Satisfying the ever-greater demands of computer-based simulation requires supercomputers of unparalleled power and complexity. The emergence of new ways of organizing hardware [such as multicore), new architectures (like IBM's CELL), and new programming models] makes it difficult to ensure that future machines will meet the needs of future applications. Because of the extreme cost of deploying a high-end supercomputer, as well as the high cost of developing the complex multi-physics codes that run on them, it is important to have tools that provide insight into how novel hardware and software will interact.

Sandia has developed the Structural Simulation Toolkit (Figure 1) to provide early insight into how future supercomputers will perform. The SST is a whole system simulator that can model complicated applications and libraries, advanced processors (such as the proposed Light Weight Processor), multilevel memory hierarchies, and high-speed system interconnection networks. This ability to support cycle-by-cycle simulation of key hardware components allows us to evaluate proposed architectural features [like full/empty bits (FEB) in memory], identify current performance bottlenecks, and research novel solutions. The capability of running entire programs,

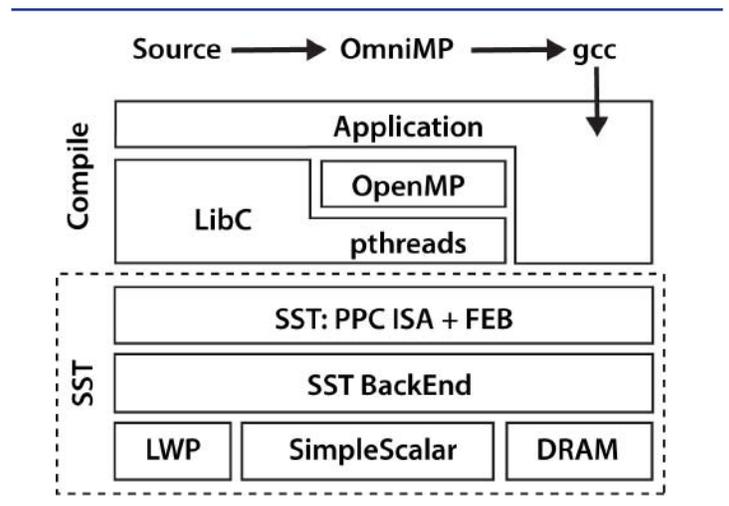


Figure 1. SST's whole system simulation model captures software and hardware performance.

including the complex communication and support libraries they require, gives us the means to experiment with new programming models and to optimize for new hardware.

To model a supercomputer, detailed models must be constructed for each critical component. These models must then be connected together in a flexible and generic manner to allow the user to explore a range of hardware options. To do this, the SST project has developed Enkidu, a discrete event simulation package that coordinates interactions between the different hardware components of a system and has a modular structure (Figure 2) that allows different sets of components to be quickly connected and driven with a variety of inputs. This flexibility allows the exploration of a diverse set of problems:

- The SST has been used to model the network used on Sandia's Red Storm supercomputer. These simulations helped identify bottlenecks and also reveal how future enhancements to network interfaces and components could improve performance (Figure 3).

- The simulator has been used to design and test a hardware matching engine that could dramatically accelerate message processing in high-speed supercomputers. This matching engine also removes work from the computer's CPU, freeing it to perform other calculations. Several supercomputer vendors have expressed interest in adopting this technology.
- Analysis of several engineering codes used at Sandia, performed with the simulator, has revealed that the mix of instructions and data access patterns in realistic scientific workloads is noticeably different from the benchmarks often used by processor manufacturers to optimize their designs.
- Simulations of the memory system revealed that the industry trend towards multicore processors must be accompanied by dramatic increases in memory capacity and bandwidth, or computers may actually get slower as more processor cores are added.
- Experiments with new applications models, such as graph-based informatics, have revealed performance limitations on conventional processors. The SST is being used to analyze new hardware techniques, such as processor-in-memory, and software models such as thread migration to address these shortcomings.

Understanding the performance limitations of future supercomputers is key to supporting a range of simulation-based science activities. The field of computer architecture is currently exploring a range of new hardware and software techniques. Powerful tools are required to understand the implications of these new designs. The Structural Simulation Toolkit provides a modular framework to explore these novel hardware and software systems.

**References**

Moore, S. (2008). Multicore Is Bad News for Supercomputers, *IEEE Spectrum*, November 2008.

Murphy, R., Rodrigues, A., Kogge, P., and Underwood, K. (2005). The Implications of Working Set Analysis on Supercomputing Memory Hierarchy Design, *International Conference on Supercomputing*, June 20-22, 2005.

Rupnow, K., Rodrigues, A., Underwood, A., and Compton, K. (2006). Scientific Applications vs. SPEC-FP: A Comparison of Program Behavior, *ICS '06*, New York, NY, pages 66-74.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

Figure 2. SST uses a modular structure to enhance flexibility.

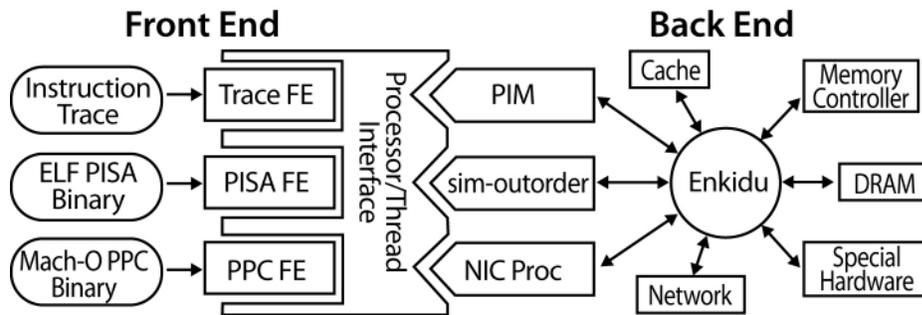
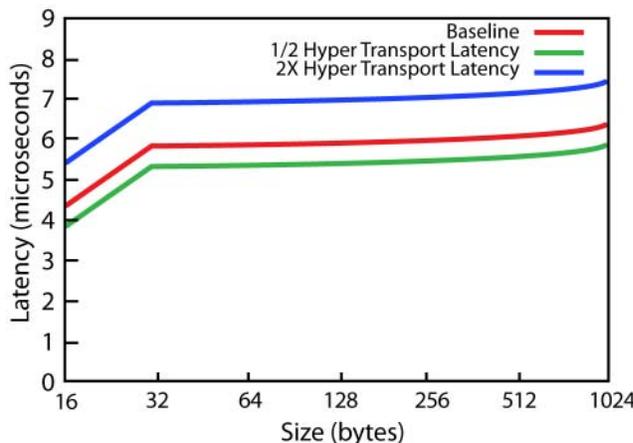


Figure 3. Simulations reveal potential enhancements to the Red Storm network.



# EXPEDITING DEVICE MODELING USING VERILOG-A COMPACT MODEL COMPILER

Contacts: Thomas Russo, Electrical and Microsystem Modeling, [tvrusso@sandia.gov](mailto:tvrusso@sandia.gov)  
Eric Keiter, Electrical and Microsystem Modeling, [erkeite@sandia.gov](mailto:erkeite@sandia.gov)

*Verilog-A is a high-level modeling language for analog circuit simulation and is well suited for device modeling. It also has become the industry standard format for public-domain models.*

The goal of this project is to produce tools for compact model development, in which a compact physics/device model can be specified in a high-level language such as Verilog-A. The Verilog-A description is translated via a so-called “model compiler” into the lower-level data structures of a high-performance simulation tool such as the Xyce parallel electrical simulator. By taking this approach, model developers can spend the bulk of development time focused on the specific physics and/or equations of the model rather than the numerical implementation details specific to the transistor-level simulator.

Modern circuit simulators, such as Xyce, require sophisticated compact models of electrical components, ranging in complexity from very simple models, such as the linear resistor, to highly sophisticated transistor models, such as the MEXTRAM. Many transistor models are industry standard models, and must be supported for a production simulator to be useful to analysts and designers. Currently, a source code implementation of a modern transistor device model such as the MEXTRAM requires at least 10K lines of C/C++ code, which needs to be customized to the specific simulator design. Once the model author has released a model, it may take several months of work to install and test it in a simulator, many more months may pass before the simulator passes through its release process and becomes available to customers.

To mitigate some of these issues, software designers at Sandia have been working on developing a Verilog-A model compiler for the Xyce circuit simulator. Verilog-A is a high-level modeling language for analog circuit simulation, and is well suited for device modeling. It also has become the industry standard format for public-domain models. A Verilog-A model description is substantially smaller than a hand-coded C++ description, in part because the Verilog-A description excludes solver data structures (which are different for every simulator), and details of the numerical implementation such as Jacobian derivatives. The project has used the open source ADMS model compiler, originally developed at Motorola, as a starting point, and has required the customization of the back end to work correctly with Xyce.

This capability will result in expedited Xyce model development, especially for non-experts. Historically, developing models in Xyce has required considerable expertise in C++ code design, OO code development, Xyce-specific data structures, and the time devoted to such issues detracted from the model itself. Typically, two general types of compact models have been incorporated into Xyce:

1. Industry standard models, like the MEXTRAM, VBIC, etc.
2. Original Sandia-specific models. Typically these models have included unique effects due to hostile and abnormal environments, including prompt ionizing radiation, total dose, and neutron damage.

For both types of models, the development costs can be considerable, and often prone to error.

While this project should ultimately result in expedited model development for Xyce, it is not without technical challenges. The project is using the Saccado library for automatic differentiation (AD) to generate the matrices of derivatives necessary to solve the circuit equations. Developing correct Jacobian derivatives is one of the most time-consuming and error-prone aspects of traditional “by hand” model implementations, so using AD has the potential to substantially reduce debug time. However, using Saccado requires that many operations be done using C++ templates on complex object types rather than simple function definitions on real variables. One ongoing challenge to the project is that in the current implementation, this templating has not been sufficient to represent the full range of Verilog-A input required. Overcoming this challenge will require careful software engineering to improve the use of Saccado by the model compiler.

In the current version of the model compiler, the Xyce source produced by the compiler has to be compiled and linked at compile time. However, models produced from Verilog-A input ideally will be loadable in Xyce without recompiling or re-linking the rest of the code. This is ultimately the goal of this work, but it will require the use of dynamic module loading and a “plug-in” API. Incorporating

such a feature in a portable manner will be challenging given the range of platforms that Xyce supports.

Device models in circuit simulators typically require numerical tricks to improve solver convergence and make them more robust. These techniques are traditionally applied directly in the device model code on a case-by-case basis, an approach that is not amenable to the automated translation required by this project. An approach is needed for assuring robust models without ad hoc convergence treatments.

### **References**

Accelera (2008). *Verilog-AMS Language Reference Manual Version 2.3*, August 2008. Napa, CA.

Kundert, K., Automatic Model Compilation: An Idea Whose Time Has Come, [www.designers-guide.org/Perspective/modcomp.pdf](http://www.designers-guide.org/Perspective/modcomp.pdf).

Lamaitre, L. McAndrew, C. C., and Hamm, S. (2002). ADMS – Automatic Device Model Synthesizer, *Proc. IEEE CICC*.

MEXTRAM home page. <http://mextram.ewi.tudelft.nl/index.html>.

***This work has been principally supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

# TITAN TOOLKIT MAKES LARGE-SCALE ANALYTICS POSSIBLE

Contacts: Brian Wylie, Data Analysis and Visualization, [bnwylie@sandia.gov](mailto:bnwylie@sandia.gov),  
 Timothy Shead, Data Analysis and Visualization, [tshead@sandia.gov](mailto:tshead@sandia.gov)

*The domain of cyber intrusion analysis is of particular importance to national security and infrastructure defense.*

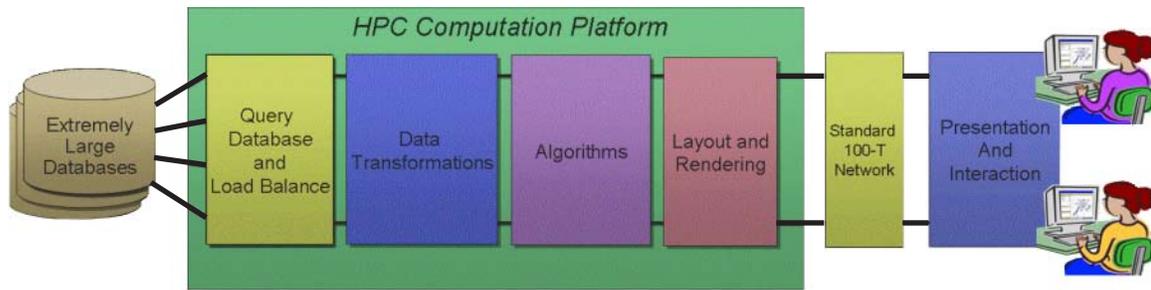


Figure 1. Titan's high-performance client/server architecture delivers scalable information analysis and advanced algorithmic capabilities to the desktop.

The open-source Titan Analytics Toolkit (Figure 1) is designed to address some of the world's largest and most challenging informatics problems. Based on the *ParaView* scalable client/server framework, the Titan Toolkit provides functionality for scalable ingestion, processing, and display of informatics data. The flexible, component-based pipeline architecture enables rapid integration and deployment of algorithms in the fields of intelligence, semantic graph, and information analysis.

The in-depth analysis of data important to national security requires a broad variety of scalable algorithms. Titan's flexible build system, use of efficient adapters, and generic programming techniques make an ideal integration platform for some of the world's most advanced algorithmic libraries:

- **MTGL** (Multithreaded Graph Library): Graph algorithms for hyper-threaded hardware.
- **Trilinos**: Scalable distributed memory linear algebra.
- **MapReduce**: Framework for distributed computing.
- **ParaText**: Scalable text ingestion and processing.
- **OVIS**: Scalable distributed statistical modeling.
- **BGL** (Boost Graph Library): Generic graph algorithms.
- **PBGL** (Parallel BGL): Distributed graph algorithms.
- **Qt**: User interface and application development.

Creating an analytics toolkit with this broad set of algorithms requires a flexible, well-designed framework that enables developers to couple new functionality in a straight forward manner. The list above continues to grow; we are currently looking at the integration of the Apache Lucene

*Mahout* machine learning libraries as part of our desire to provide a broad range of scalable algorithms to support information analysis.

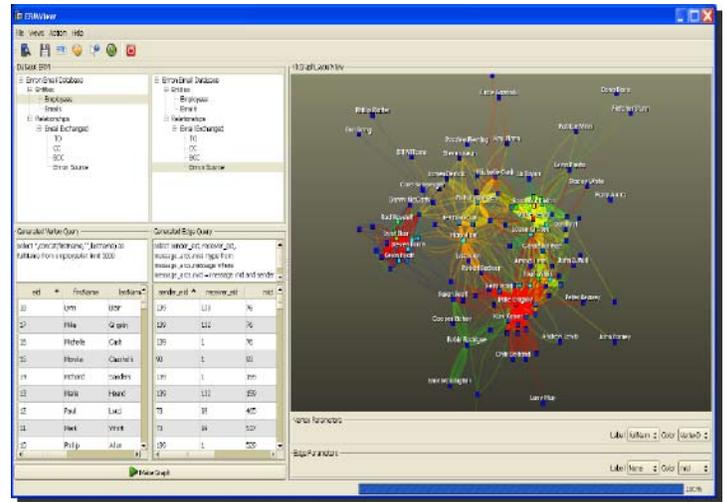


Figure 2. Titan applications can provide graph algorithms, linear algebra, and statistics to help analysts extract information from large data collections, such as this analysis of the publicly available Enron email database.

The flexibility of the Titan component architecture allows the toolkit to be applied to a wide variety of problem domains and data scales. The applications shown in Figures 2, 3, and 4 were built using Titan to address specific problem areas including biological cell development, email transaction analysis, and monitoring of network packet data. The domain of cyber intrusion analysis is of particular importance to national security and infrastructure defense.

Figure 4 demonstrates the use of graph algorithms, statistics and visualization for the analysis of internet traffic across political borders.

**References**

Cedilnik, A., Baumes, J., Ibanez, L., Megason, S., and Wylie, B. (2008). Integration of Information and Volume Visualization for Analysis of Cell Lineage and Gene Expression during Embryogenesis, *Visualization and Data Analysis 2008*, San Jose, CA.

Wylie, B., Baumes, J. (2009). A Unified Toolkit for Information and Scientific Visualization, *IS&T/SPIE Electronic Imaging, Visual Data Analytics*, San Jose, CA.

Wylie, B., Baumes, J., Shead, T. (2008). Titan Informatics Toolkit, *IEEE Visualization Tutorial*, Columbus, OH.

Wylie, B., Baumes, J., Shead, T. (2008). GSpace: A Linear Time Graph Layout, *IS&T/SPIE Electronic Imaging, Visualization and Data Analysis*, San Jose, CA.

**Titan Web Page:** [www.sandia.gov/Titan](http://www.sandia.gov/Titan)

**ParaView Web Page:** [www.sandia.gov/ParaView](http://www.sandia.gov/ParaView)

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

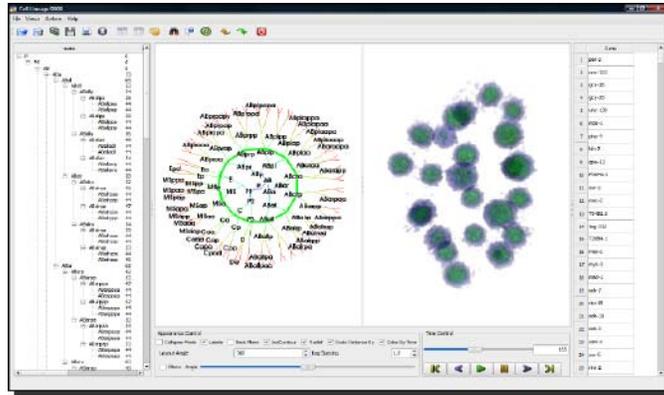


Figure 3. A Titan application used to analyze the development (cell lineage) of a *C. elegans*.

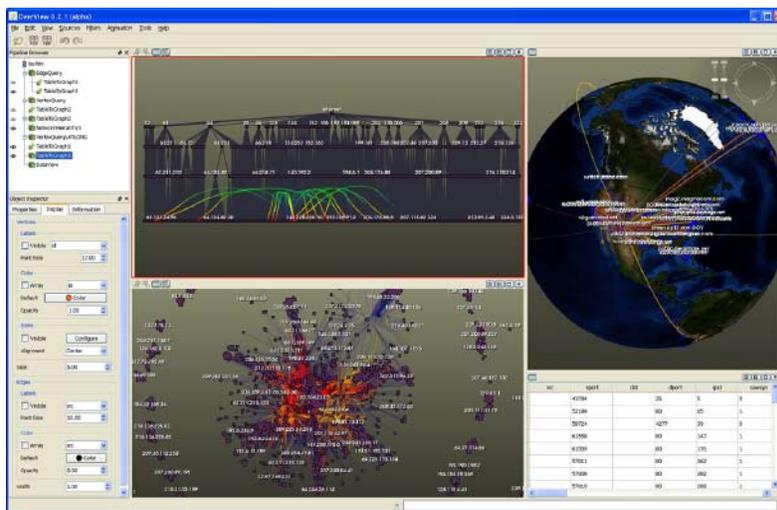


Figure 4. A Titan application for the monitoring and analysis of network packet capture data. The toolkit provides database drivers that allow analysis applications to directly interact with large, dynamic data collections. The three views above illustrate network packet transfers across internet routers and political borders.

# EXTENDING PARAVIEW FOR FRAGMENT ANALYSIS ON EXTREMELY LARGE DATA

Contact: Kenneth Moreland, Data Analysis & Visualization, [kmorel@sandia.gov](mailto:kmorel@sandia.gov)

*Recently Sandia extended ParaView to provide an innovative capability for the analysis and understanding of material fragmentation in explosive environments.*

ParaView, the open-source visualization tool and framework, tackles some of the world's largest and most challenging scientific visualization problems. As a post-processing architecture, ParaView provides advanced functionality for subsetting, drill-down, visualization, and plotting. Recently, Sandia extended ParaView to provide an innovative capability for the analysis and understanding of material fragmentation in explosive environments. This new capability operates in parallel on large distributed-memory supercomputers, enabling the processing and analysis of some of the world's largest simulation datasets.

Analyzing shock physics, which can involve high energies, high velocity materials, and dynamic phase transitions, is challenging. To better understand the physics and survey the effects, Sandia National Laboratories employs CTH, a multimaterial, large deformation, strong shock wave, solid mechanics simulation hydrocode. Although we can achieve high-fidelity simulations with CTH, we require a significant amount of post processing. Physical structures and their accompanying data must be derived from the volumetric properties computed by the simulation. To capture small fragmentation effects, simulations are run on very large scales using adaptive meshes, which further complicate the post processing. These simulations must be validated against experiments. Very little can be measured during a shock physics experiment, and most experimental data are collected in the aftermath. By using the scalable visualization tool ParaView coupled with customized feature identification, we are able to provide both the analysis and verification of these large-scale simulations.

To better analyze devices in explosive environments, Sandia is using the CTH shock physics analysis software [1]. In our experiments, a high-fidelity mesh is important. Even the simple interactions shown in Figure 1 yield thousands of fragments, many of which are less than one microgram. Such small fragments cannot be represented if the computation grid is too coarse. We achieve the necessary resolution by the adaptive mesh refinement (AMR) capabilities of CTH [2] on large-scale computers like Sandia's Red Storm, a Cray XT3 supercomputer with over 10,000 compute nodes

Making scientific queries from simulation data requires multiple post-processing tasks that have recently been incorporated into ParaView [3]. First, we need to extract the geometry of the fragments. Although it is often used to simulate fragmenting solid objects, CTH neither tracks nor derives the geometry of the solids. Instead, CTH uses an axis-aligned grid that does not conform to the geometry. Within each cell of the grid, CTH computes volume fractions, the portion of the cell filled by each material. Fragments must be identified by isolating connected cells containing a material, and then statistics such as mass and volume can be derived from the shape. The second post-processing task is to validate the simulation. The results of a simulation must match the results of an experiment with an equivalent initial condition. Without this verification step, it is impossible to know if inferences from the simulation are correct. The third post-processing task is to extract statistical information to provide better understanding of the consequences in a real-world environment.

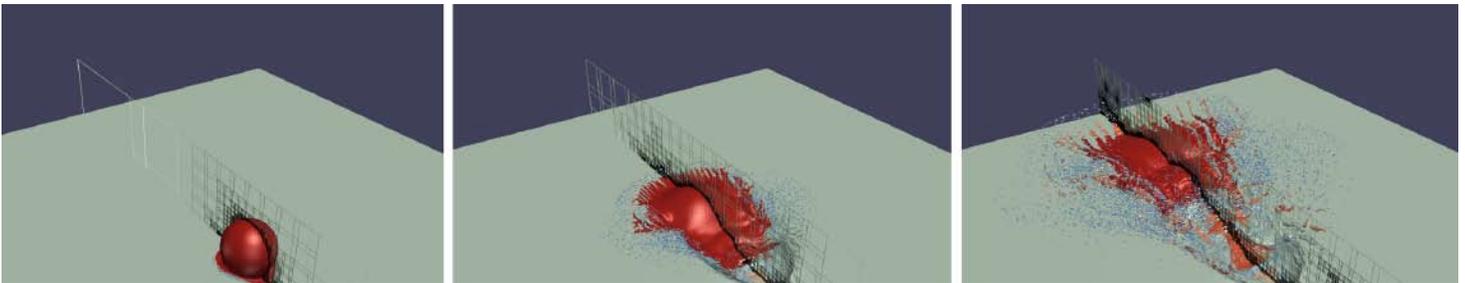


Figure 1. Simulation of a high-velocity metal ball striking a metal brick.

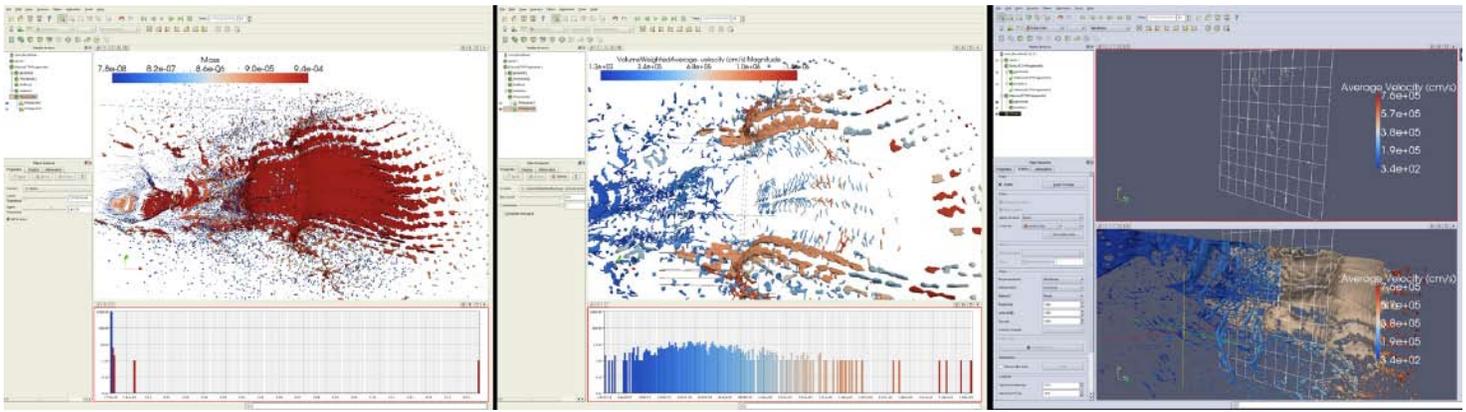


Figure 2. Fragment identification enables further analysis including parameter characterization (left), property thresholding (middle), and spatial intersection queries (right).

ParaView's highly scalable visualization and post-processing framework makes it possible to process large-scale data such as that generated by simulations run on Red Storm. By leveraging ParaView's modular design we were able to implement efficient parallel fragmentation analysis algorithms with contributions from Sandia, Kitware, Inc., and the U.S. Army Research Laboratory (ARL). With these additions to ParaView, we can now derive fragment shapes from volume fractions and extract associated physical properties such as mass, volume, temperature, pressure, and velocity. The extracted properties provide handles that allow us to select, drill-down, and analyze the data, as demonstrated in Figure 2. Additionally, these data provide physical measurements that are comparable to experimental results — critical for the verification of the simulation's physics and the analysis of virtual experiments.

### References

1. Hertel, E. S., Bell, R. L., Elrick, M. G., Farnsworth, A. V., Kerley, G. I., Mcglaun, J. M., et al. (1993). CTH: A Software Family for Multi-Dimensional Shock Physics Analysis, *Proceedings of the 19th International Symposium on Shock Waves*, pp. 377-382.

2. Schmitt, R. G., Crawford, D. A., Bell, R. L., and Hertel, E. S. (2002). Adaptive Mesh Refinement and Multi-Phase Flow in the CTH, *Workshop on Numerical Methods for Multi-Material Fluid Flows*, Paris, France.

3. Karelitz, D. B., Ice, L., Wilke, J., Attaway, S. W., and Moreland, K. D. (2008). *Post-Processing V&V Level II ASC Milestone (2843) Results*, Sandia National Laboratories, Albuquerque, NM.

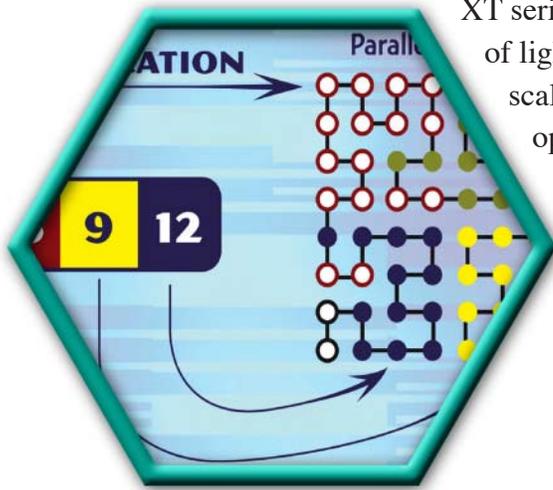
***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

*End of Section II*



# ARCHITECTURES & SYSTEMS

A key to Sandia's success in high performance computing has been a multidisciplinary approach that brings together computational scientists, mathematicians and computer scientists to tackle nationally important problems. Computer science efforts have included architectures and system software R&D. Sandia and Cray co-developed the Red Storm system that forms the basis for one of the most successful supercomputers ever — the Cray



XT series. In system software, Sandia pioneered the development of light-weight operating systems that are critical to achieving scalability on large systems. Sandia helped develop the operating systems for the Intel Paragon, ASCI Red (the world's first teraFLOPS computer), and Cray's Red Storm. Now, new efforts range from petascale enablement to cycle-accurate system simulations to resilience.

# THE NEW MOORE'S LAW AND SANDIA'S ROLE IN THE FUTURE OF COMPUTING

Contacts: Erik P. DeBenedictis, Scalable Computer Architecture, [epdeben@sandia.gov](mailto:epdeben@sandia.gov)  
 Kwok Kee Ma, Advanced Microelectronics and Radiation Effects, [makk@sandia.gov](mailto:makk@sandia.gov)  
 Tony Lentine, Applied Photonic MicroSystems, [allenti@sandia.gov](mailto:allenti@sandia.gov)  
 Paul Dodd, Radiation Physics, Technology, & Assurance, [pedodd@sandia.gov](mailto:pedodd@sandia.gov)

*...the physics of computation is an unexpected new force demanding a rewrite of old code... This project has identified the source of this change in physics and is predicting Sandia's future role systematically rather than on an ad hoc basis.*

This project developed a refinement to Moore's Law that better projects how complementary metal-oxide semiconductor technology (CMOS) and other enabling technologies will evolve over time as well as their likely impact on Sandia's creation and use of computers for national security purposes. In 1965, Gordon Moore predicted that integrated circuit density, speed, and power efficiency would scale exponentially for a decade [1], but the prediction was actually valid for four decades. The "original Moore's Law" curve in Figure 1 became a defining force for the computer industry in this period, leading to petaFLOPS supercomputers and expectations of exaFLOPS supercomputers. From everyday experience, we know that microprocessor clock rate leveled off around 2005 and that computer power consumption has become a profound concern. Only density continues to rise significantly in "new Moore's Law" scaling, also illustrated in Figure 1.

The shift in Moore's Law is having a profound implication on supercomputing. The scaling predicted by the original Moore's Law enabled supercomputers with rapidly growing performance (gigaFLOPS, teraFLOPS, petaFLOPS, mostly through increased single-node performance. Node count grew slowly enough that parallel software and algorithms could be ported from one generation of supercomputer to the next. In the new Moore's Law, rising device count must be accompanied by an almost equal rise in parallelism (because of near-constant clock rate) and a decreasing duty cycle per device (because of flat lining power efficiency). Changes in Moore's Law for the enabling technology are thus creating a strong force towards supercomputers more like "memories that compute" than the "FLOPS engines" that have dominated supercomputing in past years. The rise of Informatics at Sandia is consistent with this trend (also see p. III-16, "*Enabling Large-Scale Informatics Using Novel Architectures: Pythia and X-Caliber*").

Industry responds in the commercial marketplace through increasing specialization of architectures, a lesson that needs to translate to supercomputers. It has been

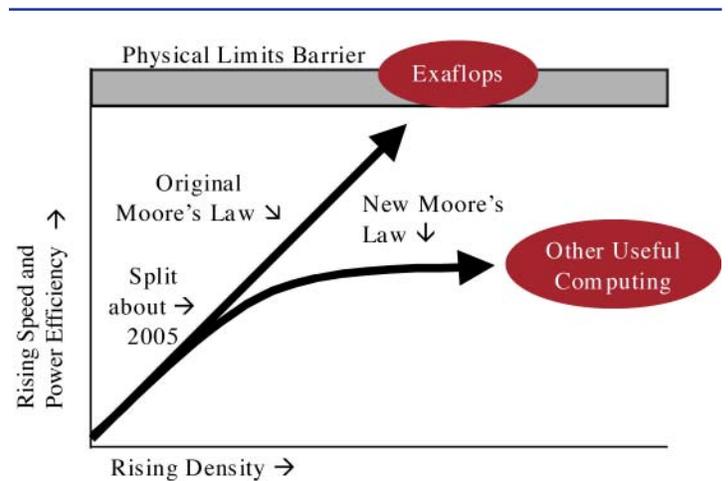
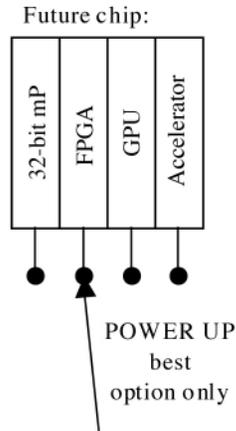


Figure 1. The original Moore's Law was headed straight towards exaFLOPS, but the physics of the underlying technology are shifting in other directions.

known for decades that specialized accelerators are faster and more power efficient than microprocessors. Industry is adopting this paradigm through computers containing Graphics Processing Units (GPUs) and with future plans for acceleration of other functions, such as encryption. The general architecture shown in Figure 2 dedicates rising device count to additional specialized architectures, "turning on" only the units that need to be active at any one time to prevent overheating. The architecture of Figure 2 is present in some top supercomputers [2] in an *ad hoc* way, but we argue that this must become a future trend.

Algorithms and complexity theory can have a role. Computer scientists have been very successful in finding algorithms that minimize the number of arithmetic operations (ops/FLOPS) needed to solve a particular type of problem, such as matrix multiplication. This was very relevant in the past when logic gates were precious, but is less relevant today now that power consumption is more important. However, some of the principles of complexity theory can be redirected towards power consumption, such as algorithms that minimize memory bandwidth while

**Figure 2.** Instead of a universal microprocessor that performs all tasks in software, industry is headed to a multi-architecture for speed and power savings.



multiplying matrices [3]. Accommodating to the new Moore’s Law by altering the nature of algorithms will make progressively more sense as technology advances.

Computer science will have an important impact on semiconductors, including at Sandia/DOE facilities such as MESA (Microsystems Engineering Sciences Applications). In the recent past, the semiconductor industry sought a drop-in replacement for CMOS to extend the life of the original Moore’s Law. This quest is losing momentum now that it is understood that the resulting computer systems would just overheat faster than today’s computers. In its place, there is device physics R&D in sublithographic nanotechnology aimed at arrays of components that compute or store data in different ways. The recent publicity surrounding a new class of nanodevices called memristors is an example [4]. The future vision is for computer chips that are augmented by new layers of nanodevices that perform data storage and highly power efficient but specialized logic functions. Figure 3 illustrates some ideas where new nanodevices form

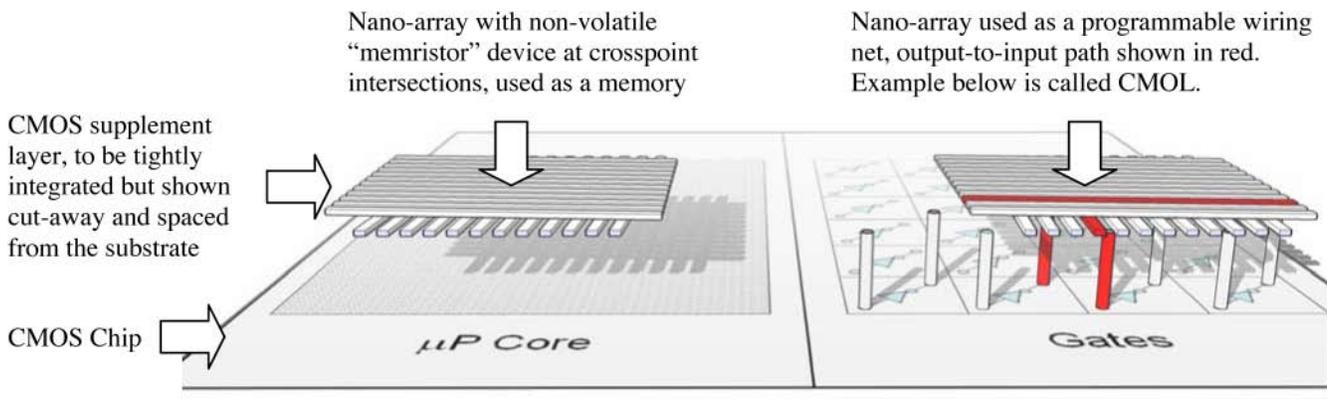
a programmable interconnect in a future computing device, displacing the need for a microprocessor and software.

Computing used to be limited only by the imaginations of the computer architect and programmer, yet the physics of computation is an unexpected new force demanding a rewrite of old code to accommodate multi-core architectures, power-efficient algorithms, and new nanotechnology. This project has identified the source of this change in physics and is predicting Sandia’s future role systematically rather than on an *ad hoc* basis.

**References**

1. Moore, E. G. (1965). Cramming More Components Onto Integrated Circuits, *Electronics*, Vol. **38**, pp. 114-117.
2. Barker, K. J., Davis, K., Hoisie, A. et al. (2008). Entering the PetaFLOP Era: the Architecture and Performance of Roadrunner, *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, Austin, TX: IEEE Press, pp.1-11.
3. Singh, M. and Prasanna, V. K. (2003). Algorithmic Techniques for Memory Energy Reduction, *Experimental and Efficient Algorithms: Second International Workshop Proceedings*, WEA 2003, Ascona, Switzerland, May 26-28, 2003. p. 622.
4. Sniderland, G. S. and Williams, R. S. (2007). Nano/CMOS Architectures Using a Field-Programmable Nanowire Interconnect, *NANOTECHNOLOGY*, Vol. **18**, p. 35204.

*This work has been supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program and Work for Others.*



**Figure 3.** The future of chip architecture is very much up in the air, but the diagram illustrates the richness of the design space. A future chip may comprise both conventional and new architectures (also see Figure 2) as well as CMOS supplemented with new types of nanodevices (such as memristors). Designs in this direction may reestablish the historical scaling of Moore’s Law, but it is quite likely that the nature of programming and algorithms will have to change.

# COMPARISON REVEALS RELATIVE STRENGTHS OF MULTITHREADED AND DISTRIBUTED MEMORY ARCHITECTURES

Contacts: Karen Devine, Scalable Algorithms, [kddevin@sandia.gov](mailto:kddevin@sandia.gov)  
Jonathan Berry, Scalable Algorithms, [jberry@sandia.gov](mailto:jberry@sandia.gov)  
Steve Plimpton, Scalable Algorithms, [sjplimp@sandia.gov](mailto:sjplimp@sandia.gov)

*Our results show both that distributed memory architectures are viable, even on power-law data, and that massively multithreaded architectures can be competitive doing floating-point computations.*

This project makes the first apples-to-apples comparisons of traditional distributed memory and massively multithreaded architectures using realistic input data. The algorithmic kernel of our comparisons is Google's PageRank method on synthetic datasets with power-law vertex degree distributions.

With their faster clock speed and good floating-point support, distributed-memory supercomputers typically dominate the slower multithreaded machines in scientific computations on physics simulation data. On the other hand, with hardware latency tolerance for random memory references, multithreaded machines have dominated in unstructured graph search experiments [1].

PageRank represents an interesting compromise. The algorithm reduces to matrix-vector multiplication with floating point arithmetic, yet the data we choose are representative of informatics problems that inspired the development of multithreaded architectures. Our results show both that distributed memory architectures are viable, even on power-law data, and that massively multithreaded architectures can be competitive doing floating-point computations.

PageRank computes the importance of a web page [2]. Page  $s$  is important if other important pages point to it. The share of importance that  $s$  receives from page  $t$  is inversely proportional to the number of pages that  $t$  links to. Thus, PageRank models the web as a directed graph with one vertex per web page and one edge for each link from one page to another. Edges from a given page are weighted by the probability of following a link on that page. Thus, page importance is propagated through the graph edges.

This graph model is used in our massively multithreaded implementation of PageRank in Sandia's MultiThreaded Graph Library (MTGL) [1, 6]. Rank propagation is accomplished through adjacency list traversal. We used a compressed sparse row data structure, but the same code would run on other graph representations.

Our distributed-memory implementation of PageRank represents the graph as a sparse matrix  $A$  with nonzero matrix entries  $A_{st}$  equal to the weight of the edge from vertex  $s$  to vertex  $t$ . The PageRank algorithm is simply a power-method iteration in which the dominating computation is matrix-vector multiplication [3].

Our experimental data are R-MAT graphs [4], which are recursively generated graphs with power-law degree distributions. They are often used to represent Web and social networks. A small set of parameters affects the resulting distribution of vertex degrees. We used two different R-MAT parameter sets: one "nice," resulting in a maximum vertex degree of roughly 1000, and the other "nasty," yielding an inverse power law degree distribution with a maximum degree of roughly 200,000. The latter distribution is representative of social networks, for example. Each dataset consists of  $2^{25}$  vertices and roughly a quarter of a billion directed edges. For both implementations, randomly permuting the vertex numbering is crucial in order to achieve good parallel speedup.

We ran our experiments on a 128-processor Cray-XMT, as well as on Sandia's Red Storm (Cray-XT3) distributed-memory supercomputer and on Sandia's Thunderbird cluster. We present the time for one PageRank iteration for the "nice" and "nasty" data sets in Figures 1 and 2, respectively. For the nice data set, performance is comparable between architectures, with each achieving nearly perfect strong scaling, as well as similar raw times. Randomization allows the distributed memory runs on "nasty" data to scale almost perfectly out to at least 1024 processors.

In 2004, Yoo et al. [5] performed an experiment on the IBM BlueGene/Light, a distributed-memory machine and the fastest supercomputer in the world at that time. They implemented " $s$ - $t$  connectivity," a simple graph search algorithm and ran it on "Erdos-Renyi" random graphs, which are, in effect, a "nicer" version of our "nice" data. They were able to find a path between two nodes in such a graph with

roughly 20 billion edges in about one second using 32,000 processors. Our experiments with the Cray-XMT indicate that this computation could be done with roughly 10 processors of an XMT. Figure 3 provides some evidence of this claim: as the number of edges increases to roughly 2 billion, ten-processor XMT runs approach the performance obtained by Yoo et al. on the BlueGene platform.

We are in the process of extending our “PageRank Derby” to include graph algorithms, such as *s-t* connectivity. We will apply both distributed memory and massive multithreading in order to discover the strengths and limitations of each platform on such problems.

**References**

1. Berry, J. W., Hendrickson, B., Kahan, S., et al. (2007) Software and Algorithms for Graph Queries on Multithreaded Architectures, *Proceedings. 21st International Parallel and Distributed Processing Symposium*, March 2007.
2. Page, L., Brin, S., Motwani, R., et al. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*, Technical report, Stanford Digital Library Technologies Project.

3. Langville, A. N. and Meyer, C. D. (2005). A Survey of Eigenvector Methods for Web Information Retrieval, *SIAM Review*, Vol. 47, No.1, pp. 135-161.
4. Chakrabarti, D., Zhan, Y., Faloutsos, C. (2004). R-mat: A Recursive Model for Graph Mining. *Fourth SIAM International Conference on Data Mining*.
5. Yoo, E. Chow, K., Henderson, B., et al. (2005). A Scalable Distributed Parallel Breadth-First Search Algorithm on BlueGene/L, *Supercomputing 2005*.
6. Berry, J. W., MTGL Website: <https://software.sandia.gov/trac/mtgl/>. 2008.

*This work has been supported by Sandia’s Defense Systems & Assessments Strategic Management Unit.*

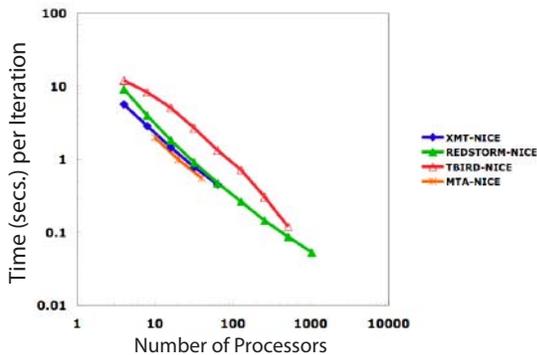


Figure 1. PageRank Derby scaling results for the “Nice” data set.

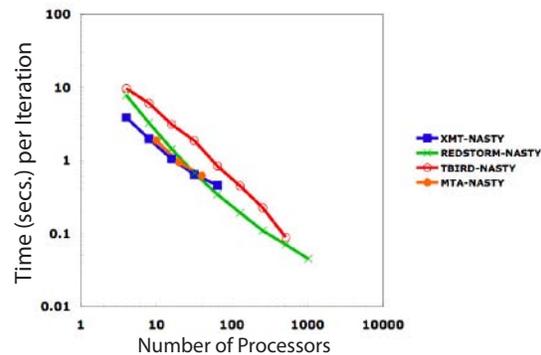


Figure 2. PageRank Derby scaling results for the “Nasty” data set.

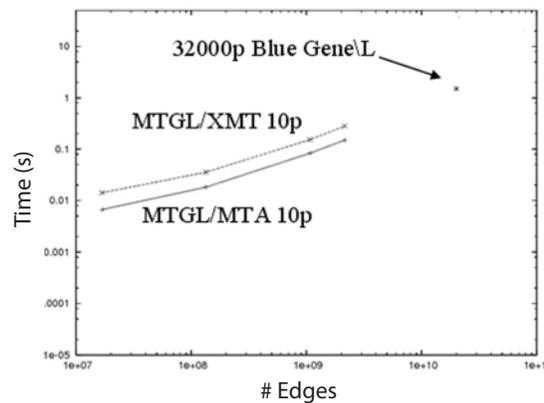


Figure 3. Performance comparison of graph search algorithms on IBM Blue Gene/L and Cray XMT/MTA using the Multi-Threaded Graph Library (MTGL).

# ENHANCING NEXT GENERATION INTERCONNECTION NETWORKS FOR HIGH-PERFORMANCE COMPUTING

Contacts: K. Scott Hemmert, Scalable Computer Architecture, [kshemme@sandia.gov](mailto:kshemme@sandia.gov)  
 Brian W. Barrett, Scalable Computer Architecture, [bwbarre@sandia.gov](mailto:bwbarre@sandia.gov)  
 Arun F. Rodrigues, Scalable Computer Architecture, [afrodri@sandia.gov](mailto:afrodri@sandia.gov)

*...it is critical that each new generation of machine continues to improve message passage performance in order to enable high levels of scalability. Our research addresses how this can be done in an environment where realized MPI performance is increasingly difficult to achieve.*

Message passing mechanisms (via MPI) are critical components for the success of high performance computing. Recent research at Sandia, namely, the development of a new queue processor to manage message lists, has potential to provide much higher message passing rates and much lower latency. This research is now impacting multiple vendors' product developments and may result in vastly improved supercomputer performance.

Over the past decade, billions of dollars have been invested in MPI codes for DOE mission critical application areas. As such, it is critical that each new generation of machine continues to improve message passage performance in order to enable high levels of scalability. Our research addresses how this can be done in an environment where realized MPI performance is increasingly difficult to achieve.

When measuring MPI performance, three primary components are considered: bandwidth, latency, and message throughput. Of these, the mechanisms for improving bandwidth are probably the best understood (though actually providing this increase can be quite expensive). On the other hand, usable reductions in latency and increases in message rate are becoming progressively more difficult in each generation. Unfortunately, the well-known LogP model for interconnect performance suggests that overhead (a significant component of latency) and gap (a corollary to message throughput) are at least as important as bandwidth for application performance.

These issues are further complicated by the way in which latency and message throughput are measured. MPI latency is typically measured using a simple "ping-pong" test, which usually results in an ideal, rarely realized measurement. The MPI ping-pong test results in significantly lower MPI processing requirements than typical MPI because of the idealized nature of the benchmark.

Another important network characteristic is message throughput. Message throughput is often characterized by "streaming bandwidth" benchmarks and is the primary factor in how quickly bandwidth curves ramp to their maximum

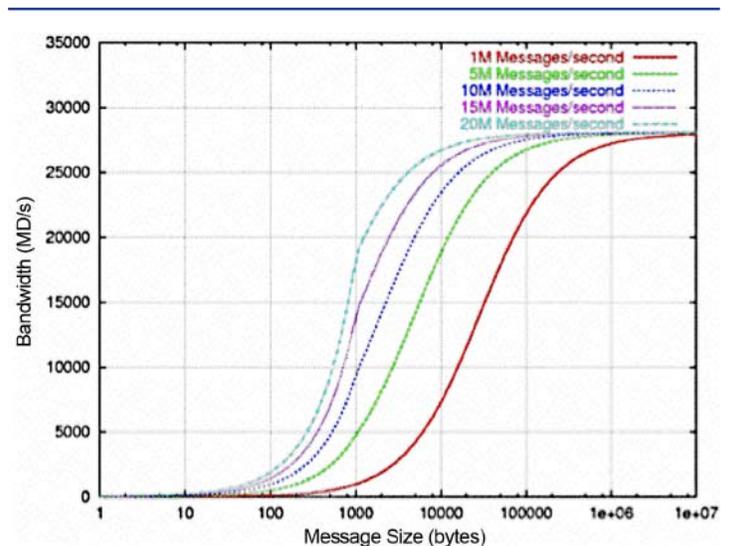


Figure 1. Streaming bandwidth results for a theoretical network with a peak bandwidth of 30 gigabytes.

value. As an example, Figure 1 shows results for streaming bandwidth for different message throughputs for a theoretical network with 30 GB/s of peak bandwidth. Comparing the effective bandwidth for the two extremes in the figure shows that a network capable of processing 20 million messages per second reaches a half bandwidth at a 1 kilobyte message size and is within 10% of peak with 10 kilobyte messages. At 1 million messages per second, these performance levels are achieved at 10 kilobytes and 1 megabyte, respectively. This vast discrepancy necessitates that interconnection networks achieve high MPI message throughput.

Our research investigates techniques for improving both latency and message throughput through the use of novel techniques to offload MPI processing to the network interface controller (NIC). This work primarily focuses on MPI matching. The two-sided nature of communication in MPI leads to two primary communication lists: the posted receive list and the unexpected message list. The posted receive list is a list of messages that the node expects to receive from other nodes. The unexpected list is a list of messages that

were received, but no matching item in the posted receive list was present.

The two MPI lists are tightly coupled, and the processing of these lists is one of the key performance bottlenecks for both latency and message throughput. The result of this research was the development of the queue processor, shown in Figure 2. The queue processor offloads the processing of both the posted receive and the unexpected message lists. MPI offload not only enables lower latency and higher message rate, but also provides mechanisms for allowing communications to proceed even when the host central processing unit (CPU) is otherwise occupied (independent progress). Moving this processing to the NIC also results in the MPI library interfering less with the caching properties of the application.

The queue processor is made up of three primary units: the associative list processing unit (ALPU), the list manager, and the match unit. The ALPU is a dedicated hardware unit that performs parallel searches. It allows for a small number (32 to 64 per process) of items to be searched in constant time,

providing improved performance for real-world operating scenarios. If the lists become too long to fit in the ALPU, the list manager, coupled with the match unit, processes the extra items at a constant cost per item. This differs from the performance characteristics of a traditional CPU, which are driven largely by hits/misses of caches and are not predictable.

Largely because of the success of this research, we now have collaborations with multiple vendors to transfer Sandia/DOE developed IP and concepts into vendor roadmaps. The incorporation of these technologies into large-scale machines will contribute to platforms with high levels of scalability to many DOE mission critical applications.

### References

Culler, D., Karp, R., Patterson, D., Sahay, A., Schauser, K. E., Santos, E. Subramonian, R., and von Eicken, T. (1993). LogP: Towards a Realistic Model of Parallel Computation, *Proceedings of the Fourth ACM SIGPLAN Symp., Principles and Practices of Parallel Programming*, pp. 1-12.

Hemmert, K. S., Underwood, K. D., and Rodrigues, A. (2007). An Architecture to Perform NIC Based MPI Matching, *Proceedings of the 2007 IEEE International Conference on Cluster Computing*, pp. 211-221.

Underwood, K. D., Hemmert, K. S., Rodrigues, A., Murphy, R., & Brightwell, R.A. (2005). Hardware Acceleration Unit for MPI Queue Processing, *Proceedings of the 19th International Parallel and Distributed Processing Symposium*, pp. 96b-96b.

Underwood, K. D., Rodrigues, A., Hemmert, K. S. (2005). Accelerating List Management for MPI, *Proceedings of the 2005 IEEE International Conference on Cluster Computing*, pp. 1-10.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

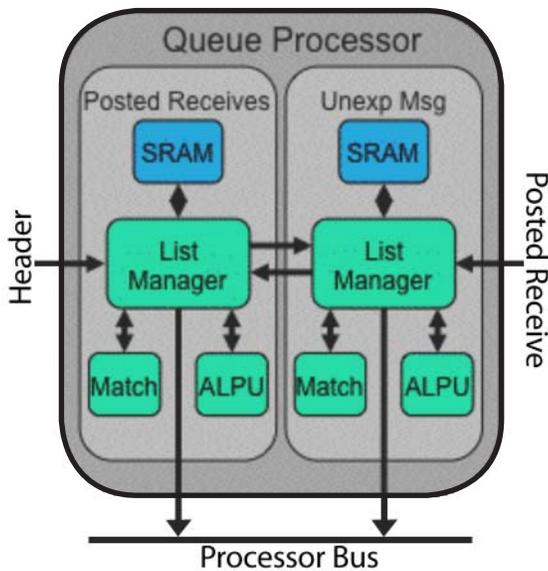


Figure 2. Block diagram of the Queue Processor.

# OPERATING SYSTEM JITTER CAN MAKE APPLICATIONS NERVOUS

Contacts: Ron Brightwell, Scalable System Software, [rbbrigh@sandia.gov](mailto:rbbrigh@sandia.gov)  
Kurt Ferreira, Scalable System Software, [kbferre@sandia.gov](mailto:kbferre@sandia.gov)

*Our results thus far demonstrate the impact of OS noise, the importance of how noise is generated (in terms of frequency and duration), and how this impact changes with scale.*

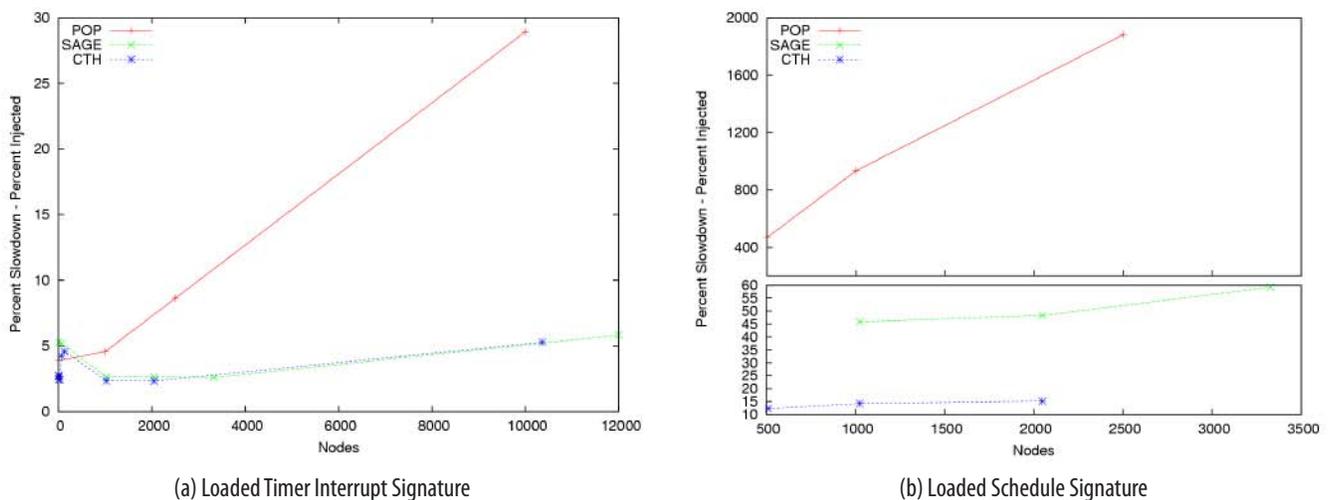
Not only is hardware important to the success of large-scale parallel computing, but the operating system (OS) running on each node is also critical in determining overall application performance. Our research shows that slight interference from the operating system—also referred to as or OS jitter or OS noise—can make some applications run nearly twenty times slower (Figure 1). This study reinforces the need for nonintrusive operating systems, like the lightweight kernels designed and developed at Sandia, which focus on meeting the demands of scalable, high-performance applications.

Lightweight kernels (LWK) have been critical to enabling performance on capability supercomputers for nearly twenty years [1], with Sandia playing a leadership role throughout their history. One feature that distinguishes LWKs from commodity operating systems, like Linux, is that LWKs avoid interfering with progress of the application by running only when requested. Current research has shown that OS interference is an important limiter of application performance in large-scale systems. This impact is attributed to interference with synchronization between application instances on distributed hosts by OS services and associated daemons. There are few published studies on the effect of different kinds of kernel-generated noise on application performance at scale. In addition, the measured impact of

noise has varied widely between systems, with studies on some systems showing relatively little performance impact from noise, and others showing substantial performance impacts.

In this research, we quantify the application performance costs of a variety of local OS interference patterns on a range of real-world, large-scale applications at scales exceeding over ten thousand nodes. Our results thus far demonstrate the impact of OS noise, the importance of *how* noise is generated (in terms of frequency and duration), and how this impact changes with scale. Also, we have been able to determine how the characteristics affect the applications studied; for example, computation/communication ratios, collective communication sizes, and other characteristics relate to the tendency to amplify or absorb noise. In addition to application characteristics affecting noise, we are studying the influence certain architectural parameters of a computer system have on an applications impact to noise. These parameters include such things as the balance ratios of a node in the system (e.g., relative processing and network performance) to hardware-based communication mechanisms.

To develop a better understanding of the influence that OS noise has on large-scale parallel applications, we built a



**Figure 1.** Performance slowdown for two common commodity noise signatures. Each signature corresponds to 2.5% net processor noise. “Loaded Timer Interrupt” (left) is a high-frequency, low duration signature while “Loaded Schedule” (right) is a low-frequency, high-duration signature.

kernel-level noise injection framework into the Catamount lightweight kernel that runs on the Cray XT Red Storm machine at Sandia [2]. Catamount is an ideal choice for this research because of its extremely low native noise signature and demonstrated record of scalability. In addition, integration within Catamount for Red Storm allows for testing noise effects on a well-balanced machine (in terms of relative compute to communication performance) at scales of over ten thousand nodes for real applications important to Sandia and DOE.

Our noise injection framework provides the ability to specify a per-job noise pattern to be generated by the operating system during application execution. Parameters for noise generation pattern include the frequency of the noise, duration of each individual noise event, the set of participating nodes, and a randomization method for noise patterns across nodes.

Using this noise injection framework, we are able to examine how various OS interference patterns impact the performance of large-scale applications. For testing, we have identified a number of simulation workloads that represent a range of different computational techniques critical to Sandia's software mission of ensuring our nation's competitiveness and security. Highlights of this project's achievements are outlined below.

Figure 1 illustrates the performance slowdown of three representative large-scale applications with interference patterns similar to those measured on a modern Linux cluster. From the figure, we see that each application reacts differently to noise, and we have been able to link that reaction to a number of communication and computational properties of the code.

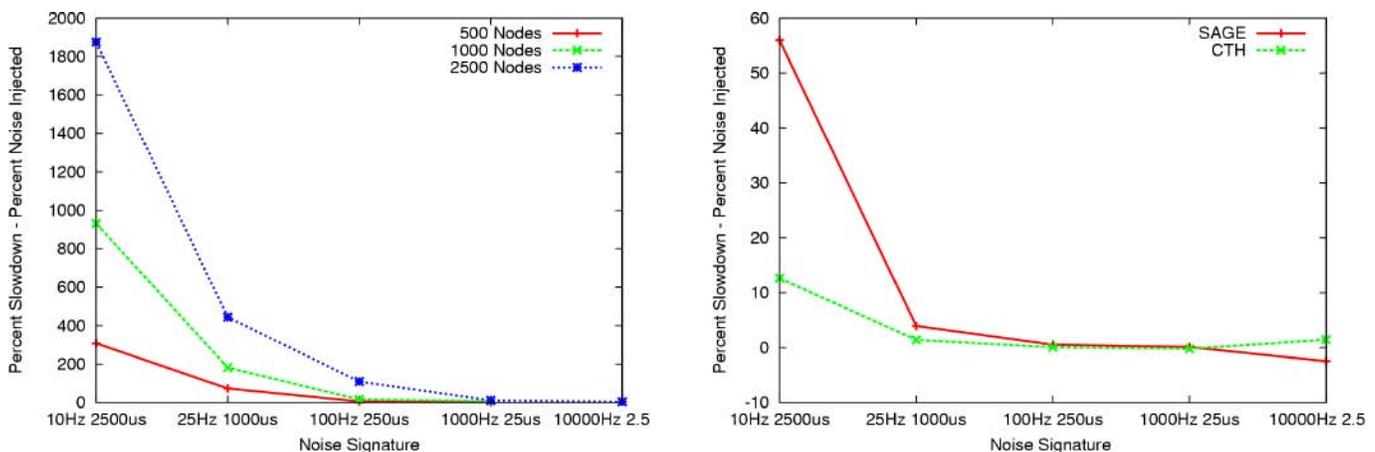
In Figure 2, we show that all OS noise is not equal for our large-scale applications. While each point on the X-axis represents the same aggregate noise, the frequency and duration of the noise differ for each point. We see that each of our applications is more sensitive to the lower frequency, higher duration noise patterns, similar to an intermittent kernel daemon on a commodity operating system.

We are now investigating the impact of various hardware features on noise. Preliminary work has shown that computing platforms, where the balance ratio is shifted towards excess computation, are more tolerant of operating system noise than those with excess network capacity. In addition to these hardware parameters, we are looking into the design of alternative application communication libraries that show less sensitivity to OS interference.

### References

1. Riesen, R., Brightwell, R., Bridges, P. G., et al. (2008). Designing and Implementing Lightweight Kernels for Capability Computing, *Concurrency and Computation: Practice and Experience*, accepted for publication.
2. Ferreira, K. B., Brightwell, R., Bridges, P. G., et al. Characterizing Application Sensitivity to OS Interference Using Kernel-Level Noise Injection, *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing (Supercomputing'08)*, November 2008, accepted for publication.

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***



**Figure 2.** Sensitivity of POP, SAGE, and CTH to noise frequency and duration parameters for a fixed 2.5% net processor noise signature. For each of the three applications, the lower frequency, higher duration OS noise signature shows the greatest performance slowdown.

# PORTALS NETWORKING SOFTWARE POWERS PETAFLUPS PERFORMANCE

Contact: Ron Brightwell, Scalable System Software, [rbbrigh@sandia.gov](mailto:rbbrigh@sandia.gov)

*Portals is the only DOE-developed native network programming interface to be deployed on a highly popular and commercially successful system.*

The Portals networking software library developed at Sandia has enabled the Cray XT series of machines to achieve record-breaking performance, including breaking the petaFLOPS barrier at Oak Ridge National Laboratory. Portals is the only DOE-developed native network programming interface to be deployed on a highly popular and commercially successful system. The Portals programming interface enables applications to communicate using high-speed networking hardware, similar to how UNIX sockets allow internet applications to communicate using commodity Ethernet networks. Unlike sockets, Portals is specifically designed to meet the scalability and performance requirements of machines with hundreds of thousands of processors.

The dominant execution model for parallel programming is communicating sequential processes, where independent tasks perform local computation and periodically exchange data. The ability to efficiently exchange data is critical to the performance and scalability of parallel computing applications. Our research in scalable, high-performance data movement of ultra-scale systems focuses on the application programming interfaces that provide these interprocess communication facilities. In particular, our recent work has been in two areas: (1) developing a next-generation, low-level network programming interface (Portals) for ultra-scale capability computing platforms and (2) enhancing support for the Message Passing Interface (MPI) library within the Sandia lightweight kernel environment. Our work on Portals and MPI are interrelated, as MPI is the most important higher-level target for Portals.

The Portals network application programming interface (API) was initially developed jointly by Sandia and the University of New Mexico as an integral component of early lightweight kernel operating systems for massively parallel processing machines. The Portals API has evolved as an independent component that provides scalable, high-performance data movement functionality for system-area networks. Rather than just encapsulate functional semantics, the Portals API exposes elementary building blocks that can be combined to support a wide variety of higher-level network transport protocols and operations. Cray, Inc., chose Portals as the lowest-level network programming interface for the SeaStar network [2] on its XT series of machines (Figure 1).

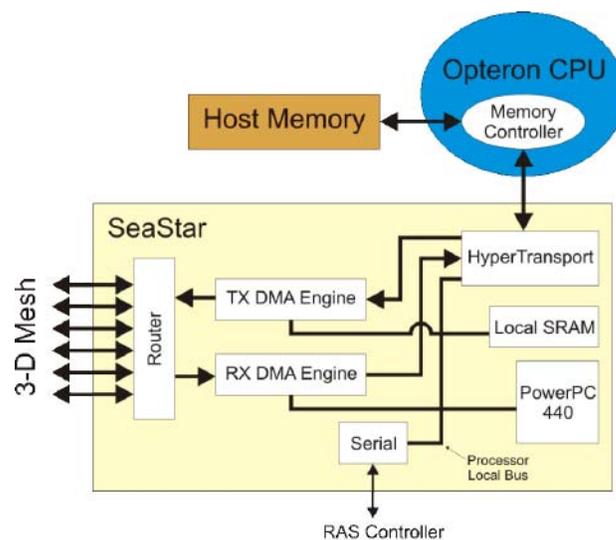


Figure 1. Sea Star network.

Sandia is a member of the Open MPI project, the purpose of which is to develop and support a scalable, high-performance, open-source implementation of the MPI Standard. MPI is the application-level communication library upon which nearly all scientific parallel computing applications are based. It is a fundamental software component that applications use for data movement operations. Sandia's work in this area involves developing and supporting Open MPI for the Cray XT platform using Portals and the Catamount lightweight kernel operating system.

There are many significant challenges associated with designing and developing a low-level network programming interface like Portals. The network programming API is where the capabilities of the underlying network hardware combine with the resource management policies and mechanisms of the operating system to fulfill the functional requirements and performance demands of parallel computing applications. In designing a network API for massively parallel computing, it is important to understand the requirements of higher-level communication interfaces like MPI. It is also important to provide mechanisms that fully exploit the capabilities of the underlying network hardware. A low-level network API not only needs to deliver low latency, high bandwidth,

and high small-message rate, but it also needs to provide scalable network resource management methods to support communication among tens of thousands, to a few hundreds of thousands of endpoints. Traditional communication protocols, such as those used for internet applications, are not suitable for a variety of reasons. Most notably, they provide redundant reliability features that degrade performance and consume too many resources on each endpoint.

Our work on high-performance network programming interfaces improves science-based engineering design capabilities by increasing the scalability, performance, and effectiveness of parallel computing platforms used for delivering advanced modeling and simulation solutions.

The development of the Portals API has had a significant impact on the parallel computing community through its implementation and deployment on the Cray XT series of machines. These machines are now used throughout the DOE and the DoD community. Portals is the only DOE-developed native network programming interface to be deployed on a highly popular and commercially successful system. The role of Portals in enabling the Cray XT system at Oak Ridge National Laboratory to break the petaflops barrier cannot be overstated.

While highly successful, the implementation of Portals on the Cray XT exposed some opportunities for enhancement on future-generation networks. We continue to work on evolving Portals in close collaboration with Intel as part of its efforts to develop a next-generation interconnect fabric based on network interface hardware designs done at Sandia. As part of this collaboration, Sandia and Intel completed work on a Cooperative Research and Development Agreement (CRADA) to implement and evaluate the next-generation Portals API on SeaStar. This work was important for validating the new semantics and capabilities added to Portals.

In addition to Portals, we made several important contributions to the Open MPI project for the Cray XT platform. We initially provided support for using Accelerated Portals, which is an optimized implementation of Portals that offloads all network processing to the SeaStar network interface. Next we extended the support for using shared memory for intra-node message passing in Open MPI for the Catamount lightweight kernel, leveraging a new shared memory capability called SMARTMAP. This work is important because it can potentially reduce the load on the network interface, especially for nodes containing large multi-core processors. Finally, our ability to exploit this capability for MPI has laid the foundation for other important work on proposed multi-core extensions for MPI being explored under the auspices of the DOE Institute for Advanced Architectures and Algorithms.

### **References**

1. Brightwell, R., Hudson, T., Pedretti, K., et al. (2005). Implementation and Performance of Portals 3.3 on the Cray XT3, *Proceedings of the IEEE International Conference on Cluster Computing*, Boston, Massachusetts.
2. Brightwell, R., Hudson, T., Pedretti, K., et al. (2006). SeaStar Interconnect: Balanced Bandwidth for Scalable Performance, *IEEE Micro*, Vol. 26, No. 2.

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program and a Cooperative Research and Development Agreement with Intel.***

# CATAMOUNT LIGHTWEIGHT OPERATING SYSTEM MAXIMIZES MULTI-CORE COMPUTE RESOURCES WHILE MINIMIZING POWER

Contacts: Suzanne M. Kelly, [smkelly@sandia.gov](mailto:smkelly@sandia.gov)  
 John VanDyke, [jpvandy@sandia.gov](mailto:jpvandy@sandia.gov)  
 Courtenay Vaughan, [ctvaugh@sandia.gov](mailto:ctvaugh@sandia.gov)  
 James H. Laros III, [jhlaros@sandia.gov](mailto:jhlaros@sandia.gov)  
 Ron Brightwell, [rbbrigh@sandia.gov](mailto:rbbrigh@sandia.gov)

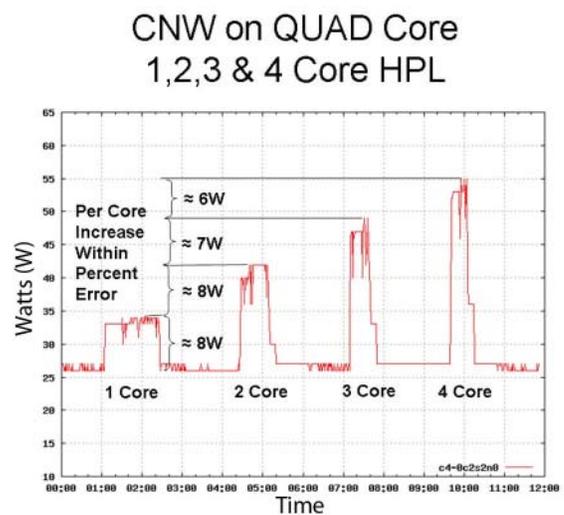
*Our work on the Catamount lightweight kernel increases the scalability, performance, effectiveness, and efficiency of the parallel computing platforms used for delivering advanced modeling and simulation solutions for science-based engineering.*

We have developed an efficient, high-performance operating system for massively parallel computers based on multicore processors: the Catamount N-way Lightweight Kernel. In high-performance computing, the operating system is critical for delivering the full capabilities of the system to applications. Parallel computing applications are characterized by being resource constrained. The need to parallelize an application is driven by the desire to exceed the resources (e.g., compute cycles or memory capacity) available on a single computer. As such, the role of the operating system is to maximize the amount of available resources delivered to parallel applications so that they can finish as quickly as possible.

Sandia began addressing the need for specific-purpose operating systems more than a decade ago by exploring lightweight kernels for some of the first massively parallel processing (MPP) systems. Our lightweight kernels optimize application performance by employing resource management policies and mechanisms that are specific to the needs of scientific parallel applications. The Catamount lightweight kernel is a third-generation operating system developed and deployed on the Cray XT series of machines. Recently, Catamount was enhanced to provide support for multi-core processors.

There are many significant technical challenges in developing and supporting a production-quality operating system for scientific calculations running in parallel mode across thousands of processors (nodes). The most recent technical challenges facing the Catamount lightweight kernel were providing support for multi-core processors and reducing CPU power consumption. The internal workings of Catamount were extended to support running application processes on more than two cores per node. These multi-core enhancements to Catamount resulted in a new version,

Catamount N-Way (CNW), which can support current and future large-scale multi-core nodes (Figure 1). The name is derived from the fact that the implementation was done assuming N cores per processor socket. It is targeted for values of N between 1 and 16, although it has only been tested on up to four cores per socket. A highly efficient shared memory capability, called SMARTMAP was introduced in the kernel. This feature has been exploited using the OpenMPI library with excellent results [1]. To address the escalating power requirements of MPP systems, CNW was modified to reduce power draw when cores are idle — this work will result in significant power savings for Red Storm. Individual cores are halted when not in use, which dramatically reduces their power draw. Traditional operating systems consume so much of the processor that it is nearly impossible to idle the cores sufficiently to halt them all.



**Figure 1.** Catamount N-Way power usage during and between executions of High Performance Linpack (HPL). Power drops to a minimal level when cores are not in use.

Our work on the Catamount lightweight kernel increases the scalability, performance, effectiveness, and efficiency of the parallel computing platforms used for delivering advanced modeling and simulation solutions for science-based engineering. In 2008, Sandia’s Cray XT Red Storm machine was upgraded to contain a mixture of dual- and quad-core nodes, and the Catamount N-Way lightweight kernel was a key technology that allowed the system to continue to function and deliver maximum performance to applications.

Figure 2 compares the performance of four applications on dual- and quad-core processors. Sage and LAMMPS were run on 2048 cores, while POP was run on 4800 cores and CTH was run on 6000 cores. The results for each application running different problem sets show the relative performance on dual-core processors (in blue) with the same number of cores using quad-core processors (in red). For most applications, using quad-core processors was only slightly worse than using dual-core processors. Based on the data, using quad-core processors increases the machine capacity by more than 38%, which is significant given the 8% increase in cost for the upgrade for the quad-core processors.

Additionally, our power modifications have allowed us to realize a 50% per CPU savings, which, based on past metrics, will reduce our facility power costs by 100s of thousands of dollars.

**References**

1. Brightwell, R., Hudson, T., Pedretti, K. T. (2008). SMARTMAP: Operating System Support for Efficient Data Sharing Among Processes on a Multi-Core Processor, International Conference for High Performance Computing, *Networking, Storage, and Analysis (SC’08)*, Austin.

*This work has been supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program.*

**Dual vs Quad Comparison: Red Storm**

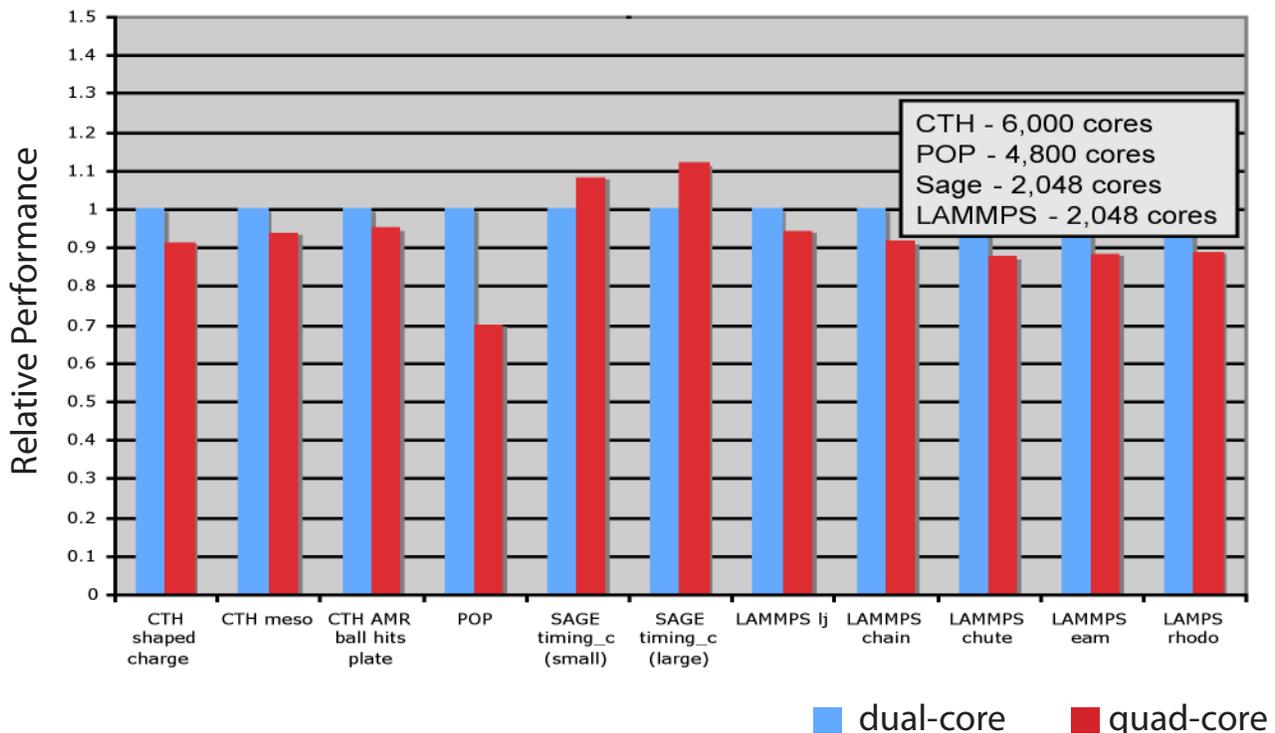


Figure 2. Performance of four applications on dual- and quad-core processors.

# ALLOCATION ALGORITHMS MAXIMIZE THROUGHPUT IN PARALLEL SUPERCOMPUTERS

Contacts: Vitus Leung, Computer Science & Informatics, [vjleung@sandia.gov](mailto:vjleung@sandia.gov)  
Cindy Phillips, Discrete Math & Complex Systems, [caphill@sandia.gov](mailto:caphill@sandia.gov)

*The CPA (at less than one percent of the cost of a parallel computer) is an example of how a relatively small investment in computer algorithms can dramatically leverage the return on a large investment in computer hardware.*

In collaboration with researchers from Stony Brook University and the University of Illinois, Urbana-Champaign, CCIM has developed innovative solutions to resource allocation for parallel processing on supercomputers [1, 2] (Figure 1). Remarkably, the allocation algorithms use space-filling curves to reorder a network of processors so that locations remain close in the physical network of processors after reordering (Figure 2). They also use bin packing (Figure 3) for contiguous allocation and span minimization for noncontiguous allocation to find open processors that are close to each other. In experiments, this optimized node allocation strategy increased throughput by 23%, thus processing five jobs in the time it normally takes to process four.

The first production uses of some of these algorithms were in Cplant system software released in 2002 and 2003.

One of the current production uses of some of these algorithms is in the Compute Process Allocator (CPA). After extensive prototype development, the CPA's innovative solution was carried to the commercial sector in 2005 when CPA was licensed to Cray, Inc. The breadth of impact has been extended through software licensing to numerous laboratory and research centers that bought XT3/4 systems from Cray. Sites with supercomputers using CPA include SNL's Red Storm. The CPA (at less than one percent of

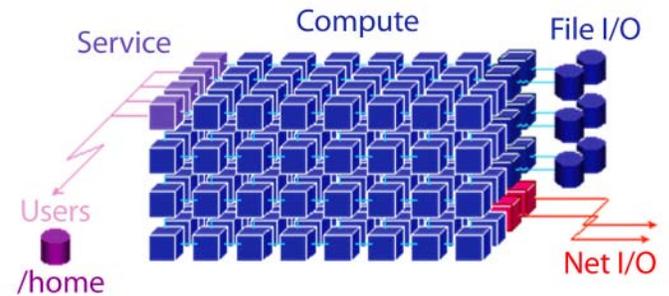


Figure 1. Parallel supercomputer.

the cost of a parallel computer) is an example of how a relatively small investment in computer algorithms can dramatically leverage the return on a large investment in computer hardware.

With CPA, Sandia has leveraged research sponsored by the Laboratory Directed Research and Development (LDRD), Computer Science Research Foundation (CSRF), and Advanced Scientific Computing Research (ASCR), filed two patent applications on the technical advances underlying CPA, and affected the external community through software licensing. For its superior strategy and scalability over other allocators, the CPA won a prestigious 2006 R&D 100 Award.

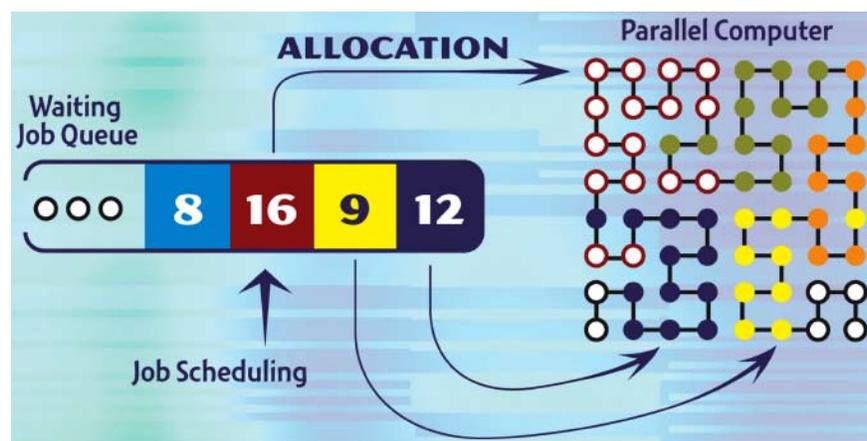


Figure 2. Allocation Algorithm with Hilbert Space-Filling Curve and Span Minimization.

The other current production use of one of these algorithms is in LLNL's Simple Linux Utility for Resource Management (SLURM). SLURM provides an Applications Programming Interface (API) for integration with external schedulers such as the Moab Cluster Suite. SLURM provides resource management on about 1000 computers worldwide, including some of the most powerful computers in the world.

### References

1. Bender, M. A., Bunde, D. P., Demaine, E. D., et al. (2008). Communication-Aware Processor Allocation for Supercomputers: Finding Point Sets of Small Average Distance," *Algorithmica*, Vol. 50, No. 2, pp. 279-298.

2. Leung, V. J., Arkin, E. M., Bender, M. A., et al. (2002). Processor Allocation on Cplant: Achieving General Processor Locality Using One-Dimensional Allocation Strategies, *Proceedings IEEE International Conference on Cluster Computing*.

*This work's original support was from SNL's Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the Department of Energy Advanced Scientific Computing Research (ASCR) Program and the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

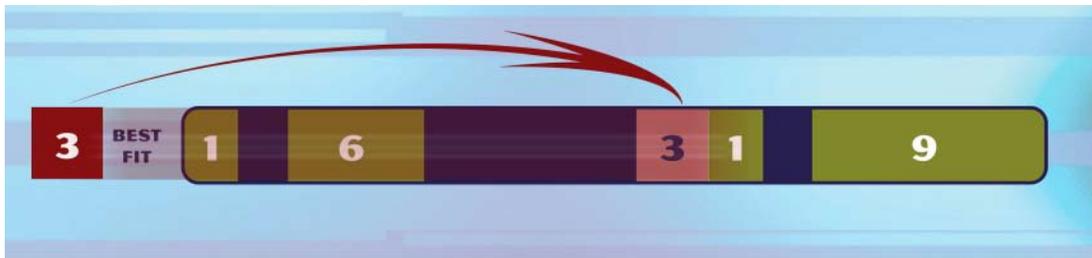


Figure 3. Best-fit packing.

# ENABLING LARGE-SCALE INFORMATICS USING NOVEL ARCHITECTURES: PYTHIA AND X-CALIBER

Contacts: Richard C. Murphy, Scalable Computer Architecture, [rcmurph@sandia.gov](mailto:rcmurph@sandia.gov)  
 Arun F. Rodrigues, Scalable Computer Architecture, [afrodri@sandia.gov](mailto:afrodri@sandia.gov)  
 K. Scott Hemmert, Scalable Computer Architecture, [kshemme@sandia.gov](mailto:kshemme@sandia.gov)  
 Brian W. Barrett, Scalable Computer Architecture, [bwbarre@sandia.gov](mailto:bwbarre@sandia.gov)  
 Subhash Shinde, Advanced Microsystem Packaging, [slshind@sandia.gov](mailto:slshind@sandia.gov)  
 Doug Trotter, Radiation Hardened CMOS Technology, [dctrott@sandia.gov](mailto:dctrott@sandia.gov)

*Sandia is addressing this new application domain and the associated scalability challenges with a two-pronged approach: first, in the Pythia project we are working to influence the commodity architecture roadmap to support informatics applications; and second, we have developed the more custom X-caliber Architecture to be designed from the ground-up to address large-scale informatics problems.*

High Performance Computing (HPC) applications have become more data intensive than in the past. Emerging *Informatics* applications are characterized by very large irregular data sets. Rather than attempting to model the world mathematically, informatics applications examine data already produced by either real world collection or simulation in an attempt to derive knowledge. They exhibit significantly different and more difficult data movement patterns than do traditional applications. Lower spatial locality (the use of data near data already being used), lower temporal locality (the reuse of data that has already been used), and larger data sets are the norm. This is exhibited in Figure 1 (the size of the dots represents the relative data set size). Because computer performance is primarily bound by data movement (through the memory system and the network), this is a recipe for disastrously poor performance and scalability.

Sandia is addressing this new application domain and the associated scalability challenges with a two-pronged approach: (1) in the Pythia project we are working to influence the commodity architecture roadmap to support informatics applications, and (2) we have developed the more custom X-caliber architecture to be designed from the ground up to address large-scale informatics problems.

Pythia and X-caliber are engagements among Sandia, Micron Technologies, Sun Microsystems, Louisiana State University, University of Southern California/ISI, Indiana University, the University of Notre Dame, and the University of Maryland.

*Pythia.* The Pythia architecture addresses the data movement challenges enumerated above at all levels: memory, network, and storage subsystem. The processor provides a balance between heavily multithreaded high throughput computing and single thread performance (unlike architectures such

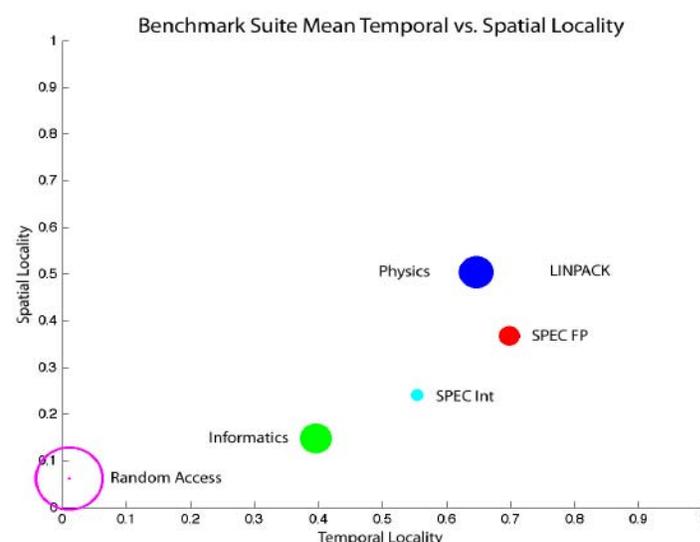


Figure 1. Spatial and temporal locality of key application suites.

as the Cray XMT that focus almost exclusively on the former, or accelerators more oriented toward the latter). As compared to conventional architectures, the memory system provides significantly increased concurrency, bandwidth, and capacity. We are developing a custom network interface (NIC) that provides an improved messaging rate, higher bandwidth, and hardware support for both MPI acceleration and Partitioned Global Address Space (PGAS) languages. The storage system includes both technological and programmability improvements. Rather than a flat disk array, Pythia provides flash-based storage in the memory hierarchy (with a resulting 2 to 5X improvement in IOPS). The machine additionally provides intelligent data query capabilities via LexisNexis' declarative Enterprise Control Language (ECL) for moving data through the storage hierarchy. This facilitates a programming model in which petabytes of data on disk can be reduced to tens of

terabytes for further processing in-core during each phase of computation.

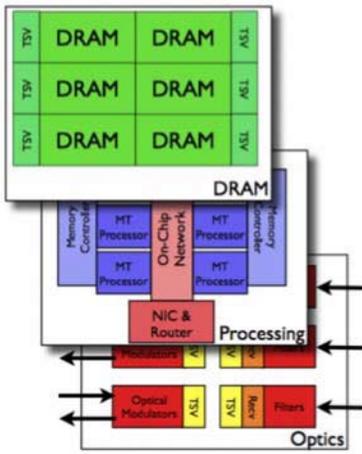


Figure 2. The X-caliber 3D stacked architecture.

*X-caliber* (Figure 2). In the longer term, we are interested in architectures that both address informatics and physics applications in a more power-efficient fashion. The *X-caliber* architecture does so with an eye towards meeting DOE's exascale computing requirements in the next decade. Simply scaling today's peta-scale Massively Parallel Processor (MPP)

architectures to the exascale would produce a machine that consumes 200 or more acres of space and exceeds a 4 gigawatt power budget. *X-caliber* addresses many of these power problems through the use of 3D integration, a key strength of Sandia's microelectronics organization. Additionally, *X-caliber* extends Sandia's long-term research into NIC/router design and silicon photonics. Each of these technologies represents a critical approach for reducing power in the overall compute system. A DARPA exascale report explains that most power consumption in exascale machines will result from data movement operations through the network and memory hierarchy. The 3D integration enabled by *X-caliber* reduces centimeter length wires to micron length, enabling the design of simpler, more efficient communication mechanisms.

Since the inception of the electronic computer in the mid-1940s, memory access has restricted performance (typically known as the *von Neumann Bottleneck* or *The Memory Wall*). The *X-caliber* project focuses on significantly lower power multithreaded processor cores tightly bound to dense DRAM with the use of Through Silicon Vias (TSVs), enabling more efficient local memory access. Distance is one of the more critical challenges to the construction of modern supercomputers. It is seen in many forms: between nodes connected by a cable, or between two chips on a board connected by wires, or even from one functional unit to another on a single chip. In today's machines, where the power to perform a mathematical operation is significantly less than the power required to move the data necessary to support that operation to the appropriate functional unit, reducing distance enables the creation of more powerful machines. The last three decades have seen processors go from being constructed from discrete transistors on a board,

to the modern microprocessor consuming a single chip, and finally to the multicore era with multiple processors on a single die. The integration of heterogeneous things, however, is significantly harder (requiring more complex fabrication or longer distance communication). For example, processors and memory are fabricated differently, with the former emphasizing transistor performance and the latter density. Optical interconnects, such as the silicon photonics proposed in *X-caliber*, are similarly different in fabrication. Historically, creating a single, unified fabrication process that encompasses all steps from all types of component technology has proven economically unfeasible.

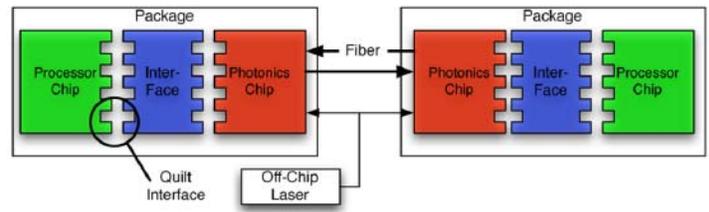


Figure 3. A quilt packaged architecture.

Sandia's approach is to examine two key technologies for very tight integration to address both the problems of distance and heterogeneity. The 3D integration proposed in *X-caliber* stacks chips (of potentially different types) on top of each other in three dimensions. The TSVs cutting through the stack reduce wire lengths from centimeters to microns. We are also examining Quilt Packaging (QP, depicted above), which integrates multiple chips in a planar fashion, similarly reducing wire lengths. Each enables heterogeneous components to be inexpensively combined into a single, lower power device.

## References

- Kogge, P. M. (ed.). *ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems*, University of Notre Dame, CSE Dept. Tech. Report TR-2008-13, Sept. 28, 2008.
- Murphy, R. C., On the Effects of Memory Latency and Bandwidth on Supercomputer Application Performance, (2007). *IEEE International Symposium on Workload Characterization 2007 (IISWC07)*, Boston, MA, September 27-29, 2007.
- Murphy, R. C., and Kogge, P. M. (2007). On the Memory Access Patterns of Supercomputer Applications: Benchmark Selection and Its Implications, *IEEE Transactions on Computers*, Vol. 56, No. 7, pp. 937-945.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

# KITTEN LIGHTWEIGHT KERNEL MAKES MULTI-CORE PROCESSORS PURR

Contact: Kevin Pedretti, Scalable System Software, [ktpedre@sandia.gov](mailto:ktpedre@sandia.gov)

*Our next-generation open-source LWK, called Kitten, is designed specifically for current and future multi-core processors, which will likely form the basis for upcoming exascale computing platforms.*

The computing industry's shift to relying on increased parallelism to increase performance, rather than increasing clock speeds, will significantly increase the number of processors in capability supercomputers (Figure 1, [1]). This will exacerbate the barriers to application scalability that we face today, such as operating system interference. Lightweight kernels (LWKs), with their very low overhead, are an effective solution to this problem.

Sandia has pioneered the design, development, and deployment of a series of highly scalable LWK operating systems for massively parallel processing systems. Unlike general-purpose operating systems, LWKs are designed to maximize the performance, scalability, and reliability of large-scale parallel computing platforms. Our next-generation open-source LWK, called Kitten, is designed specifically for current and future multi-core processors, which will likely form the basis for upcoming exascale computing platforms.

Sandia has a very successful history of deploying previous generations of highly scalable LWKs on several massively parallel computing platforms, from early systems like the Intel Paragon to the more recent Cray XT platform. LWKs have demonstrated significant advantages over full-featured commodity operating systems at the largest scale on today's capability supercomputers. We believe that LWKs will be even more important for future exascale many-core computing environments, both as a critical production system software component as well as an ideal research vehicle for exploring operating system technology.

This project is developing a new open-source LWK operating system, called *Kitten* [2], which addresses many of the common criticisms of previous LWKs, thereby significantly increasing its appeal to application developers, system vendors, and academic researchers. *Kitten* provides a user-level environment that is largely compatible with Linux, allowing application developers to continue to use standard compilers, system libraries, and programming models. Additionally, when coupled with the lightweight Palacios hypervisor from Northwestern University and the University of New Mexico, *Kitten* can launch and manage

## Challenge: Exponentially Increasing Parallelism

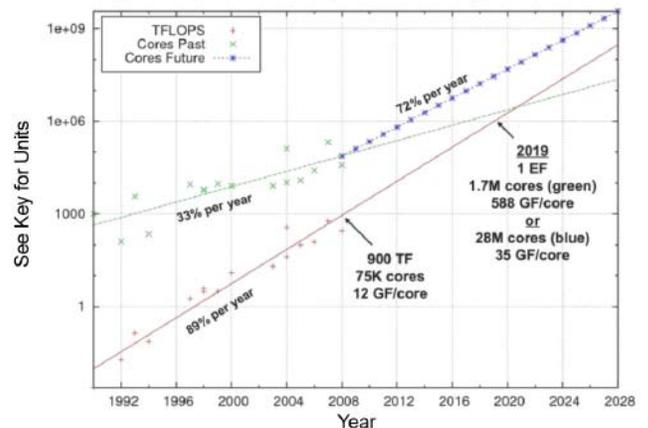


Figure 1. The number of cores in future machines is expected to grow exponentially, requiring a renewed focus on scalable system software.

guest operating systems to accommodate applications that require functionality not provided by the default user environment. Finally, we are leveraging the simplicity of the LWK environment to explore optimization opportunities for many-core processors. The first of these is the *SMARTMAP* [3] technique (Figure 2) that allows processes executing in the same many-core processor to communicate with each other directly, rather than using existing methods that require multiple inefficient memory copies. This enables MPI-based applications to cut their memory bandwidth usage for message passing in half, thereby making better use of an increasingly scarce resource on many-core processors. This software is available for download at: <http://software.sandia.gov/trac/kitten>

At the single-node scale, the performance of the underlying memory subsystem becomes more important as the number of processor cores per chip continues to increase. To make effective use of the available compute power, applications will likely have to become much more sensitive to the way they access memory. Techniques such as *SMARTMAP* allow for more efficient intra-node communication. This has been shown to result in significantly better MPI collective performance.

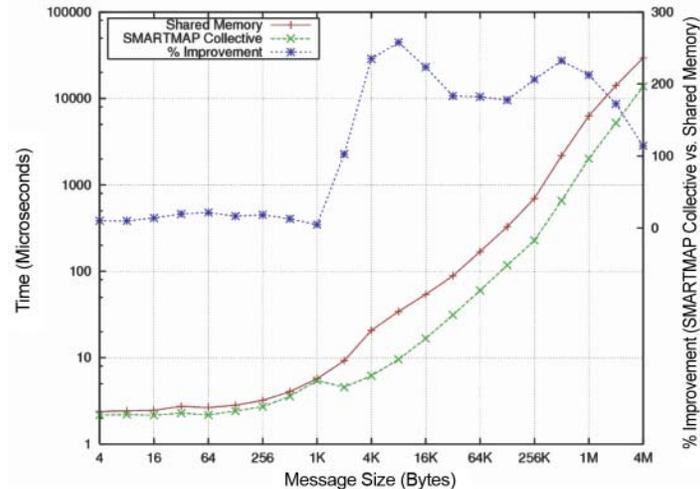
Supercomputing is of critical importance to national competitiveness and national security. This project is addressing the system software needs of capability supercomputing that are not being met by the general computing industry. We are pursuing vendor partnerships to transition the technology developed by this project into use on future capability supercomputer platforms.

### References

1. Pedretti, K. T., Kelly, S. M., Levenhagen, M. J. (May 2008). *Summary of Multi-Core Hardware and Programming Model Investigations*, SAND2008-3205, Sandia National Laboratories, Albuquerque, NM, May 2008.
2. Pedretti, K. T. (2008). *Quad-core Catamount and R&D in Multi-core Lightweight Kernels*, presentation, Salishan Conference on High-Speed Computing, April 2008.

3. Brightwell, R., Hudson, T., Pedretti, K. T. (2008). *SMARTMAP: Operating System Support for Efficient Data Sharing Among Processes on a Multi-Core Processors*, *International Conference for High Performance Computing, Networking, Storage, and Analysis (SC'08)*, Austin, TX, November 2008. Available through: <http://portal.acm.org/dl.cfm?coll=portal&dl=ACM&CFID=24376423&CFTOKEN=62367353>

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***



**Figure 2.** SMARTMAP collectives cut intra-node bandwidth usage by half compared to conventional shared memory techniques, resulting in up to 255% better performance for a four process Allreduce on a quad-core processor.

# THE ALLIANCE FOR COMPUTING AT EXTREME SCALES (ACES) TARGETS PRODUCTION CAPABILITY COMPUTING FOR NNSA

Contacts: Sudip Dosanjh, Computer & Software Systems, [sudip@sandia.gov](mailto:sudip@sandia.gov)  
Jim Ang, Scalable Computer Architecture, [jaang@sandia.gov](mailto:jaang@sandia.gov)

*Joint teams have been formed to design, architect, deploy, and operate future NNSA ASC production capability platforms. These platforms will support weapons physics and engineering calculations that require running across an entire platform.*

On March 7, 2008, Thomas O. Hunter, President of Sandia National Laboratories, and Michael R. Anastasio, Director of Los Alamos National Laboratory (LANL), signed a Memorandum of Understanding formalizing the creation of the New Mexico Alliance for Computing at the Extreme Scale (ACES). ACES will provide High Performance Computing (HPC) capability computing for NNSA's stockpile stewardship mission. Independently, Sandia and LANL have been world leaders in HPC for several decades. By bringing these two organizations together in the ACES partnership, stronger, more tightly integrated technical teams will emerge, efficiencies will be realized, and facilities at the two institutions will be optimized. Joint teams have been formed to design, architect, deploy, and operate future NNSA, Advanced Simulation and Computing (ASC) production capability platforms. These platforms will support weapons physics and engineering calculations that require running across an entire platform.

As part of ACES, both laboratories share intellectual capabilities and capitalize on their existing expertise in developing architectures and designs for future platforms. Los Alamos's Strategic Computing Complex (SCC) facility will house high-performance capability computing assets needed to support NNSA's ongoing stockpile stewardship mission, and to meet the ASC Roadmap timeline requirement for an exascale capability by 2018. The first of these ACES capability platforms will be named Zia and is targeted for installation in FY10. It will support the needs of all three NNSA laboratories.

The goals of ACES are to provide production capability computing for NNSA and national leadership in HPC. Production requires high reliability and usability; capability means that a single application must execute effectively across the entire machine. These requirements place challenges on scalability, resiliency, and energy-efficiency of capability systems, while providing support for the broad portfolio of MPI-based ASC-developed applications. Key

strategies include the following:

- Aligning with and influencing industry roadmaps
- Co-architecting platforms, applications, and algorithms, recognizing that new architectures are likely to influence applications and algorithms
- Ensuring pragmatic migration of ASC codes to new platforms with significant performance gains
- Encouraging and fostering credible competition in the supercomputing industry (procurements will be open and competitive)
- Focusing on a broad range of applications
- Impacting the supercomputing industry through market acceptance of designs and component technologies
- Being driven by cost, risk, and benefit analyses
- Partnering with the DOE's Office of Advanced Scientific Computing Research, by leveraging the Institute for Advanced Architectures and Algorithms initiative (IAA), and other government agencies (notably DARPA and other DoD agencies).

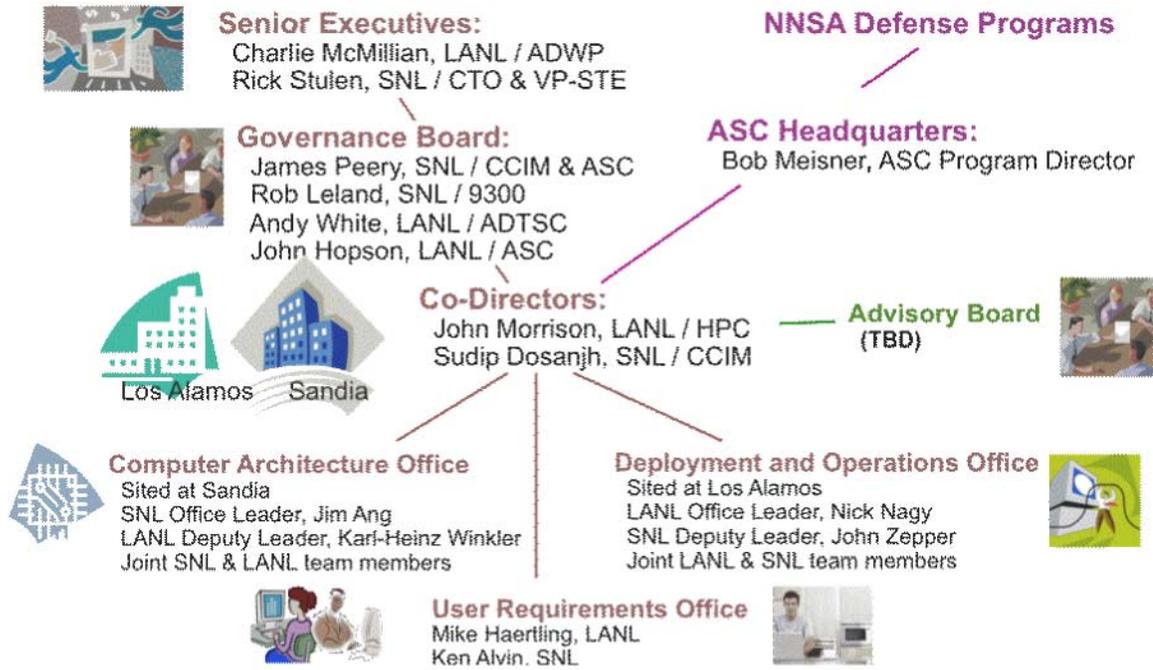


The goals of ACES partnership include designing and deploying Zia, the 2010 ASC production capability platform, and fostering the development of a technology foundation for future capability platforms. The ACES Architectures Office was established to develop technical requirements and specifications for Zia, develop the formal statement of work, issue the RFP for Zia, and execute the NNSA's Critical Decision (CD) process.

At the beginning of FY09, the ACES Architecture Office received a request from NNSA/ASC to assume technical oversight responsibility for a new technology development and engineering project. Because the target timeframe for

this technology development effort would intersect the Trinity timeframe for a 2015 ASC production capability platform, ACES accepted this responsibility.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

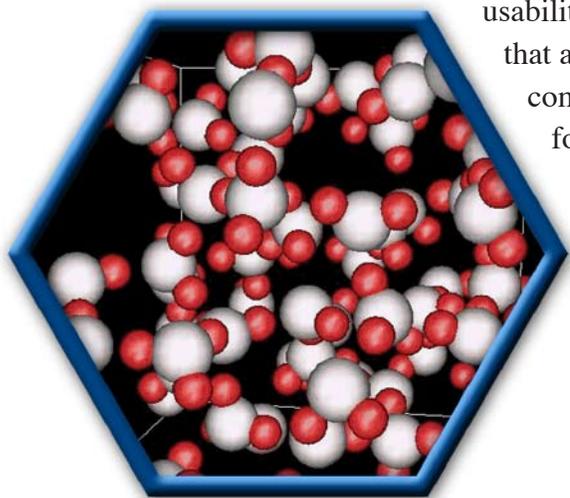


*End of Section III*



# *DRIVING APPLICATIONS*

Driving applications are important partners with algorithmic and architectural work in the Computation, Computers, Information and Mathematics (CCIM) Center 1400. Suitable applications are those that (1) have promising significant impact on Sandia and DOE missions, (2) have significant complexity in the underlying subject matter domain, and (3) demand HPC computing capabilities that are on the frontier of what can currently be achieved. Our ideal applications stress our algorithms and the most capable existing hardware architectures. Moreover, the applications work includes the role of uncertainty quantification and formal verification and validation. We are increasingly emphasizing the usability of delivered application work to inform complex decisions that are relevant to national security policies. Applications of our computational capabilities supporting nuclear weapons work, for example, are increasingly applied within the context of Quantification of Margins and Uncertainties (QMU). In other areas, such as our analysis of Arctic climate change, the connection to national security policy decision-making is the basic rationale for the work.



# UNPRECEDENTED SCALE SEMICONDUCTOR MODELING CONDUCTED ON MASSIVELY PARALLEL PLATFORMS

Contacts: Paul Lin, Electrical and Microsystem Modeling, [ptlin@sandia.gov](mailto:ptlin@sandia.gov)  
John Shadid, Electrical and Microsystem Modeling, [jnshadi@sandia.gov](mailto:jnshadi@sandia.gov)

*Sandia scientists and computational methods researchers are performing simulations of damage to semiconductor devices due to irradiation at unprecedented scale. This has been achieved by a combination of advanced solution techniques and access to massively parallel platforms.*

Sandia scientists and computational methods researchers are performing simulations of damage to semiconductor devices resulting from irradiation at an unprecedented scale. This has been achieved by a combination of advanced solution techniques and access to massively parallel platforms.

The performance of electronics in hostile environments is of critical importance to the weapons program. As semiconductor devices are being irradiated, displacement damage produces a large number of defect species. The combination of a mesh fine enough to provide sufficient resolution and this large number of defect species rapidly leads to numerical simulations with a very large number of unknowns. For realistic 2D transient simulations, these problems can be of the order of 10 million to one billion unknowns. Present 2D transient simulations require 2000 to 8000 processors and a week of run time. Figure 1 shows the steady-state electric potential solution of a typical bipolar junction transistor (BJT). Three-dimensional simulations will have two to three orders of magnitude more unknowns than 2D simulations.

Numerical solution of semiconductor problems with such a large number of unknowns required the development of a massively parallel finite element fully implicit Newton-Krylov simulation code. Semiconductor devices are modeled by the drift-diffusion equations plus additional

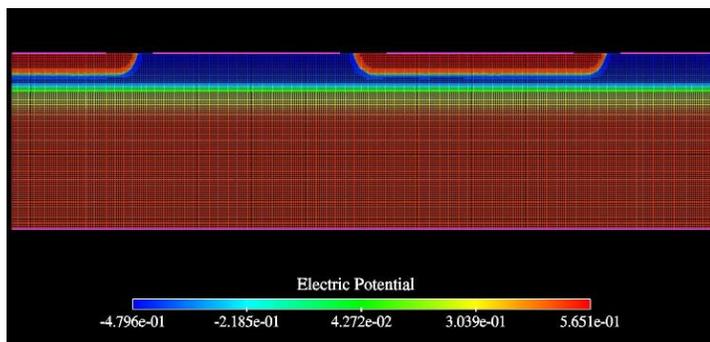


Figure 1. Steady-state electric potential solution for BJT.

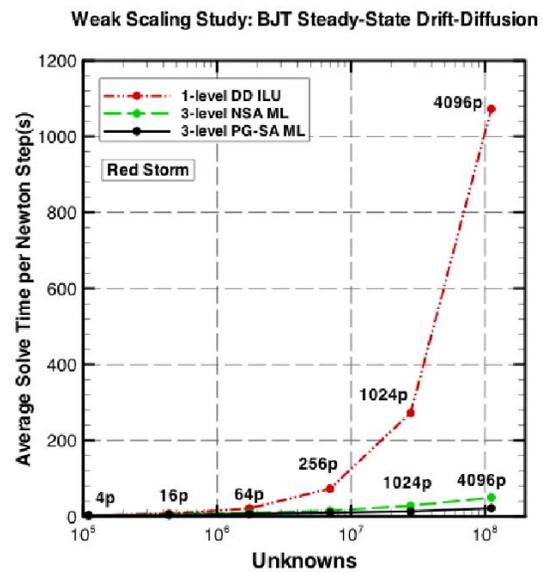


Figure 2. Comparison of one-level with two multigrid preconditioners.

species transport equations for the defect species. The Newton method solution approach generates large linear systems that are solved by Krylov methods. The choice of the preconditioner is critical to the parallel scaling and reducing the solution time for these linear systems. This research team has been applying multigrid preconditioners to the drift-diffusion system and another PDE system [1] via the Trilinos solver project. Figure 2 shows a weak scaling study on the SNL Red Storm machine for a BJT comparing the linear solve time for a one-level incomplete lower/upper (ILU) factorization (red line) with a nonsmoothed aggregation ML (green line) and a Petrov-Galerkin smoothed aggregation ML preconditioner (black line). The steady-state problem is scaled up to 4000 processors and 100 million unknowns, and for the largest problem, ML is almost 50 times faster than the one-level preconditioner. This research team has also been developing approximate block factorization and physics-based preconditioners.

In addition to advanced solution methods, high-fidelity simulation requires access to large NNSA parallel platforms such as the ASC Red Storm Cray XT3/4 machine at SNL and the ASC Purple IBM machine at LLNL. Access to these platforms allows 2D transient simulations that enable investigation into two-dimensional effects that are not captured by the 1D transient simulations on smaller Linux clusters. Future platforms will allow investigation into three-dimensional effects. Simulations have been performed on ASC Purple and ASC Red Storm on problems as large as one billion unknowns and the semiconductor device simulator has been tested on 38,000 cores of Red Storm.

It is worth noting that these unprecedented scale semiconductor modeling simulations will also be of interest to the satellite community because of concerns of damage from cosmic rays and other radiation.

### **Reference**

1. Lin, P. T., Sala, M., Shadid, J. N., Tuminaro, R. S. (2006). Performance of Fully-Coupled Algebraic Multilevel Domain Decomposition Preconditioners for Incompressible Flow and Transport, *International Journal for Numerical Methods in Engineering*, Vol. **67**, No. 2, pp. 208-225.

***This work has been supported by National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

# NEW SOLUTION METHODS FOR MULTIPLE-TIME-SCALE RESISTIVE AND EXTENDED MAGNETOHYDRODYNAMICS

Contacts: John N. Shadid, Electrical and Microsystem Modeling, [jnshadi@sandia.gov](mailto:jnshadi@sandia.gov)  
 Roger P. Pawlowski, Applied Math and Applications, [rppawlo@sandia.gov](mailto:rppawlo@sandia.gov)  
 Paul T. Lin, Electrical and Microsystem Modeling, [ptlin@sandia.gov](mailto:ptlin@sandia.gov)  
 Luis Chacon (Co-PI), Fusion Energy Division, Oak Ridge National Laboratory, [chaconl@ornl.gov](mailto:chaconl@ornl.gov)

*This work is a step towards developing scalable petascale solution methods for the next generation of predictive MHD simulations for the DOE Office of Science.*

Sandia and Oak Ridge scientists, and applied mathematics, and computational science researchers, have demonstrated advanced scalable parallel solution methods for MHD fluid models of plasma systems with strong electromagnetic effects. These systems are common in the natural physical world and are highly important for many advanced technological applications. Examples of the former include astrophysical systems, such as stellar interiors and atmospheres, gaseous nebulae, and the interaction of the Earth's magnetosphere with the "solar wind," producing space weather. Among the scientific and technological applications of importance to DOE are fusion energy devices (such as tokamak reactors, e.g., ITER) and some aspects of important high-energy-density physics experiments (e.g., NIF and Z-pinch). The resulting mathematical models for these systems are strongly coupled, highly nonlinear, and characterized by multiple physical phenomena that span a very large range of length and time scales and are extremely challenging to solve numerically.

To deal with this extreme range of scales, a research team has developed and demonstrated high-resolution and adaptive spatial discretizations with fully implicit time integration techniques that can be stable and accurate when run at the dynamical time-scale of interest in these multiple-time-scale systems. However, the significant benefits of these techniques come at a high price — very large-scale strongly coupled nonlinear systems that must be solved. To make the solution of these large-scale problems tractable, the Sandia and Oak Ridge teams have developed unique approaches based on parallel multi-level preconditioned Newton-Krylov methods.

These techniques utilize physics-based and block-oriented preconditioners that reduced these complex coupled equations to systems that are solvable by scalable multilevel solution strategies [1, 2]. These methods have been demonstrated on MHD systems with up to 100 million unknowns on 4096 processors [1]. Recently, these methods have been demonstrated at an extreme scale of 1 billion unknowns and executed on up to 24,000 cores of the Cray XT system at

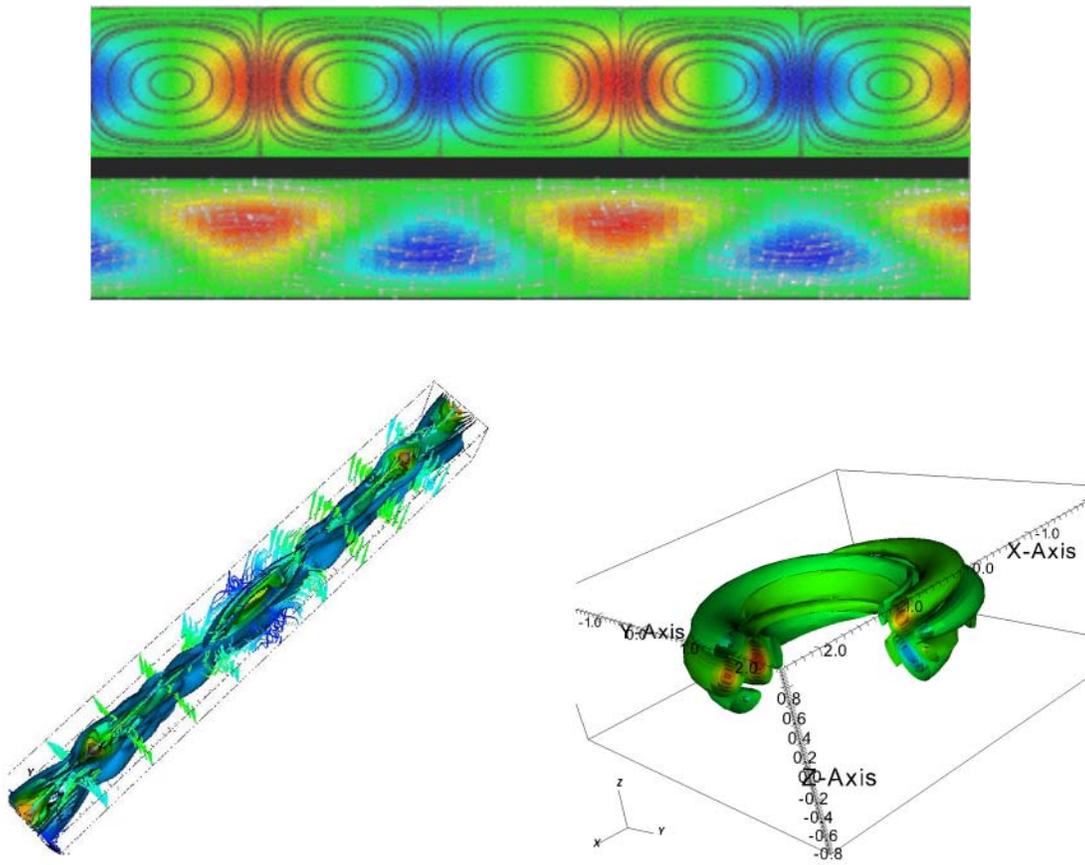
Sandia for solution of a prototype MHD pump. This work is a step towards developing scalable petascale solution methods for the next generation of predictive MHD simulations for the DOE Office of Science.

The Sandia and Oak Ridge research team has employed these robust and scalable solution methods on a number of challenging MHD systems. These include internal duct flows with external and induced magnetic fields (e.g., a prototype MHD pump); thermal-buoyancy induced flows with magnetic fields (e.g., mechanisms important in the geo-dynamo and Earth's magnetic field); fast magnetic reconnection studies (e.g., magnetic-island-coalescence and shear-generated instabilities) and magnetic confined fusion reactor scale instabilities (Figure 1).

## References

1. Chacon, L. (2008). An Optimal, Parallel, Fully Implicit Newton-Krylov Solver for Three Dimensional Visco-Resistive Magnetohydrodynamics, *Phys. Plasmas*, Vol. **15**, No. 05.
2. Lin, P. T., Sala, M., Shadid, J. N., Tuminaro, R. S. (2006). Performance of Fully-Coupled Algebraic Multilevel Domain Decomposition Preconditioners for Incompressible Flow and Transport, *International Journal for Numerical Methods in Engineering*, Vol. **67**, No. 2, pp. 208-25.

***This work has been supported by the Department of Energy's Advanced Scientific Computing Research (ASCR) Program.***



**Figure 1.** Large-scale Fundamental Studies of MHD Instabilities: Above: Plot of streamlines and vertical velocity color contours along with the plasma current for a nonlinear solution for a resistive MHD bifurcation study for the Hydromagnetic Rayleigh-Bernard stability problem. Below: 3D fully implicit extended MHD simulations. The left figure depicts densities and velocities in the nonlinear stage of a 3D Kelvin-Helmholtz unstable configuration with differential rotation. The figure on the right depicts temperature during the evolution of a helical perturbation in a 3D toroidal fusion device.

# RIGOROUS CODE VERIFICATION DRIVES MAJOR IMPROVEMENTS IN COMPUTATIONAL SHOCK MAGNETOHYDRODYNAMICS

Contacts: John Niederhaus, Comp. Shock and Multiphysics, [jhniede@sandia.gov](mailto:jhniede@sandia.gov)  
 Allen Robinson, Comp. Shock and Multiphysics, [acrobin@sandia.gov](mailto:acrobin@sandia.gov)  
 Bill Rider, Comp. Shock and Multiphysics, [wjrider@sandia.gov](mailto:wjrider@sandia.gov)  
 Tom Gardiner, HEDP Theory, [tagardi@sandia.gov](mailto:tagardi@sandia.gov)

*Systematic verification provides compelling evidence that the extended DeBar correction implemented in ALEGRA shock MHD removes errors in simulations that would otherwise remain deeply flawed — a clear example of the benefit to be gained by exercising the discipline of verification in scientific code development.*

Numerical modeling for magnetohydrodynamics (MHD) is a core competency within the computational physics effort at Sandia, and the 2D/3D MHD code ALEGRA is the vehicle for this work. Fundamental to many of the applications for which ALEGRA is used is accurate treatment of MHD shock waves. Here we describe two test problems that have emerged as particularly useful for evaluating the accuracy of shock-MHD treatment. These provide rigorous verification to facilitate aggressive test-driven code development for ALEGRA while ensuring fidelity to the underlying physics of MHD.

The two problems are magnetized reformulations of one-dimensional test problems from pure shock hydrodynamics: the Noh and Woodward-Colella (WC) problems. These involve strong MHD shock waves and provide stringent tests of the ability of the code to maintain energy conservation. This is of concern because it is well known that if energy is not conserved in the simulation, shocks arising in the computed solution may propagate with unphysical speeds and jump conditions, leading to erroneous results.

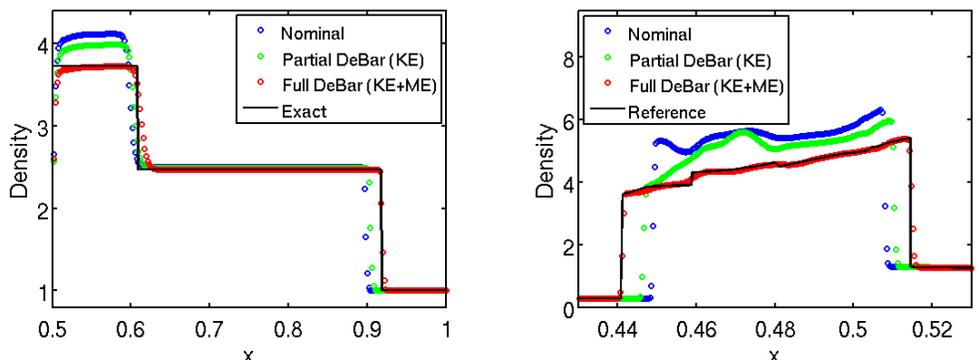
To test energy conservation in ALEGRA for shock-MHD, the magnetized Noh and WC problems have been simulated on a series of increasingly fine meshes. The fineness of the mesh is described by  $N$ , which is the number of elements spanning the domain. The results for  $N = 512$  and 1280 are

shown in Figure 1 with exact or highly resolved reference solutions [1]. The density discontinuities apparent in the solutions are strong MHD shock waves.

In these simulations, energy defects on the order of several percent have been observed. These losses give rise to significant errors in shock propagation speeds and post-shock states, as seen in the “nominal” cases in Figure 1. The energy loss has been traced to the advection scheme, which is the portion of the ALEGRA algorithm where motion relative to the mesh is computed. The advection scheme used by ALEGRA and similar codes is known to allow some energy to be lost when strong shocks are present, but is necessary for simulating certain classes of problems.

To address this issue, the DeBar energy correction has been implemented in many of these codes [1]. This correction accounts only for kinetic energy lost during advection, and the data in Figure 1 (“partial DeBar”) indicate that it is insufficient for the shock-MHD test problems here. The ALEGRA team has extended the correction to account for magnetic energy losses. With the full DeBar (kinetic + magnetic) correction, the energy-conserving property of the equations of MHD is maintained. As seen in Figure 1, shock speeds and post-shock states are thus computed accurately only when the full DeBar correction is used.

**Figure 1.** Density profiles computed using ALEGRA: (left) magnetized Noh problem with  $N=512$  elements and exact solution at  $t=0.75$ ; (right) magnetized Woodward-Colella (WC) problem at  $N=1280$  and reference computed solution at  $t=0.025$ .



These tests have been made more sensitive and rigorous using order verification for the density field. To perform order verification here, the total difference in the density relative to exact and reference solutions is summed. The convergence rate  $p$  for the calculation is then computed, which is the rate at which this error diminishes as the mesh is refined (as  $N$  is increased). The physics of these problems and the mathematics of the algorithms used in ALEGRA — if implemented correctly — require that convergence rates near 1.0 and 0.75 be obtained for the Noh and WC problems, respectively.

The total error in the density field is plotted versus the mesh density  $N$  in Figure 2. Here we observe a clear divergence in error trends between the cases where energy conservation is lost or is maintained via the DeBar correction. Only in the case where energy conservation is fully enforced in the simulations do we observe the expected convergence rates. The observed convergence rates for  $N > 512$  vary between 0.88 and 0.98 for the Noh problem, and between 0.7 and 0.75 for the WC problem.

Systematic verification provides compelling evidence that the extended DeBar correction implemented in ALEGRA shock MHD removes errors in simulations that would otherwise remain deeply flawed. This is a clear example

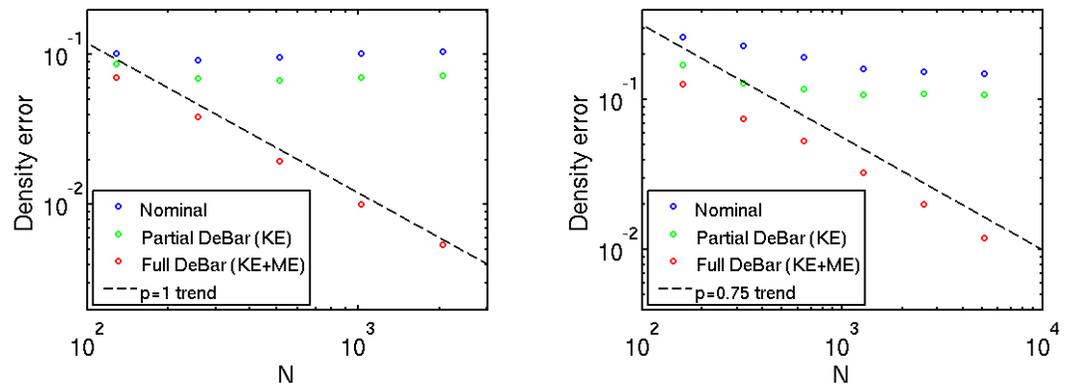
of the benefit to be gained by exercising the discipline of verification in scientific code development. Using order verification, success or failure in implementation of a new code feature is captured in a quantitative metric, reflecting the fidelity of the code to underlying physics.

### Reference

I. Niederhaus, J. H., Gardiner, T. A., Rider, W. J., Robinson, A. C. (2008). *Energy Conservation in Lagrange-Remap Modeling for Ideal-MHD Shock Propagation*, SAND-2008-8194C, Sandia National Laboratories, Albuquerque, NM.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*

**Figure 2.** Density error convergence trends for the (left) magnetized Noh and (right) magnetized Woodward-Colella (WC) problems at  $t=0.75$  and  $t=0.025$ . Dashed black lines indicate the expected convergence behavior for each problem.



# DEVELOPING SCALABLE PRECONDITIONERS FOR PARALLEL ELECTRICAL SIMULATION

Contacts: Heidi K. Thornquist, Electrical and Microsystem Modeling, [hkthorn@sandia.gov](mailto:hkthorn@sandia.gov)  
 Erik G. Boman, Scalable Algorithms, [egboman@sandia.gov](mailto:egboman@sandia.gov)  
 David M. Day, Applied Math and Applications, [dmday@sandia.gov](mailto:dmday@sandia.gov)  
 Ray S. Tuminaro, Scalable Algorithms, [rstumin@sandia.gov](mailto:rstumin@sandia.gov)

*The Sandia research team has integrated this new linear solution strategy into Xyce, Sandia's parallel electronic simulator, demonstrating its effectiveness on several large CMOS memory circuits. Previous to this work, Xyce was not able to simulate several of these circuits with standard preconditioning techniques.*

Sandia researchers are developing robust and scalable preconditioners for parallel electrical simulation to enable the simulation of very highly integrated circuits containing tens-to-hundreds of millions of devices. These new preconditioners combine effective global reordering strategies with state-of-the-art partitioning techniques to accelerate the solution of large-scale linear systems.

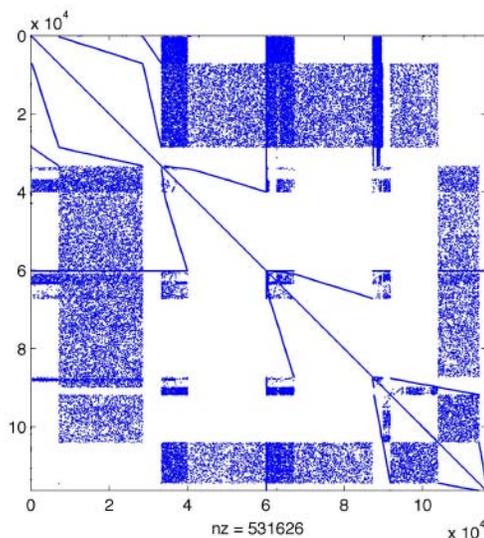
While advances in manufacturing enable the fabrication of integrated circuits containing tens-to-hundreds of millions of devices, the time-sensitive modeling and simulation necessary to design these circuits poses a significant computational challenge. This is especially true for mixed-signal integrated circuits where detailed performance analyses are necessary for the individual analog/digital circuit components as well as for the full system. When the integrated circuit has millions of devices, performing a full system simulation is practically infeasible using currently available Electrical Design Automation (EDA) tools.

The principal reason that traditional transistor-level circuit simulation does not scale is the time required to compute the solutions of large-scale linear systems. Linear systems

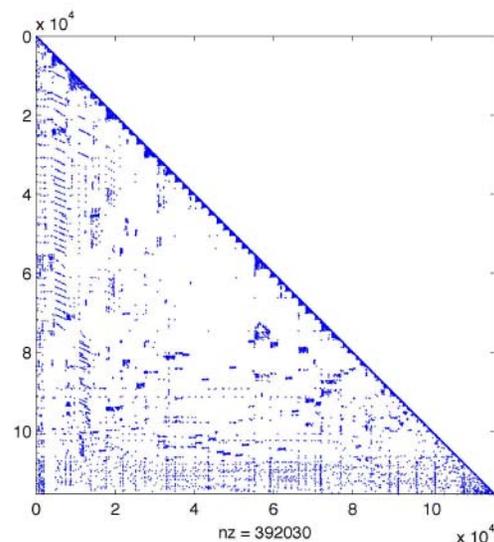
resulting from the simulation of circuits are sparse, typically have heterogeneous nonsymmetric structure (Figure 1), and are often ill-conditioned. Direct sparse linear solvers have been the industry standard approach for solving these linear systems because they are robust in the face of ill-conditioning and easy to use in practice. However, direct solvers suffer from poor scaling and become impractical when the linear system has hundreds of thousands of unknowns or more. Iterative solvers scale well, but their effectiveness is dependent upon the appropriate choice of matrix filtering, reordering, partitioning, and preconditioning, so conventional wisdom has been that such methods are not applicable for circuit simulation.

To motivate the use of iterative solvers in parallel electrical simulation, this research team has developed a new linear solution strategy that incorporates intelligent choices of matrix filtering, reordering, and partitioning to generate a scalable preconditioner [1]. Often, the most difficult linear systems to solve are those generated during the DC operating point (DCOP) calculation. Therefore, this new linear solution strategy finds an effective reordering and partitioning for the DCOP linear systems, which can then be reused

**Figure 1.** Matrix structure for CMOS memory circuit.



**Figure 2.** As discussed on the next page, block triangular form (BTF) structure for CMOS memory circuit.



for the transient simulation. The methods for reordering and partitioning come from the underlying combinatorial algorithms that make direct solvers so effective on these types of problems. However, the first step in this strategy is to filter any dense rows and columns, typically resulting from power supply and ground nodes, because they have the potential to increase communication costs dramatically. The reordering step of this strategy takes advantage of the fact that the Jacobian matrices generated during the DCOP calculation are often reducible and can be permuted into block triangular form (BTF). Exploiting the BTF structure (Figure 2) often gives great performance gains for both direct and iterative methods. Hypergraph partitioning [2, 3] is a state-of-the-art method that is employed in this strategy to partition the block structure of the resulting matrix structure (Figure 3). The resulting block diagonals are used to generate a block Jacobi preconditioner.

The Sandia research team has integrated this new linear solution strategy into Xyce, Sandia's parallel electronic simulator, demonstrating its effectiveness on several large CMOS memory circuits. Previous to this work, Xyce was not able to simulate several of these circuits with standard

preconditioning techniques. Using the preconditioner generated by this new linear solution strategy, Xyce can easily simulate this memory circuit. Furthermore, Xyce shows an impressive parallel speedup using this strategy (Figure 4).

### References

1. Thornquist, Heidi K., Keiter, Eric R., Hoekstra, Robert J., Day, David M., Boman, Erik G. (2009). A New Parallel Strategy for Transistor-Level Circuit Simulation, submitted to the Design Automation Conference (DAC).
2. Çatalyürek, Ü., and Aykanat, C. (1999). Hypergraph-Partitioning-Based Decomposition for Parallel Sparse-Matrix Vector Multiplication, *IEEE Trans. Parallel Dist. Systems*, Vol. 10, No.7, pp. 673-693.
3. Devine, K. D., Boman, E. G., Heaphy, R. T., Bisseling, R. H., Çatalyürek, Ü. (2006). Parallel Hypergraph Partitioning for Scientific Computing, *IEEE Proceedings of the 20th International Parallel and Distributed Processing Symposium (IPDPS'06)*.

*This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.*

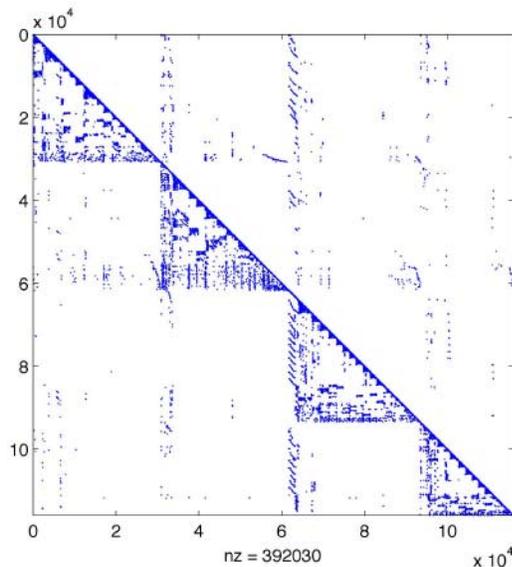


Figure 3. Hypergraph partitioning of BTF structure for CMOS memory circuit on four processors.

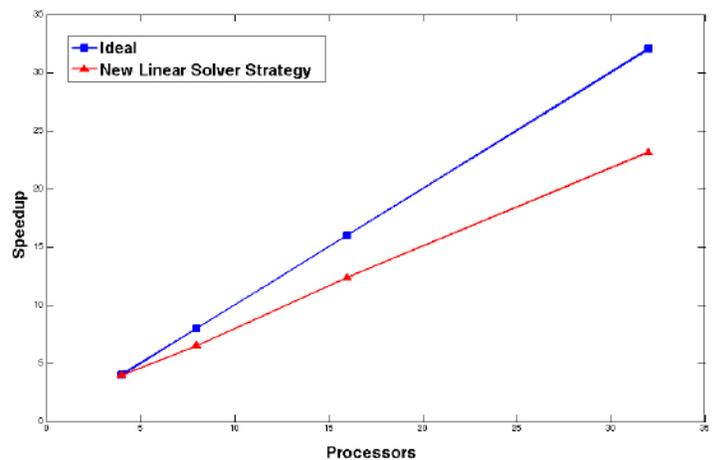


Figure 4. Linear solver speedup for CMOS memory circuit on 4, 8, 16, 32 processors.

# PREDICTING DEFECT PHYSICS FOR RADIATION DAMAGE IN SEMICONDUCTOR DEVICES

Contact: Peter A. Schultz, Multiscale Dynamic Materials Modeling, [paschul@sandia.gov](mailto:paschul@sandia.gov)

*We are developing SeqQuest, a fast, compact, parallelized, first principles density functional theory (DFT) code, as a powerful computational tool for integrating into multiscale efforts.*

*SeqQuest has been an important contributor to multiple Sandia missions.*

Developing computational tools to enable multiscale modeling bridging between physical scales in engineering and materials systems has been a cornerstone of efforts within the CCIM to achieve predictive science-based materials simulations. When phenomena at the macroscopic scale are determined by processes occurring at the atomic scale, accurate and efficient quantum mechanics methods incorporating the correct physics are the foundation upon which an effective multiscale simulation hierarchy is based. We are developing SeqQuest, a fast, compact, parallelized, first-principles density functional theory (DFT) code, as a powerful computational tool for integrating into multiscale efforts. SeqQuest has been an important contributor to multiple Sandia missions.

With the retirement of the Sandia Pulsed Reactor, the Qualification Alternatives for the Sandia Pulsed Reactor (QASPR) project at Sandia was chartered with the challenge of predicting response of neutron-irradiated electrical systems. Electrical response of irradiated devices are ultimately determined by processes occurring at the atomic scale; by the capture and emission of carriers by an evolving chemistry of atomic defects initially created during the neutron pulse. To this end, we developed a new, robust theory for computing energies of charge transitions of asymptotically isolated defects, incorporated this theory into the DFT code SeqQuest, and, using improved parallel implementations, applied it to a wide sampling of defects in silicon (>15) and in gallium arsenide (8). The Si calculations predicted new defect states crucial to accurate simulations of radiation damage. The GaAs calculations are providing important early guidance to QASPR planning going forward concerning which defects are electrically active and which defects are likely to be mobile and drive the evolving defect chemistry.

Prediction of a new defect level for the phosphorus-vacancy pair (P-v) defect in silicon proved timely for QASPR continuum device simulations. The Si defect validation study using DFT/SeqQuest established a mean absolute average error in predicted defect energy levels (electronic capture/emission energies) of 0.1 eV, and a maximum error of ~0.2 eV, for the first time demonstrating that a DFT method can attain predictive accuracy for a broad range

of defect levels. We predicted the existence of previously unsuspected charge transitions in the phosphorus-vacancy pair defect (0/+) and in the boron-vacancy pair (-/0). In device simulations, the P-v pair emerged as a crucial defect impacting the performance (gain) in an irradiated transistor, and identification of the new P-v mid-gap state was important for reliable simulations. This pure prediction of a new (0/+) defect level in P-v, enabled by demonstrating sufficient validated accuracy, was subsequently confirmed by experiment. The predictive accuracy achieved in these defect simulations vindicates a science-based approach in QASPR's goal of qualifying nuclear weapons.

Extending science-based modeling in QASPR beyond silicon-based devices, to GaAs-based III-V hetero junction bipolar transistors (HBTs), is additionally challenging because only minimal experimental characterization of GaAs defects exists and information about the defect chemistry required for device simulations can only be provided by theory. Following the protocols used for Si, our DFT/SeqQuest calculations determined that almost all the primary defects in GaAs are electrically active, many possessing mid-gap states potentially impacting device performance, but that only one defect, the interstitial-As (an extra As in the lattice) was likely to be mobile and therefore be important in the evolving defect chemistry and transient electrical response. The calculations reproduced detailed properties of the EL2 defect in GaAs, here unambiguously confirmed to be the anti-site As (As replacing Ga in the lattice). Validated by this result and building on the success in silicon, these results may justify the use of simplifying assumptions in the device simulations, making the challenge of modeling HBT devices in QASPR much less daunting.

The defect level calculations were performed with SeqQuest (<http://dft.sandia.gov/Quest>). Conventional DFT calculations for defect level calculations run afoul of two connected problems. First, the "band gap problem" causes the fundamental gap in semiconductors (the energy scale for defect levels) to be severely underestimated, and second, DFT codes are designed as "supercell" calculations: an infinitely replicated periodic array of defects rather than an isolated

defect. Together, these two issues, along with the very expensive computational cost of the many large-scale DFT calculations needed, have historically precluded predictive simulations of charged defect energies in semiconductors. We developed and implemented in SeqQuest a Finite Defect Supercell Model (FDSM). The FDSM removes the infinities created by a periodically replicated charge in the supercell, provides a more rigorous treatment of the Coulomb boundary conditions, and removes other artifacts of the supercell approximation. The FDSM, with its rigorous boundary conditions, was implemented in SeqQuest, and with the supercell approximation errors properly removed, the defect level calculations yield an agreement with experiment to within an average 0.1 eV, unencumbered by a band gap problem. The FDSM capability is unique to SeqQuest.

An additional challenge of conventional DFT methods for defect calculations is the high computational cost. The smallest systems giving reasonably converged results for defect calculations entail a few hundred atoms, and to survey computationally the large number of different defects required to assemble a credible validation test suite requires hundreds and even thousands of these calculations. The Si study above required more than 1000 separate, fully energy-optimized structures with 250 atoms, a scale of computing not practical with conventional DFT methods. SeqQuest uses optimized local-orbital basis sets, highly tuned algorithms that scale linearly with number of atoms for the construction of electron interactions, and a steadily improving parallel implementation. Calculations of 250 atoms and even 500 atoms are routine with even rather modest parallel computational resources, in contrast to plane-wave-based methods conventionally used, without compromising any meaningful accuracy.

### **References**

Schultz, P. A. (2008). Theory of Defect Levels and the ‘Band Gap Problem’ in Silicon, *Physical Review Letters*, Vol. **96**, Article No. 246401.

Schultz, P. A. (2008). Schultz replies (to Comment on “Theory of Defect Levels and the ‘Band Gap Problem’ in Silicon, *Physical Review Letters*, Vol. **103**, Article No. 089702.

***This work has been supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program and Hostile Environment Campaign.***

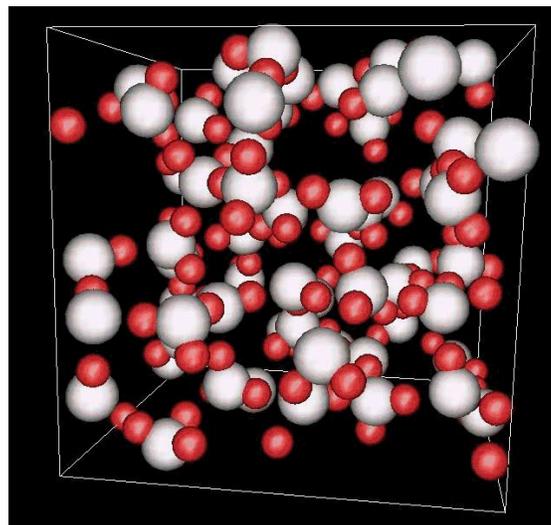
# ENABLING QUICK AND FAST PREDICTIVE MULTI-SCALE MODELING AND SIMULATIONS

Contact: Ann E. Mattsson, Multiscale Dynamics Materials Modeling, [aematts@sandia.gov](mailto:aematts@sandia.gov)

*...AM05 is emerging as the functional of choice for accuracy, as well as speed, for essentially all materials*

Quantitatively accurate results for a broad range of systems, in combination with a relatively low computational cost, have made density functional theory (DFT) the foundation of most large-scale quantum mechanical simulations in science. Thus, DFT is also the workhorse method for the high-fidelity quantum calculations that are frequently relied on as the basis for predictive material multi-scale simulations (Figure 1). DFT simulations help solve some of the most difficult materials problems: describing radiation effects in semiconductor materials, as is done in the Sandia QASPR project, understanding materials under the extreme conditions encountered in high-energy-density physics experiments on Sandia's "Z" pulsed power machine, or revealing how ion selection occurs in biological systems.

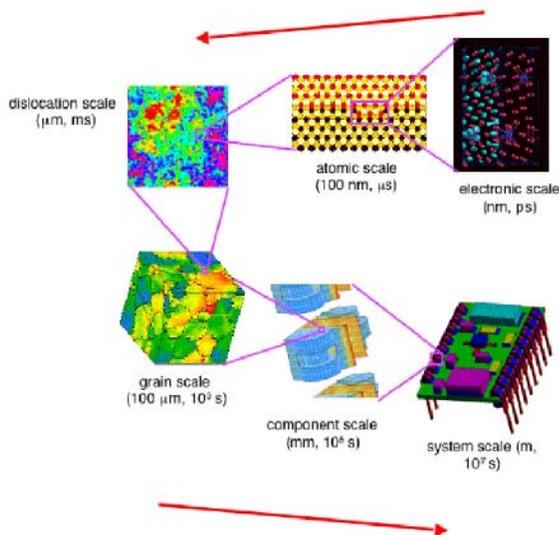
At the core of every DFT calculation lies the exchange-correlation (XC) functional, which sets the limit for the accuracy of the calculations. Theory tells us that a "divine" XC functional exists, with which DFT calculations provide the correct ground state results. Since this functional is not known, many approximate functionals have been developed. A remaining issue, however, has been the lack of systematic improvement in accuracy of functionals, in



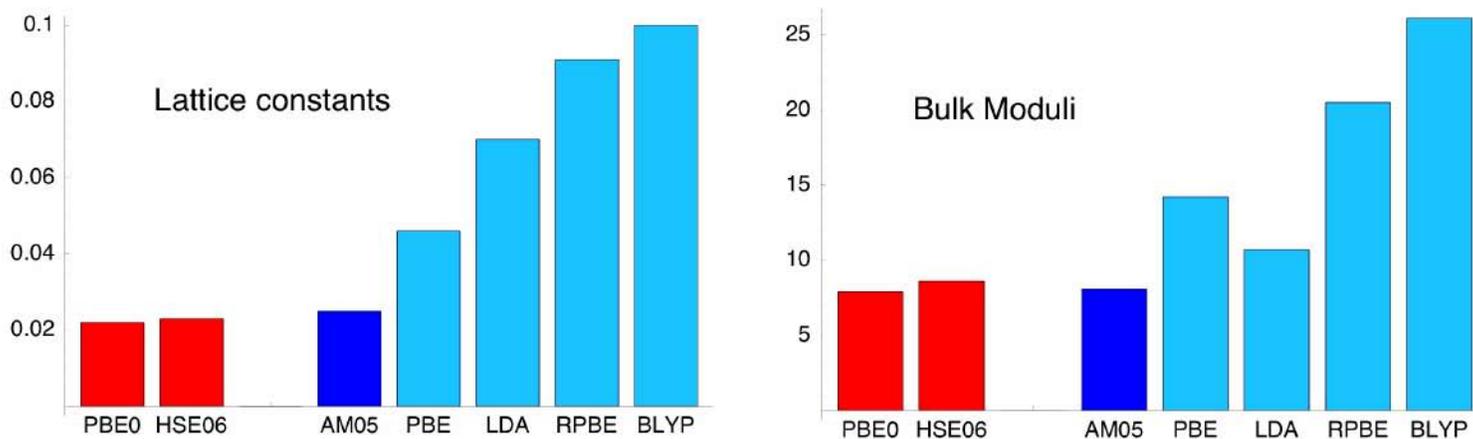
**Figure 2.** Obtaining an accurate description of water dynamics requires very long (10s of ps) molecular dynamics simulations on large supercells. This requires an XC functional that is both fast and accurate. The picture is a snapshot of a DFT-MD simulation using AM05 for 64 water molecules.

particular for solid-state systems; it has not been possible to predict which functional is most accurate for a given system. The development of new functionals is therefore of basic importance to the progress of not only computational materials science, but also that of physics and chemistry in a broad sense. Since DFT is increasingly being employed for large systems (several hundred atoms) and for long *ab initio* molecular dynamics (MD) simulations (tens of ps) (Figure 2), the trade-off between speed and accuracy is arising as an additional major concern in functional development.

There is a fundamental difference in the behavior of electronic wave functions in a bulk solid and in regions with greatly reduced electron density (surfaces, vacancies in metals, and interstitial regions in semiconductors). In the former, the wave functions are periodic propagating waves, while in the latter they are damped evanescent surface waves. By first identifying and then focusing on this duality, we succeeded in developing a functional with high accuracy while retaining low complexity and, thus, high speed. For solids, the 2005 Armiento and Mattsson (AM05) XC



**Figure 1.** DFT calculations on the electronic scale are the underpinning of many predictive multiscale efforts at Sandia.



**Figure 3.** Comparison of mean absolute errors (MAE) for properties of 20 solids calculated with seven different functionals. The left panel shows lattice constants MAE in Ångström ( $10^{-10}$  m) and the right panel shows bulk moduli MAE in GPa. GGA type functionals (blue) are one to three orders of magnitude faster to use than hybrids (red). AM05 has the same accuracy as hybrids for solids and thus enables accurate and fast DFT calculations of, for example, defects in semi-conductors (QASPR). It also allows for the use of DFT-MD as an accurate tool in Equation of State construction (HEDP).

functional is proving to be as accurate as the best available XC functionals (so called hybrid functionals, Figure 3), while allowing one to three orders of magnitude faster calculations; a calculation of a defect in a semiconductor like GaAs that runs overnight (12 hours) using AM05 would take almost six weeks using the fastest of the accurate hybrid functionals (HSE03/HSE06). Hence, AM05 is emerging as the functional of choice for accuracy, as well as speed, for essentially all materials.

Even so, there is more work to be done. Neither AM05 nor any other existing functional can treat dispersive forces or van der Waals bonding, such as are prevalent in two important classes of materials: biomolecules and molecular solids, the latter of which includes high explosives and energetic materials. Nor can existing functionals reliably predict properties of materials containing f-electron elements (lanthanides and actinides). Functional development therefore continues at Sandia to systematically extend AM05 to accurately treat van der Waals and f-electron materials. One focus is on efforts in rational compound design, where it is of utmost importance to be able to obtain the right trends so that a target property can be calculated as a function of chemical composition.

### References

- <http://dft.sandia.gov/functionals/AM05.html>
- Armiento, R., Mattsson, Ann E. (2005). Functional Designed to Include Surface Effects in Self-Consistent Density Functional Theory, *Physical Review B*, Vol. **72**, No. 8, 5108.
- Mattsson, A. E. (2002). In Pursuit of the “Divine” Functional, *Science*, Vol. **298**, p. 759.
- Mattsson, A. E., Armiento, R., Paier, J., Kresse, G., Wills, J. M., Mattsson, T. R. (2008). The AM05 Density Functional Applied to Solids, *Journal of Chemical Physics*, Vol. **128**, 084714.
- Mattsson, A. E., Schultz, P. A., Desjarlais, M. P., Mattsson, T. R., Leung, K. (2005). Designing Meaningful Density Functional Theory Calculations in Materials Science—a Primer, *Modelling Simul. Mater. Sci. Eng.*, Vol. **13**, R1-R31.

***This work’s original support was from SNL’s Laboratory Directed Research and Development (LDRD) Program. It is currently supported by the National Nuclear Security Administration’s (NNSA) Advanced Simulation and Computing (ASC) Program.***

# UNCERTAINTY QUANTIFICATION METHODOLOGY FOR DAMAGED SEMICONDUCTOR DEVICES

Contacts: Joseph Castro, Elect'I & Microsystem Modeling, [jpcastr@sandia.gov](mailto:jpcastr@sandia.gov)  
 Robert Hoekstra, Elect'I & Microsystem Modeling, [rjhoeks@sandia.gov](mailto:rjhoeks@sandia.gov)  
 Brian Rutherford, Ind Surv Assessment & Statistics, [bmruthe@sandia.gov](mailto:bmruthe@sandia.gov)  
 Vicente Romero, Validation & Uncertainty Quant, [vjromer@sandia.gov](mailto:vjromer@sandia.gov)

*Since modeling predictions have never been used as direct evidence for nuclear weapons electrical systems qualification for hostile environments, we must establish the confidence necessary to trust these models to support warhead qualification.*

To be deemed qualified, nuclear weapons systems must meet military requirements including survival in radiation environments. Fast burst reactor (FBR) testing of the electrical systems within these weapons has been the preferred technique for qualification to short pulse neutron environment requirements. The SPR (Sandia Pulse Reactor)-III reactor at Sandia National Laboratories was used historically for the purposes of qualification. In a move to improve operations at Sandia, SPR-III was shut down and dismantled. Since weapons systems still must be demonstrated to meet short pulse neutron requirements for qualification, the QASPR (Qualification Alternatives to the Sandia Pulse Reactor) project was initiated to compensate for the loss of SPR-III. To fulfill this need, the QASPR approach is to develop a science-based engineering methodology that integrates available experimental capabilities with modeling and simulation and uncertainty quantification to produce a best-estimate plus uncertainty prediction as evidence for qualification.

Since modeling predictions have never been used as direct evidence for nuclear weapons electrical systems qualification for hostile environments, we must establish the confidence necessary to trust these models to support warhead qualification. So, success of the QASPR program relies on not only predictive capability of the models but on providing significant supporting evidence for confidence in those predictions. Validation of the model against a broad range of experimental data including a detailed quantification of the uncertainty and variability is the avenue to this goal. Only then can the prediction be used in the quantification of margins for qualification. Three key difficulties arise in this process: lack of detailed knowledge, which may result in large uncertainties, accounting for systematic biases, and characterizing and quantifying observed experimental variability and uncertainty. A UQ Methodology (Figure 1, bottom) was developed using calibration, uncertainty characterization,

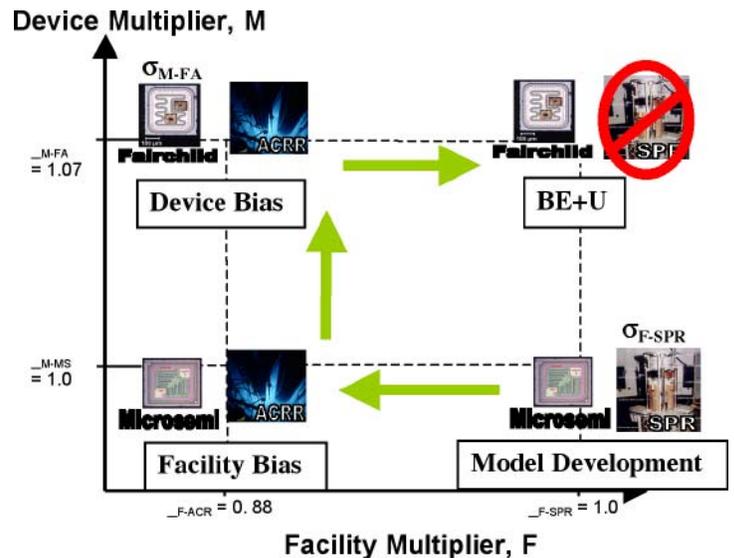
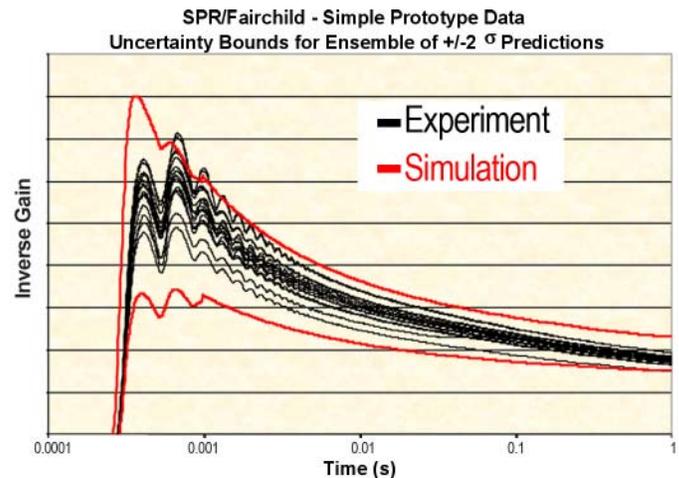


Figure 1. Top: Successful results of blind test comparing ensemble of experimental data with best-estimate plus uncertainty prediction generated with UQ Methodology. Bottom: UQ methodology developed for simple prototype.

design of experiments, experimental testing, and statistics to overcome these difficulties and generate a best-estimate plus uncertainty prediction of an electrical device's response to the SPR-III environment.

The paradigm that was used to process uncertainty sources into a form that supported the model validation comparisons in this work is outlined in [1] and further development of the principles and procedures are explained in [2]. A series of QASPR prototypes or "proof of concept" exercises were developed to exercise these newly developed methodologies. The first of these prototypes is the Simple (Device) Prototype with a Silicon device, which has now been successfully completed (Figure 1, top) and represents an important step toward providing that confidence. A Sandia SAND report is in process describing the details of the UQ methodology [3].

### **References**

1. Romero, V. J. (2007). *A Paradigm of Model Validation and Validated Models for Best-Estimate-Plus-Uncertainty Predictions in Systems Engineering*, Paper 200-01-1746 for Soc. Automotive Engineers 2007 World Congress, April 16–20, 2007, Detroit, Michigan.
2. Romero, V. J. (2009). *Application of a Pragmatic Interval-Based "Real Space" Approach to Fire-Model Validation Involving Aleatory and Epistemic Uncertainty*, Paper AIAA-2009-2279, 11th AIAA Non-Deterministic Approaches Conference, 4-7 May 2009, Palm Springs, CA.
3. Rutherford, B., Romero, V., Castro, J., Hoekstra, R. (2009). The Simple Prototype UQ Methodology, in preparation.

***This work is funded by the National Nuclear Security Administration's (NNSA) via C7 (Nuclear Survivability), Advanced Simulation and Computing (ASC) Program ASC, and RTBF (Readiness in Technical Base and Facilities).***

# ADDRESSING AGING PROBLEMS THROUGH ENHANCED LOW DOSE-RATE SENSITIVITY (ELDRS) AND OTHER RADIATION EFFECTS CALCULATIONS

Contact: Harold P. Hjalmarson, Multiscale Computational Materials Methods, [hphjalm@sandia.gov](mailto:hphjalm@sandia.gov)

*The [ELDRS] problem affects several types of bipolar transistors...used in low-level radiation environments such as satellites in space or nuclear weapons in storage.*

*It is regarded as a transistor problem... .*

The enhanced low dose-rate sensitivity (ELDRS) problem has been the focus of much experimental and theoretical work at several laboratories since its discovery in 1992 [1]. Devices that exhibit this problem have enhanced radiation sensitivity to a given dose of ionizing radiation if it is delivered at a very low dose rate compared with the rates used for radiation tolerance testing. This problem affects several types of bipolar transistors. It is a concern for transistors used in low level radiation environments such as satellites in space or nuclear weapons in storage. It is regarded as a transistor problem because the problematic radiation environment is a normal operating environment.

Recently, we published a theoretical explanation for ELDRS in terms of radiation-induced release of hydrogen in the silicon dioxide portions of these transistors [2]. Upon release, this hydrogen migrates to the interface between the silicon dioxide and the underlying silicon. There it interacts with hydrogen-passivated silicon dangling bonds. This reaction recreates the bare dangling bond that then tends to recombine electrons and holes during the operation of the transistor. This degrades the transistor performance by reducing the gain.

Figure 1 compares the calculations with data for interface trap density as a function of dose rate. These data show that more interface traps are created at low dose rates compared with high dose rates. These data, which are representative of ELDRS data, are difficult to acquire because the irradiation times are very long. Thus, ELDRS testing becomes time consuming and expensive.

The calculations shown in the figure are in good agreement with the data. Each point represents a calculation taken to a time of approximately 30 years. In these calculations, the radiation releases hydrogen in an indirect way. First holes are trapped in the oxide. Then ever-present molecular hydrogen is “cracked” at these sites, thereby releasing atomic hydrogen. This process is very efficient at low dose rates. However, at high dose rates, the holes are neutralized by recombination with radiation-injected

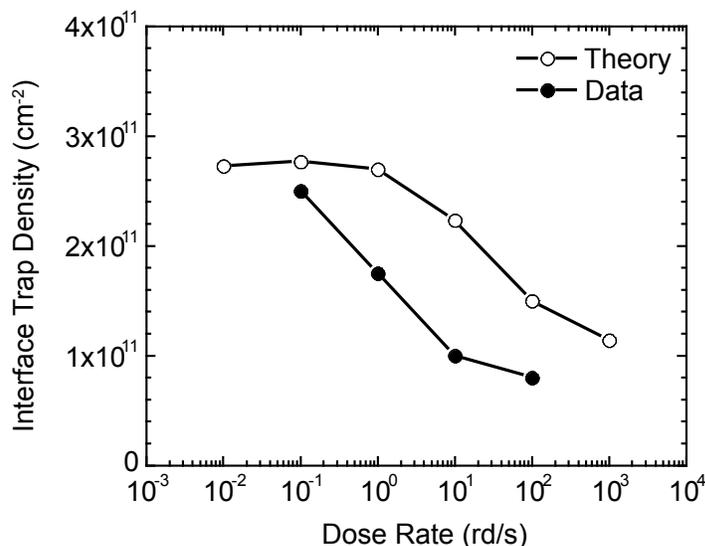


Figure 1. Shows the interface trap density as a function of dose rate.

electrons before they are able to crack molecular hydrogen. This bimolecular process dominates at high dose rates because in this case both the electron and hole populations become large.

The theory predicts that the interface trap density will depend on molecular hydrogen concentration. The calculated dependence is consistent with experiments in which this concentration is varied by using an environment of molecular hydrogen whose partial pressure is controlled.

Similar calculations are also being performed to understand the electrical effects caused by high dose rates of hot electrons for short durations of time. These continuum calculations consider the effects of subsequent hot electron-hole plasma. A set of atomistic molecular dynamic (MD) calculations is being done to understand atomistic effects of the hot electrons and holes. A set of time-dependent density-functional theory (TDDFT) calculations is being planned to obtain additional microscopic information for the continuum calculations.

Finally, we are using continuum calculations of reactive transport to understand aging effects in ultra-thin gate oxides. Such oxides use novel high-k materials to shrink device dimensions in future microelectronic technology. These new calculations are also focused on the electrical effects of interface traps.

### **References**

1. Enlow, E. W., Pease, R. L., Combs, W. E., Schrimpf, R. D., Nowlin, R. N. (1991). Response of Advanced Bipolar Processes to Ionizing Radiation, *IEEE Trans. Nucl. Sci.*, Vol. **38**, No. 6, pp. 1342-1351.
2. Hjalmarson, H. P., Pease, R. L., Devine, A. B. (2008). Calculations of Radiation Dose-Rate Sensitivity of Bipolar Transistors, *IEEE Trans. Nucl. Sci.*, Vol. **55**, No. 6, pp. 3009.

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

# CIRCUIT BASED FRAMEWORK IMPROVES MODELING AND SIMULATION OF LATENT *MYCOBACTERIUM TUBERCULOSIS*

Contact: Elebeoba E. May, Discrete Math and Complex Systems, [eemay@sandia.gov](mailto:eemay@sandia.gov)

*To combat the growing threat of latent tuberculosis to global health security, a path connecting the gap between fundamental research and rational drug design for effective treatment of latent TB is needed. As in other science and engineering fields, modeling and simulation is serving as a transformative bridge in understanding the multiscale phenomena of latency and reactivation in tuberculosis.*

According to the 2008 World Health Organization's (WHO) report, tuberculosis (TB), caused by the bacterium *Mycobacterium tuberculosis* (Mtb), continues to be a major international cause of illness and death worldwide. WHO reports indicate that in 2006 there were 9.2 million new cases of TB and 1.7 million deaths due to TB. Mtb is able to persist in host tissues in latent state. During latency, Mtb is present in the host but does not produce any overt symptoms; this presents a challenge in the treatment of latent TB. With an estimated one-third of the world's population carrying latent TB, reactivation of this highly contagious disease is of great concern particularly in individuals with weakened immune systems. With nearly four thousand genes and over nine hundred reactions in its reconstructed metabolic network, identifying which combination of genes and biochemical pathways constitute the Achilles heel of Mtb is not a trivial task. By coupling advances in high-throughput transcriptomics and metabolomics with large-scale computing, simulation, analysis, and optimization tools, Sandia scientists are meeting this challenge.

To combat the growing threat of latent tuberculosis to global health security, a path connecting the gap between fundamental research and rational drug design for effective treatment of latent TB is needed. As in other science and engineering fields, modeling and simulation is serving as a transformative bridge in understanding the multiscale phenomena of latency and reactivation in tuberculosis. Through a five-year grant award from the National Institutes of Health's National Heart, Lung, and Blood Institute, researchers at Sandia are partnering with researchers in the University of New Mexico Health Sciences Center for Infectious Disease and Immunity and Los Alamos National Laboratory's Theoretical Biology and Biophysics Division to develop models that enable a quantitative understanding of the genetic basis of latency and reactivation in a murine model of tuberculosis. The larger goal of this collaborative effort is to leverage BioXyce, developed using CCIM's large-scale circuit simulation tool Xcyce™ (<http://www.cs.sandia.gov/xyce/>), to understand the response

of Mtb within the microenvironment of a granuloma. A granuloma is an aggregation of host immune cells that function to cooperatively quarantine but not completely eliminate the mycobacterium. Understanding the genetic and biochemical mechanisms that Mtb employs to persist within the hostile environment of the granuloma will help identify viable chemotherapies to treat the extremely large number of individuals with latent tuberculosis.

With BioXyce, we can simulate large control networks consisting of entire cells, homogeneous cell cultures, or heterogenous interacting host-pathogen systems in order to understand the dynamics and stability of such systems. The input parameters for BioXyce, collected from literature and databases like BioCyc, KEGG, and BRENDA, are optimized using the DAKOTA (Design Analysis Kit for Optimization and Terascale Applications) uncertainty quantification toolkit. In collaboration with researchers in the Department of Electrical and Computer Engineering, University of British Columbia-Vancouver, we are augmenting the BioXyce/DAKOTA framework using computational reachability techniques to set initial value conditions and provide tighter parameter bounds for our model. This results in networks able to replicate behaviors more consistent with experimental data.

In a deviation from the norm with regard to the empirical sciences, modeling and simulation plays a preeminent role in this work. Using the Wayne model of hypoxia, where oxygen is slowly depleted from the environment of actively growing *M. tuberculosis*, scientists are generating high time-resolution dynamic profiles of the system in order to produce a high-fidelity quantitative model able to replicate observed behavior and perform novel predictions of Mtb response to varying environmental conditions. Current quantitative and empirical work (carried out in the Lyons Lab, UNM-HSC) focuses on *in vitro* studies of Mtb in a hypoxic microenvironment, where the tubercle bacilli can circumvent the shortage of oxygen by developing alternative energy generation mechanisms by way of the glyoxylate bypass pathway (Figure 1, pathway recreated from BioCyc,

*www.biocyce.org*). BioXyce models of this important pathway have been constructed, and simulations in the absence and presence of various inhibitory molecules have been conducted (Figure 1).

A large-scale circuit simulation framework enables the parallel probing of viable targets for small-molecule therapeutics. Rather than probing the singular effect of a small molecule on a target protein, researchers are able to probe the ripple effect a single molecule or group of small-molecules has on the entire system. In collaboration with UNM-HSC’s Biocomputing Center this approach, coined “systems chemical biology,” is being used to identify small molecules that directly or indirectly interfere with latency-related pathways (Figure 1). As advances in biotechnology propel science deeper into the –omics age, large-scale simulation is quickly becoming the enabling link that transforms biological data into scientific discoveries in the medical, environmental, and energy sciences.

**References**

May, E., Leitao, A., Faulon, J-L., Joo, J., Misra, M., Oprea, T. (2008). Understanding Virulence Mechanisms in *M. Tuberculosis* Infection Via a Circuit-Based Simulation Framework, *IEEE Engineering in Medicine and Biology Society International Conference*.

May, E., Schiek, R. (2009). BioXyce: An Engineering Platform for the Study of Cellular Systems, *IET Systems Biology Journal*. In press.

***This work is supported by the National Institutes of Health/National Heart Lung Blood Institute grant #5K25HL75105-3.***

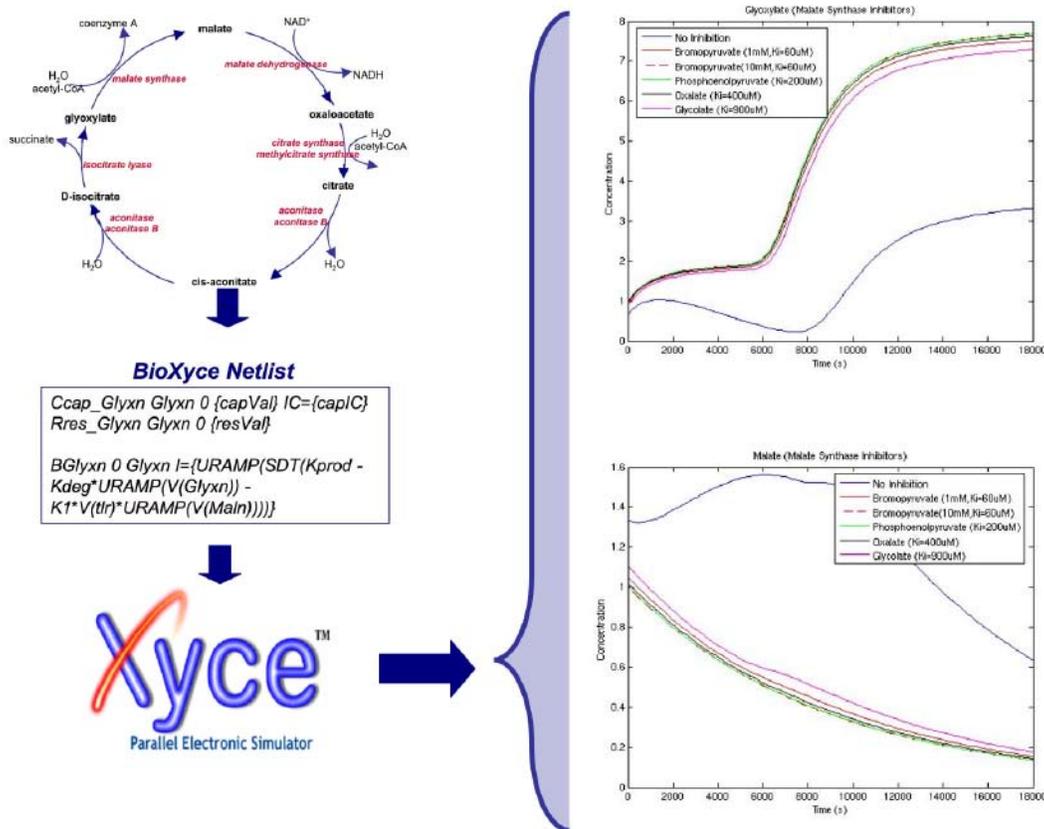


Figure 1. Simulation of the glyoxylate pathway in the presence and absence of inhibitory molecules.

# DEVELOPING TWO-FLUID MODELS FOR SIMULATING SINGLET OXYGEN GENERATOR REACTORS

Contacts: Lawrence C. Musson, Electrical and Microsystems Modeling, [lcmusso@sandia.gov](mailto:lcmusso@sandia.gov)  
 Roger P. Pawlowski, Applied Math and Applications, [rppawlo@sandia.gov](mailto:rppawlo@sandia.gov)  
 Andrew G. Salinger, Applied Math and Applications, [agsalin@sandia.gov](mailto:agsalin@sandia.gov)

*Designing SOG reactors by building and testing alone is prohibitively expensive. By applying computational simulation, designers can explore an expanded number of reactor configurations at a vastly accelerated pace.*

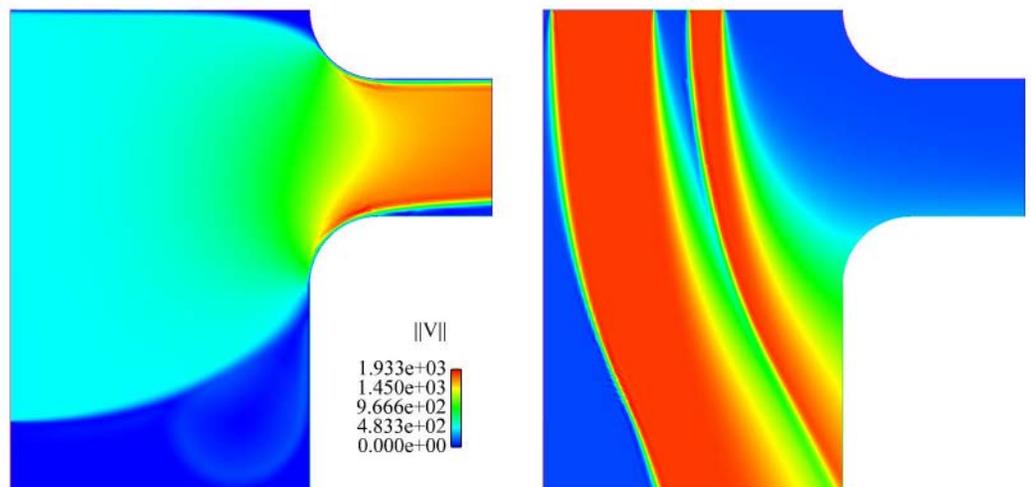
Leveraging over 15 years of investment in chemically reacting flow simulation, Sandia is impacting a new generation of national security and industrial technologies by optimizing the design of chemical oxygen iodine lasers (COIL). COILs were invented in the 1970s by the U. S. Air Force. Recently, they are being further examined for applications in missile defense and industrial cutting and drilling. Critical to this technology is the development of singlet oxygen generators (SOG) to maximize the power of the laser while minimizing the size and weight of the overall system. Designing SOG reactors by building and testing alone is prohibitively expensive. By applying computational simulation, designers can explore an expanded number of reactor configurations at a vastly accelerated pace. Sandia is leading the way by developing new models and software for the simulation of these multiphase reacting flow systems on massively parallel super computers.

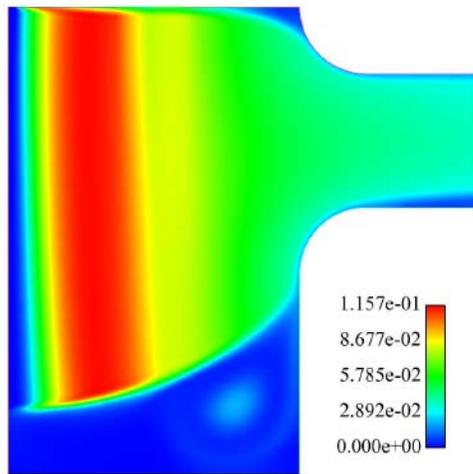
The SOG is a low-pressure, multiphase flow chemical reactor that is used to produce molecular oxygen in an electronically excited state (i.e., singlet delta oxygen). The primary product of the reactor, the energetic oxygen, can be used to dissociate and energize iodine. The gas mixture including the iodine is accelerated to a supersonic speed and lased. Thus, the SOG is the fuel generator for the COIL.

The SOG appears in various configurations, but the one in focus here is a crossflow SOG. A gas consisting of molecular chlorine and a diluent, usually helium, is pumped through a roughly rectangular channel. An aqueous solution of hydrogen peroxide and potassium hydroxide is pumped through small holes into the channel, perpendicular to the direction of the gas flow, causing the solution to become aerosolized. In the liquid droplets, dissociation of the potassium hydroxide draws a proton from the hydrogen peroxide to generate an  $\text{HO}_2^-$  radical. When chlorine diffuses from the gas phase into the liquid, it reacts with the  $\text{HO}_2^-$  ion to produce the singlet delta oxygen; some of the oxygen diffuses back into the gas phase for use in the next stage of the COIL.

The model that was chosen is the Eulerian-Eulerian form of the multiphase flow, isothermal Navier-Stokes equations wherein one set of the equations represents the gas phase and another equation set represents the liquid phase. As flows in the SOG are dominated by advection, a stabilized Galerkin finite element formulation is employed to solve the partial differential equations. The set of equations is large, even for the 2D models studied here. Thus, we are taking advantage of, and developing, algorithms to harness the power of large parallel computing architectures to solve the

**Figure 1.** Discussed on the following page, these figures show contours of the gas speed and the number density of droplets in the SOG.





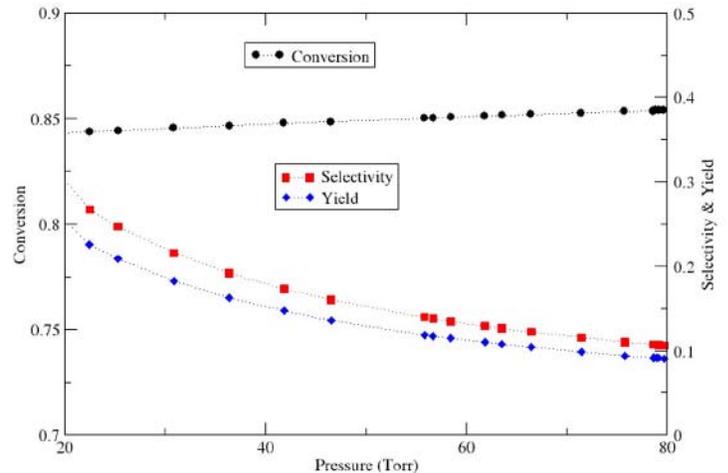
**Figure 2.** Shaded contours of the singlet delta oxygen concentration imply the spatial variation of the production of singlet delta in the reactor.

steady-state form of these equations rapidly, which enables the exploration of the large parameter space of the equations via continuation methods. For a complete description of the equations and solution methods, see Shadid [1].

The gas phase in the singlet oxygen generator is the continuous phase. In this model, the gas flow is assumed to be steady and incompressible. The appropriate continuum equation set for such a system is the steady form of the Navier-Stokes equations consisting of continuity and momentum conservation. Singlet delta oxygen is produced in the reactor by chemical conversion, and so the composition of the gas phase must also be modeled by solving mass transfer equations. In this work, the source terms that appear in the mass transfer equation arise from chemical reactions. This model follows the one proposed by Thayer [2] but has been generalized to accommodate a droplet size distribution.

The second phase is one of droplets dispersed within the gas. In the SOG, the equations model not only the spatial variation of number density, composition, and momentum of droplets, but also model the distribution of each of these quantities as a function of droplet size. A complete description of the dispersed phase flow equations is not possible in this space, but is being prepared.

The crossflow singlet oxygen generator is so named because the working fluids travel largely perpendicular to one another. In the SOG configuration in this work, the gas, a mixture of He and Cl<sub>2</sub>, enters from the left and exits to the right with oxygen as an additional component. The BHP is sprayed in from the top in two stages as seen in Figure 1 (see previous page). Contours of the speed of the gas and the



**Figure 3.** Performance of the SOG is reported by the three metrics of conversion, selectivity, and yield.

number density of the liquid show the flow field of each in the reactor. Figure 2 shows the mass fraction of the excited oxygen. The fraction reaches its maximum in the region of the reactor where the gas and droplets mingle. Downstream of this region, the excited oxygen quenches rapidly to its ground state.

SOG performance is measured by three metrics: conversion, selectivity, and yield. Figure 3 shows a typical set of these metrics for the SOG shown in Figures 1 and 2 and how that performance changes as a function of the reactor pressure. Higher pressure in the reactor increases the total conversion of chlorine, but the accompanying higher gas density accelerates the loss path of singlet delta oxygen through quenching.

### References

- Shadid, J., Salinger, A. G., Pawlowski, R. P., Lin, P. T., Hennigan, G. L., Tuminaro, R., and Lehouq, R. B. (2006). Large-scale Stabilized FE Computational Analysis of Nonlinear Steady State Transport / Reaction Systems, *Comp. Meth in App Mechanics and Eng.*, Vol. **195**, pp. 1846-1876.
- Thayer III, William J., Cousins, A. K., Romea, R. D. (1994). Modeling of Uniform Droplet Singlet Oxygen Generators, *Proceedings of the SPIE*, Vol. **2117**, pp. 71-100.

***This work has been supported by the US Air Force Research Laboratory (AFRL).***

# DEVELOPING FLEXIBLE MULTI-PARADIGM MODELING AND SIMULATION ENVIRONMENTS FOR COMPLEX PLANNING

Contact: John D. Sirola, Exploratory Simulation Technologies, [jdsirola@sandia.gov](mailto:jdsirola@sandia.gov)

*EMPaSE (Extensible Multi-Paradigm Simulation Environment) has been applied to network simulations both within Sandia and by external customers. We have leveraged the extensible nature of both the simulation engine and the integrated development environment (IDE) to rapidly prototype an integrated planning tool for external customers.*

This work aims to develop a flexible common modeling and simulation environment for creating and executing models within sequential modular (SM), system dynamics (SD), agent-based (AB), or discrete event simulation (DES) modeling paradigms, as well as hybrid models that combine components that embody more than one of these paradigms. These paradigms frequently arise when modeling a diverse range of domains, including planning and scheduling, engineered systems, and socio-technical policy models. However, while more than one of these paradigms may be applicable to a particular problem, modeling tools generally restrict the practitioner to a single modeling paradigm. This restriction can hamper model development, as the paradigm chosen for early phases of development may be inappropriate for — or even prevent — subsequent model enhancements. A better approach would be to create a single *hybrid* model that contains components that leverage different modeling paradigms.

To address this need, we are developing an Extensible Multi-Paradigm Simulation Environment (EMPaSE) (Figure 1). This environment supports not only single-paradigm modeling, but also intrusive hybridization of SM, SD, AB, and DES paradigms. Our work has centered on developing two separate, but complementary, tools: a graphical multi-paradigm integrated development environment (MPIDE) and a high-efficiency multi-paradigm simulation engine (MPSE). This work arose from an effort to develop a general-purpose modeling environment for agent-based simulations. When designing agent-based systems, we encountered a fundamental duality within individual agents: while most agent-based simulation environments rely on a messaging- or event-based framework, the internal logic dictating an agent's behavioral response was more likely to be defined causally or procedurally. In order to allow users to interactively create and define new behaviors, the modeling and simulation environment would need to be able to support both paradigms. While environments exist that allow for high-level hybridization [1], there are no environments that support the intrusive hybridization of simulation components we found necessary.

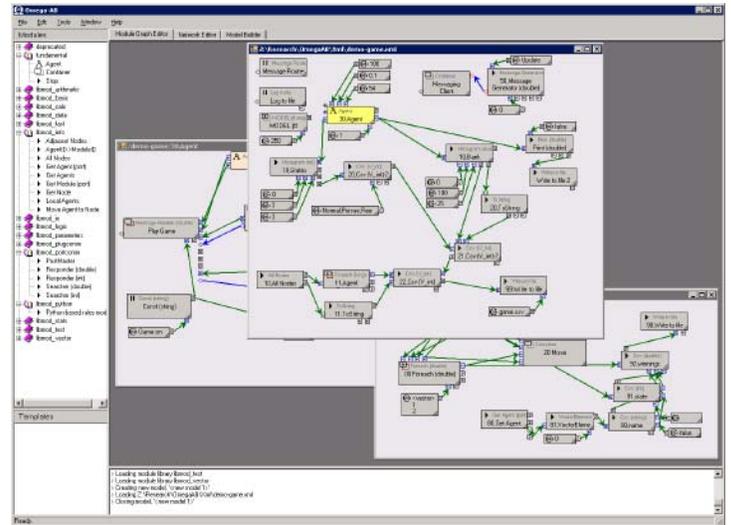


Figure 1. The EMPaSE graphical multi-paradigm integrated development environment (MPIDE) showing the interface for constructing a simple agent-based model.

The fundamental difference between an event-based and a procedural (or equation-based) simulation that prevents ready hybridization lies in the movement of data within the simulations; and specifically, whether the data moves using a “pull” or “push” model. General procedural computation invariably follows a “pull” data transfer model. Each step in the procedure (either equation or control structure) relies solely on data computed in a previous step. When evaluating a step, the simulator will read (pull) the required data in from the referenced variables, perform the computation specified by the step, and store the result back to a variable so that it is available to subsequent steps. The key property of pull-based data transfers is that it is the *consumer* of the data that initiates the data transfer. In contrast, event-driven simulation paradigms — including discrete event and the core agent-based simulation environment — rely on data transfers (communication patterns) that are initiated by the data *producer*. For example, one agent will prepare a message and send (push) it to the intended recipient agent, often triggering an immediate response from the recipient. Similarly, in a

discrete event paradigm, the main event queue will iteratively push the next event in the queue to the designated target, triggering the corresponding event processing function.

Push-based communication paradigms may be emulated using pull-based patterns by having the recipients continuously poll their data sources. However, this approach places a significant burden on the individual recipients and can complicate the verification process for both the environment and the model. Instead, EMPaSE implements two parallel low-level communication systems: a pull-based system (ports) for general computation, and a push-based system (plugs) for messages and events. By providing explicit systems for both pull and push patterns, EMPaSE can natively support both general procedural-based (SM, SD) and event-based (AB, DES) simulation paradigms, as well as intrusive hybridization of these hybrids.

While the distinguishing capability of EMPaSE is the dual-data flow model, it has several other key design features. EMPaSE models are built using a hierarchical modular structure. Modules form the fundamental building blocks of an EMPaSE model and can define anything from a basic mathematical operation to a complex behavioral response model. Modules rely on a “plug-in” system, which allows developers to build independent module libraries that are gathered, linked, and instantiated by EMPaSE at run time. Inter-module communication occurs through two complementary systems: pull-based “ports” for general computation patterns and push-based “plugs” for event and message processing. EMPaSE models organize their computational modules in a hierarchical structure. This allows modelers to visually scope portions of a model and define independent execution drivers (Figure 1). This enables separating conditional components, loops, and event or message handlers from the main execution driver. Additionally, hierarchical scoping facilitates encapsulation of component state logic necessary for constructing agent-based models.

The collection of modules, their organization, and their connectivity through “ports” and “plugs” define the computational structure for an EMPaSE model. In addition, EMPaSE natively supports simulation environments based on an abstract graph of nodes and links. Dedicated modules (Agents and Data Provider modules) act as bridges between the computational structure (module hierarchy) and the simulation environment (network). These modules reside at nodes and relate to their neighbors through directed typed links. To facilitate the construction and visualization of complex, interacting networks with dramatically different structure, EMPaSE provides a system for organizing the nodes into hierarchical trees that describe 2-D “slices” of the overall network (Figure 2).

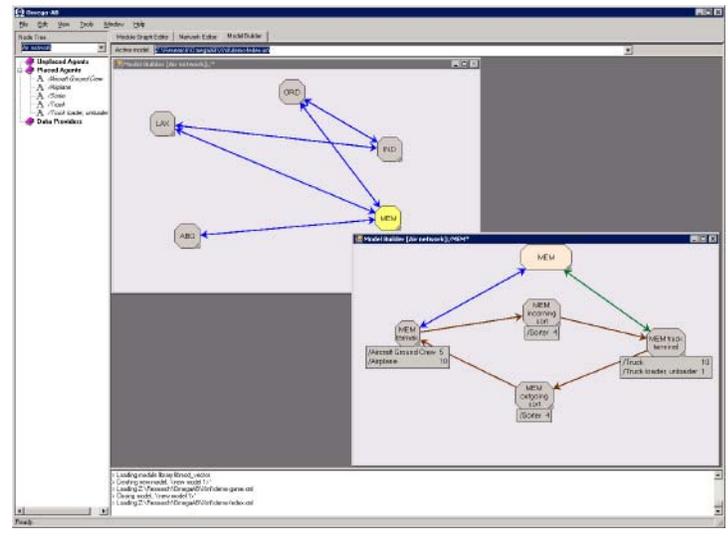


Figure 2. Visualizing a 2-D slice of a model network environment within the MPIDE.

Current research activities are focusing on developing improved capabilities for supporting extensibility within both the core simulation engine and the graphical integrated development environment. Additionally, we are investigating the application of new massively multi-threaded supercomputing architectures as a potential avenue for supporting efficient parallelization of the MPSE for large hybrid discrete-continuous models.

EMPaSE has been applied to network simulations both within Sandia and by external customers. We have leveraged the extensible nature of both the simulation engine and the integrated development environment (IDE) to rapidly prototype an integrated planning tool for external customers. Another internal project has selected EMPaSE as a candidate integration environment for coupling a system dynamics regional resource model with detailed cognitive stakeholder models.

### References

1. Borshehev, A., and Filippov, A. (2004). From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools, *Proceedings of the 22th International Conference of the System Dynamics Society*. Keble College, Oxford. July 25 – 29.

***This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.***

# EVALUATING IMPACTS OF ARCTIC CLIMATE CHANGE ON NATIONAL SECURITY

Contacts: James Strickland, Exploratory Simulation Technologies, [jhstric@sandia.gov](mailto:jhstric@sandia.gov)  
 George Backus, Exploratory Simulation Technologies, [gabacku@sandia.gov](mailto:gabacku@sandia.gov)  
 Mark Boslough, Exploratory Simulation Technologies, [mbboslo@sandia.gov](mailto:mbboslo@sandia.gov)

*Within its national security mission, Sandia National Laboratories is evaluating the impact of climate change on the Arctic as well as impacts that will potentially cascade to other parts of the globe because of those in the Arctic.*

The Arctic region is rapidly changing in a way that will affect the rest of the world. Parts of Alaska, western Canada, and Siberia are currently warming at twice the global rate. This warming trend is accelerating snow and ice loss, permafrost deterioration, coastal erosion, and other phenomena that are a direct consequence of climate change.

Within its national security mission, Sandia is evaluating the impact of climate change on the Arctic as well as impacts that will potentially cascade to other parts of the globe because of those in the Arctic. In this paper, we summarize some of the underlying climate drivers and national security implications associated with the changing Arctic [1 to 5].

Melting of Arctic sea ice has long been identified as one of the strongest signals of climate change. Ice cover is now disappearing. For example, the Arctic Climate Impact Assessment (ACIA) Scientific Report provides a very complete review of the physical impacts of climate change on the Arctic region. The most cautious ACIA model projects a “near-total melting of Arctic sea ice by 2100.” However, taking recent trends into account, there are now estimates that a seasonally ice-free Arctic could happen as early as 2013. By September 2007, according to the data shown in Figure 1, the ice cover had decreased to 4.28 million km<sup>2</sup>, nearly 40% below the long-term average.

Along with sea ice, large areas of permafrost are melting. Permafrost is defined as sediment that has remained below the freezing point of water for two or more consecutive years. The permafrost is concentrated in a geographical band at high latitudes that crosses Siberia, Fennoscandia, Greenland, Canada, and Alaska. At temperatures above the freezing point, carbon dioxide and methane are released into the atmosphere by bacteria and fungi as they process the organic material. In East Siberia, the wind-blown Yedoma permafrost contains about 450 billion tons of easily mobilized carbon. This is about as much carbon as the rest of the Arctic combined, and the same amount of carbon that has been released by all the burning of fossil fuels since the beginning of the industrial revolution. If the Arctic were to warm up enough to decompose only 1% of the permafrost per year, it would

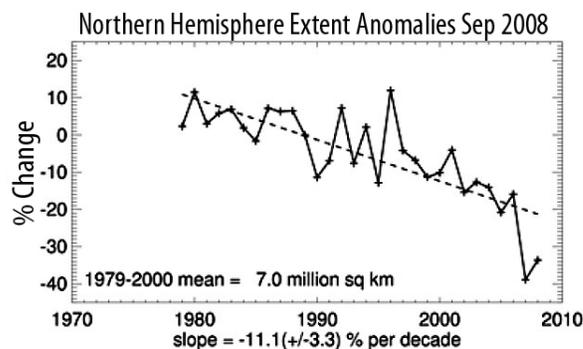


Figure 1. September Arctic sea-ice extent (National Snow and Ice Data Center).

double the carbon releases to the atmosphere from the present rate as a result of human activities (about 9 billion tons/year).

Coastlines that were protected by large expanses of sea ice in the past are now being battered by big waves during the seasonal retreat of the ice. Moreover, patterns of atmospheric circulation are changing, and stronger winds blowing across longer ice-free fetches produce higher waves with more erosive potential. Ice that remains entrained in the waves provides an additional scouring agent. Coastal erosion provides a positive feedback because it accelerates degradation of coastal permafrost that releases methane and carbon dioxide that, in turn, causes global warming.

As part of our work, we developed the climate scenario map shown in Figure 2. The associated scenarios range from “manageable adaptations” to “collapse and chaos.” The most probable scenario is that we are “on the brink.” The rate of climate change associated with this scenario can range from moderate to abrupt, whereas the severity of change falls within the range of that forecast by the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).

For the “on the brink” scenario, the climate is deteriorating toward a tipping point. Temperature increases are dependent on greenhouse gas emissions, the Arctic Ocean becomes ice-free in the summer for longer periods, permafrost deterioration accelerates, coastal erosion becomes more severe, discharge of freshwater from land ice increases

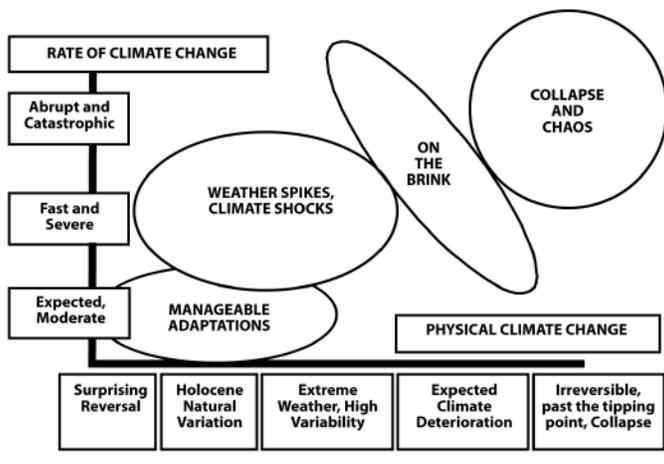


Figure 2. A Map of Arctic climate scenarios.

greatly, large areas of tundra are replaced by shrub land, and forests grow northward. Additionally, fisheries move and many mammal, fish, and bird populations suffer decline or extinction while others perhaps benefit in some way from such changes.

The opening of the Arctic presents many security challenges because of the high potential for changing global economic (and thereby, geo-political) power balances. Once the Arctic becomes economically exploitable, it may provide a large fraction of new global oil, gas, and mineral reserves. The open-water conditions of the future will also (1) allow for a dramatic increase in shipping, (2) could spur spectacular infrastructure and processing development along the route, and (3) elevate economic and strategic competition among nation states.

Estimates indicate that open Arctic shipping routes could reduce transportation costs by an average of 40% on key Asian-European routes, cutting some distances by two-thirds. A simple use of economic data indicates that such reductions as a result of Arctic open water could attract up to 80% of the global transportation market. In the near-term, the reduction in

Arctic ice coverage will quickly open the Arctic for routine, seasonal marine transport. The use of icebreakers and ships built with ice capable hulls will greatly extend the shipping season. According to the Russians, their Arctica-class nuclear icebreakers make possible year-round navigation in the western section of the Northern Sea Route (NSR) (indicated in Figure 3) from Murmansk to River Lena as well as ports on major Siberian rivers.

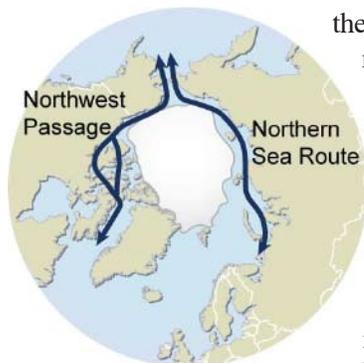


Figure 3. Arctic Shipping Routes  
Courtesy Hugo Ahlenius, UNEP/  
GRID-Arendal

In the near term, much of the ship traffic will be for resupply of the burgeoning oil, gas, and mineral industry. According to a July 2008 report by the U. S. Geological Survey, as much as 13% of the world's undiscovered oil and 30% of undiscovered gas reserves are in the Arctic seabed. In a recent Center for Strategic and International Studies (CSIS) symposium, Senator Lisa Murkowski stated that the Arctic may contain up to 100 billion barrels of oil and 25% of the Earth's remaining oil and natural gas.

The changing conditions in the Arctic Ocean and the high-latitude land areas may create a new "gold rush" that has the potential for conflict. There are five nations with territorial claims in the Arctic: Russia, Canada, the U. S., Denmark, and Norway, Sweden, Finland, Iceland, and semi-autonomous groups of indigenous peoples also have economic and strategic interests.

If Arctic trade and supply chains develop as imagined, the balance of political and economic power within the tropics and Southern Hemisphere will change because trade among northern hemisphere countries will depend more heavily on Arctic routes. The equatorial (Panama) and southern routes (Africa, Indonesia, and S. America) will experience severe dislocations. Economic expansion in, and strategic use of, the Arctic will provide challenges to the DHS (USCG), DoD, DOS, EPA, and Intelligence communities concerning law enforcement, treaties, monitoring, peacekeeping, marine safety, and protection of the environment.

### References

1. Backus, G., and Strickland, J. (2008). *Climate-Derived Tensions in Arctic Security*, SAND2008-6342, Sandia National Laboratories, Albuquerque, NM.
2. Boslough, M., Ivey, M., Taylor, M., Zak, B., Backus, G. (2008). *The Arctic as a Test Case for an Assessment of Climate Impacts on National Security*, SAND2008-7006, Sandia National Laboratories, Albuquerque, NM.
3. *Arctic Climate Impact Assessment (2005)*. Eds. C. Symon, L. Arris, B. Heal. Cambridge University Press, New York, NY, 2005.
4. *Climate Change 2007: Synthesis Report (2007)*. Eds. R. K. Pachauri and A. Reisinger, Intergovernmental Panel on Climate Change, Geneva, Switzerland.
5. Murkowski, L. (2008). *U.S. Strategy in the Arctic: Energy, Security, and the Geopolitics of the High North*, Americas Program, Energy and National Security, CSIS, Washington, DC.

***This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.***

# LARGE-SCALE ATOMISTIC SIMULATION DEVELOPED TO STUDY STOCKPILE MATERIALS CHEMISTRY

Contact: Aidan Thompson, Multiscale Dynamic Materials Modeling, [athomps@sandia.gov](mailto:athomps@sandia.gov)

*In collaboration with Caltech, the ReaxFF reactive interatomic potential was implemented in Sandia's GRASP massively parallel MD code, enabling multi-million atom molecular dynamics simulation of chemical processes in many important stockpile materials, including polymer oxidation, energetic materials, and semiconductor interfaces.*

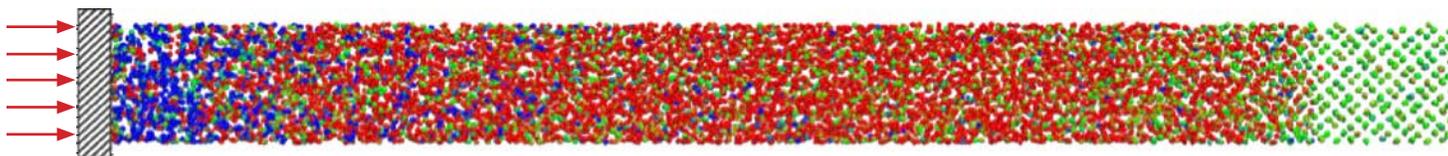
A new atomistic simulation capability was developed to study the effect of chemical processes on materials properties as part of the work performed for the Advanced Simulation and Computing Program's Physics and Engineering Models. In collaboration with Caltech, the ReaxFF reactive interatomic potential was implemented in Sandia's GRASP massively parallel MD code, enabling multi-million atom molecular dynamics simulation of chemical processes in many important stockpile materials, including polymer oxidation, energetic materials, and semiconductor interfaces. ReaxFF was previously developed at the Caltech ASC Alliance Center to model chemical reaction events with accuracy approaching that of quantum electronic structure calculations, but with much greater speed. The GRASP implementation of ReaxFF is now being used by research groups at DOE, DoD, NASA, and many universities. Most recently, the GRASP implementation of ReaxFF has been incorporated into Sandia's widely used LAMMPS package.

A key driver for developing this parallel capability is the need to perform large-scale simulations to understand how detonation occurs in explosives, which is critical to predicting how these materials perform in weapon devices. While the propagation of the steady detonation front can be accurately represented by continuum mechanics models, how this detonation front forms because of an initial mechanical or thermal excitation cannot. This is because the initiation process depends strongly on subcontinuum properties, such as intergranular voids, grain orientation, and crystal defects. For example, it has been observed that a single crystal of PETN detonates at shock pressures of 13 GPa and 31 GPa when shocked along the  $\langle 110 \rangle$  and  $\langle 100 \rangle$  direction, respectively.

The massively parallel implementation of ReaxFF in GRASP allows us to directly simulate the release of chemical energy in a PETN crystal  $[\text{C}(\text{CH}_2\text{ONO}_2)_4]$  when it is shocked along different crystal orientations. Initially, our work has focused on the relatively insensitive  $\langle 100 \rangle$  direction [1]. All of the MD simulations have been performed on Sandia's Thunderbird computer. Each simulation required integrating the equations of motion of up to 500,000 particles for more than 100 picoseconds using a 0.1 femtosecond timestep. Using several hundred processors, we have been able to achieve a throughput of about 0.5 ps/day.

Figure 1 shows the single crystal PETN sample about 10 ps after colliding on the left with a reflective boundary at 3 km/s. Only the PETN nitrogen atoms are shown, colored by potential energy. We observe an initial sharp rise in potential energy right at the shock front, which persists for some distance until it drops because of the formation of stable chemical species.

By tracking the position of the shock front over time, we observe how the shock velocity changes in response to the storage and release of chemical energy behind the shock front. The difference between the 4 km/s impact (strong shock) and 3 km/s impact (weak shock) is quite striking (Figure 2). By looking at particular chemical species concentrations as a function of time and space (not shown), we have been able to correlate the observed changes in shock-front velocity with the storage and release of chemical energy. In the case of the strong shock, net energy release occurs after very short times, and within 20 ps, the shock-front appears to have reached a constant velocity, propelled by the formation of stable



**Figure 1.** Visualization of nitrogen atoms in a PETN crystal after it has collided against a reflective boundary. The atoms are colored by potential energy, which clearly illustrates the unshocked material (green), unreacted shocked material (red), and reacted shocked material (blue).

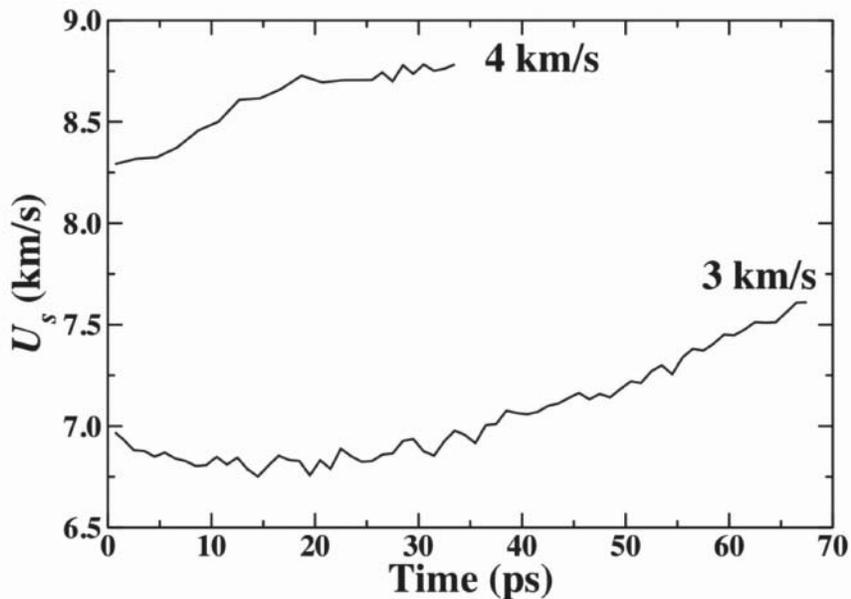
reaction products such as  $N_2$ ,  $H_2O$  and  $CO_2$ . In the case of weak shock, the shock-front initially decelerates because of the storage of energy in intermediate reaction products such as  $NO_2$ . After about 30 ps, these unstable intermediates start to convert to stable final products, releasing energy and causing the shock wave to accelerate.

The parallel implementation of ReaxFF in GRASP, along with Sandia's high-performance computing resources, have enabled us to perform atomistic simulations of detonation initiation on unprecedented length and time scales. They have provided unique insight into the role played by endothermic and exothermic chemical reactions in initiation. We are now applying the same methodology to study the more sensitive  $\langle 110 \rangle$  direction. Ultimately, we hope to identify the specific atomistic processes that determine which directions are sensitive and which are insensitive.

### References

1. Thompson, A. (2009). Reactive Molecular Dynamics Simulations of Shock Through a Single Crystal of Pentaerythritol Tetranitrate, submitted to the *Journal of Physical Chemistry B*.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.*



**Figure 2.** Plot of instantaneous shock-front velocity as a function of time for a weak shock (3 km/s) and a strong shock (4 km/s). The observed acceleration of the shock wave is a result of the net energy release that occurs when final reaction products are formed. In the case of the weak shock, the initial deceleration is a result of energy storage in intermediate reaction products.

# ADVANCED UNCERTAINTY QUANTIFICATION METHODS ENABLE UNDERSTANDING OF IMPACT OF RADIATION ENVIRONMENT ON NUCLEAR WEAPONS

Contacts: Wesley Fan, Radiation Transport, [wcfan@sandia.gov](mailto:wcfan@sandia.gov)

Laura Swiler, Optimization and Uncertainty Quantification, [lpswire@sandia.gov](mailto:lpswire@sandia.gov)

*We have performed the first large-scale UQ study of cable SGEMP, requiring hundreds of runs on high-performance computers, using approximately 270,000 processor-hours to generate simulation results. ... The sensitivity analysis and UQ provided greater understanding of input parameter influence on the cable responses, which should be incorporated in future studies in place of the worst-case analysis.*

Strategic missile weapon systems are required to survive the radiation environment produced by nuclear detonations. One potential damage mechanism to the electronics is a result of the radiation-induced transient electrical current on the internal cable conductor (a phenomenon called Cable SGEMP). In the past, extensive underground test experiments were required to design and certify that the electronic systems survive SGEMP. Today the underground test capability is no longer available, and there is no test capability that replicates the threat-level conditions. As a result, a robust computational simulation capability and implementation of certification-based design approaches (i.e., design to test) are required to complement experiments using the existing above-ground test facilities.

We used Sandia-developed codes to model radiation transport in the cable, and subsequent electromagnetic effects [1]. We used simulations to perform uncertainty quantification (UQ). UQ involves identifying and characterizing uncertainties in a physics model, propagating these uncertainties through a physics simulation code, and determining the effects on system output response metrics. In this work, the uncertainties associated with the input and model parameters were specified as interval quantities. The UQ analysis simply maps from intervals on the inputs to intervals on the response metrics.

The codes used were based on finite element models and were computationally expensive to run. We had to carefully design the uncertainty analysis to maximize the information obtained with a minimum number of runs. This study exploited the loose coupling between the radiation transport and electromagnetic calculations, and employed an iterative Latin hypercube sampling [2] method (via Sandia's DAKOTA toolkit) [3], that is, fewer samples of the more expensive radiation transport solver, and more samples on the less expensive electromagnetic solver.

The UQ results are illustrated in Figure 1, where each iteration number represents a specific set of radiation transport parameter values, while each vertical bar represents the interval-valued output responses obtained from sampling the electromagnetic simulation parameter values. Taken together, these results not only provide estimates on the upper and lower bound of the cable responses, but can also be used in a sensitivity analysis study to identify the input variables that have significant effects on the responses. Furthermore, we extended the cable analysis to include a specific circuit component so that we could assess performance of the protection devices. The performance margins were determined by comparing the circuit responses to the prescribed damage threshold.

The most important accomplishment of this milestone was developing the capability (through scripts, executables, meshes, programs, and a UQ framework) for performing uncertainty analysis for cable SGEMP given all the physics steps involved in the calculations. We have performed the first large-scale UQ study of cable SGEMP, requiring hundreds of runs on high-performance computers, using approximately 270,000 processor-hours to generate simulation results. This was also the first "end-to-end" computational study with coupled analysis of cable and circuit performance. The sensitivity analysis and uncertainty quantification provided greater understanding of input parameter influence on the cable responses, which should be incorporated in future studies in place of the worst-case analysis. This work helped achieve greater confidence in our code performance over a range of situations. We developed a process that is being used as a guideline for other SGEMP assessments and by designers to identify trade-offs in cable design studies.

### References

1. Liscum-Powell, J. L., Bohnhoff, W. J., Turner, C. D. (2007). *A Cable SGEMP Tutorial: Running CEPXS, CEPTRE and EMPHASIS/CABANA to Evaluate the Electrical Response of a Cable*, SAND2007-2548, Sandia National Laboratories, Albuquerque, NM.
2. Swiler, L. P. and Wyss, G. D. (2004). *A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library Standalone Version*, SAND2004-2439, Sandia National Laboratories, Albuquerque, NM.

3. Eldred, M.S., et al. (2006). *DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 4.0 Users Manual*, SAND2006-6337, October 2006. Updated September 2007 (HYPERLINK ["/licensing/release/Users4.1.pdf"](#) Version 4.1).

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***

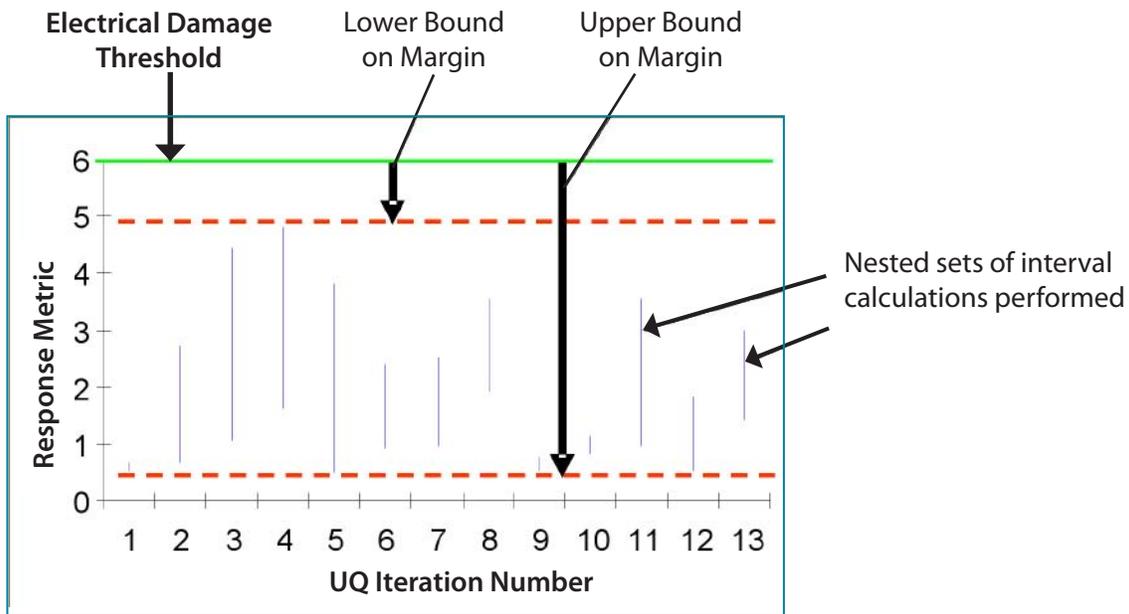


Figure 1. Upper and lower bound of cable responses.

## EFFECTIVELY MODELING Z-MACHINE WIRE ARRAY IMPLOSIONS

Contacts: Bill Rider, Comp. Shock and Multiphysics, [wjrider@sandia.gov](mailto:wjrider@sandia.gov)  
 Allen Robinson, Comp. Shock and Multiphysics, [acrobin@sandia.gov](mailto:acrobin@sandia.gov)  
 John Niederhaus, Comp. Shock and Multiphysics, [jhniede@sandia.gov](mailto:jhniede@sandia.gov)  
 Raymond W. Lemke, HEDP Theory, [rwlemke@sandia.gov](mailto:rwlemke@sandia.gov)

*With the utilization of the new methods in ALEGRA, in combination with an innovative technique for modeling the vaporizing wire array, we discovered that we could now simulate wire-array implosions with enough accuracy that they could be compared favorably with experimental data. Furthermore, ALEGRA could be used to analyze and design follow-on experiments because of this predictive capability.*

Sandia's Z machine uses very high electrical currents to create extreme magnetic fields useful for examining high energy density matter. One useful concept is a wire-array implosion. In a wire-array implosion, huge amounts of electrical current are run through a cylindrical array of tungsten wires generating an enormous magnetic field (Figure 1). This magnetic field causes the wires to vaporize into plasma and then drives this plasma toward the center axis of the array at fantastic speeds (Figure 2). When this plasma stagnates at the axis, it generates a huge amount of heat that exits the wire array as radiation. Modeling this chain of events is enormously challenging because of the wide breadth of physics involved under extreme conditions.

ALEGRA is a magnetic-radiation-hydrodynamic code that is designed to meet the challenge of modeling the wire-array implosion. ALEGRA has been around since 1994, but improvements in recent years have now enabled the effective simulation of wire-array implosions. A key

enabling technology has been the implementation of a uniquely powerful solution to the magnetic portion of the problem using what are called compatible discretizations, which in a nutshell preserve the fundamental properties of magnetic flows exactly. In addition, codes like ALEGRA typically solve for the energy of the flow concentrating on the thermal content of the fluid. The problem is that flows like those found in a wire array are dominated by the energy of the magnetic field or the motion of the flow for significant portions of their evolution.

In fact, the process is essential where the energy of the motion comes to a halt when the flow hits the center axis of the wire array and is converted into heat. Because of this, we had to modify the solution for the energy of the flow to more accurately account for this process. This change required us to rearrange the energy accounting in the flow so that this process could be accurately simulated.

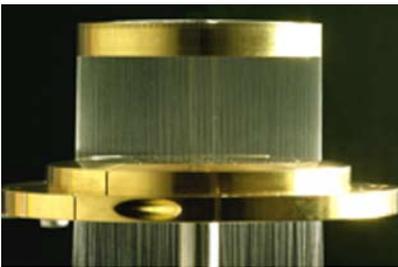


Figure 1. A wire array configuration used as a load for the Z-machine is shown above.

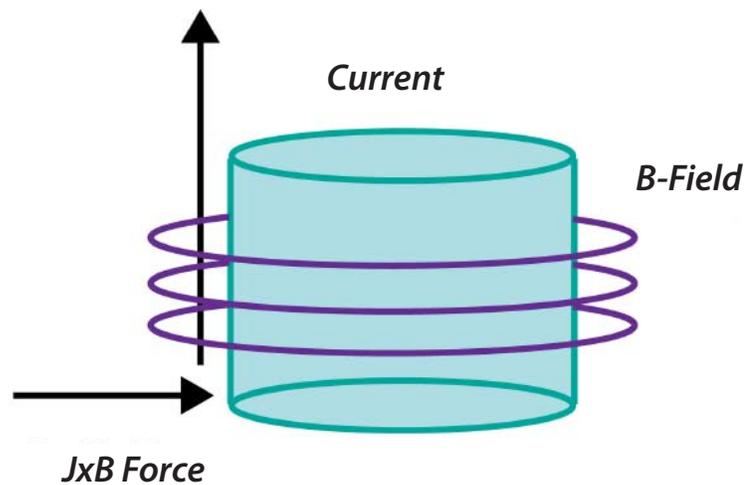


Figure 2. The current and associated magnetic field in the Z configuration drive an cylindrical implosion.

With the utilization of the new methods in ALEGRA, in combination with an innovative technique for modeling the vaporizing wire array, we discovered that we could now simulate wire-array implosions with enough accuracy that they could be compared favorably with experimental data (Figure 3). Furthermore, ALEGRA could be used to analyze and design follow-on experiments because of this predictive capability. This provided the scientists working on experiments in the Z-machine with new-found confidence in the predictive power of the simulations with ALEGRA.

We continue to improve the algorithms by careful development and detailed analysis. We have developed a similar method for conserving magnetic energy during remap. Impressive improvements in shock test problems are observed by accounting for this energy (Figure 4).

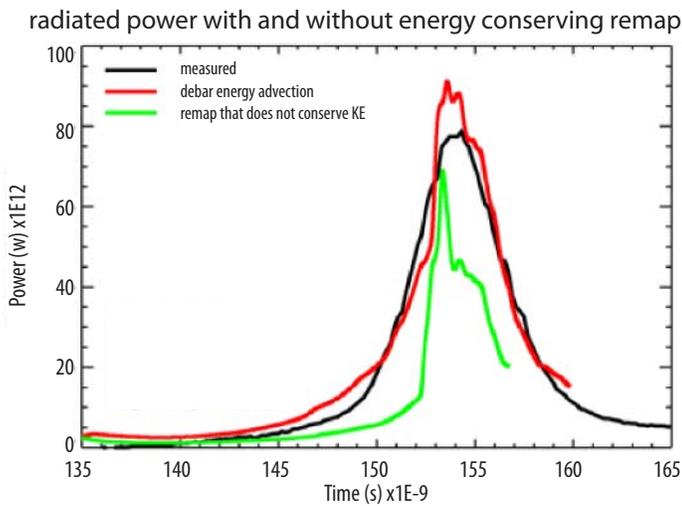
### References

Lemke, R., et al. (2009). Effects of Mass Ablation on the Scaling of X-Ray Power with Current in Wire-Array Z Pinches, *Physical Review Letters* DOI: 10.1103/PhysRevLett.102.025005. Published Online.

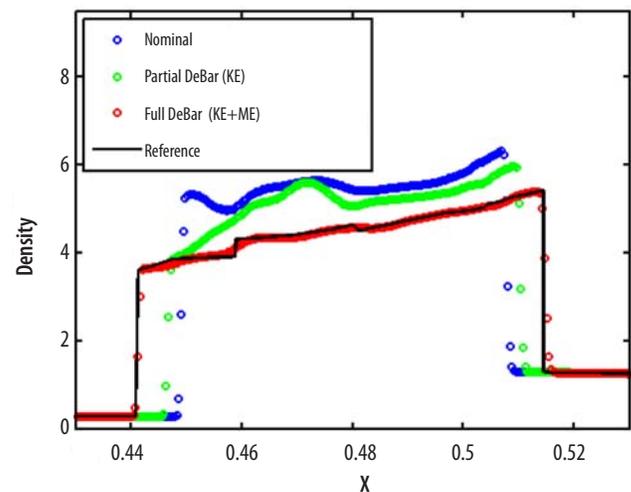
<http://link.aps.org/abstract/PRL/v102/e025005>

Robinson, A., et al. (2008). ALEGRA: An Arbitrary Lagrangian-Eulerian Multimaterial, Multiphysics Code. AIAA 2008-1235, 46th AIAA Aerospace Sciences Meeting and Exhibit, 7 - 10 January 2008, Reno, Nevada .

***This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program.***



**Figure 3.** Algorithms to ensure that kinetic energy is preserved with material motion allow for vastly improved comparison with data.



**Figure 4.** Magnetic energy conservation algorithm with material motion shows that additional quantitative jumps in solution accuracy are possible.

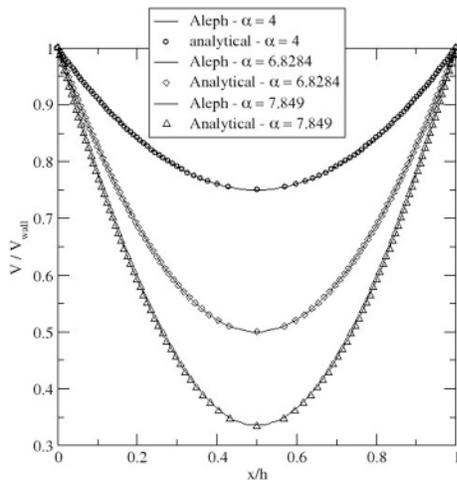
# AN ADVANCED TOOL FOR PLASMA SIMULATION

Contacts: Paul Crozier, [prcrozi@sandia.gov](mailto:prcrozi@sandia.gov)  
 Lawrence Musson, [lcmusso@sandia.gov](mailto:lcmusso@sandia.gov)  
 Matthew Hopkins, [mmhopki@sandia.gov](mailto:mmhopki@sandia.gov)

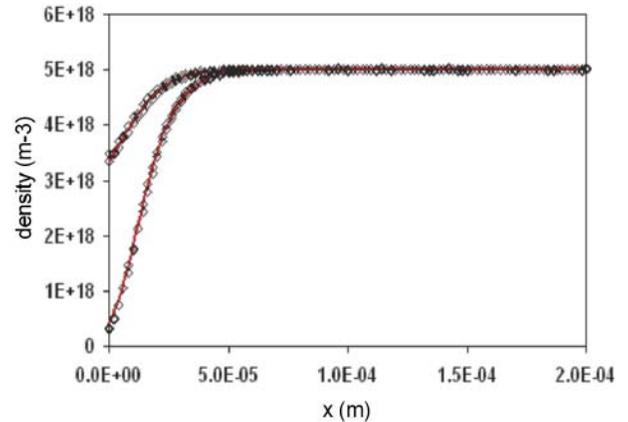
*Our team has built Aleph as part of a larger effort aimed at neutron tube simulation, with contributions from many more people. It is anticipated that use of Aleph for neutron tube simulation will result in tremendous cost savings in design and manufacturing.*

Over the past two years, we have created a new software tool (Aleph) for the simulation of low-density plasmas. Aleph is a direct simulation Monte Carlo (DSMC) plus particle-in-cell (PIC) code capable of particle representations of plasmas within arbitrary 3D geometries defined by unstructured tetrahedral meshes. We have built the software upon the foundation of existing software modules, many from the Trilinos library, such that Aleph includes dynamic load balancing and hence performs and scales well on available high-performance compute resources. Our team has built Aleph as part of a larger effort aimed at neutron tube simulation, with contributions from many more people. It is anticipated that use of Aleph for neutron tube simulation will result in tremendous cost savings in design and manufacturing. Secondary application areas with impact on Sandia's mission include modeling of ion beams, arcs, and hypersonic upper-atmosphere flows.

To establish confidence in Aleph's plasma simulation capabilities, several simple problems with analytical or otherwise well-accepted, published solutions have been used to verify the solution methods of the chosen physical models.



**Figure 1.** The graph shows the effect of the electron beam on the potential in the gap as a function of  $a$ , the ratio of the injected current to what would flow in a Child's law diode. A full description of the variables and the analytical solution can be found in Birdsall & Bridges, *Electron Dynamics of Diode Regions*.



**Figure 2.** Plasma sheath simulation density profiles of positively charged ions (upper line) and electrons (lower line). The simulation results (black symbols) compare favorably with the analytical solution (red lines).

Two highlights of the suite of verification problems are a classical electron beam diode and the formation of a plasma sheath near a conducting wall. The diode is formed by sending a beam of electrons across an evacuated gap formed by two infinite parallel plates, each charged to an equal potential. The electrons transiting the gap depress the applied field in the gap as shown in Figure 1. If the flux of electrons from the originating plate exceeds a certain threshold, a resisting electric field is formed that the electrons cannot overcome, and the current shuts off. The second notable verification problem, the sheath problem, is one in which a surface, such as that on a Langmuir probe, is plunged into an otherwise neutral plasma. The sheath is defined to be the boundary layer of unequal ion and electron densities (and resultant potential gradient) adjacent to the surface (Figure 2).

Ion beam modeling is a central thrust of this code-development work. In current models, ions are extracted from a plasma wherein ions and electrons are explicitly represented. A grounded plate with a slit (2D) or circular (3D) aperture allows the passage of ions into a gap formed by the aperture plate and a negatively charged target electrode (Figure 3). The model allows for a directed study of the effects of operating parameters and aperture geometry on the

beam's focus and the overall current delivered to the target, as well as the stability of the beam. Future ion beam model and code development will include the effects of a nonvacuum environment, sidewall charging by excited neutrals, and secondary electron emission.

Arcs can act as sources for plasmas and ion beams, so complete modeling of such an ion beam device should include modeling of the arc discharge phenomenon. To the best of our knowledge, we are the first research group in the world to attempt particle simulation of an arc discharge. We have built a simulation model of a hot anode vacuum arc (HAVA) that includes the key physical phenomena: high-energy electrons originating from the cathode bombard the titanium anode; the anode heats up and emits neutral titanium atoms; some of the neutrals are then ionized by the incoming electrons; production of low-energy secondary electrons makes additional electron production more likely; an avalanche of electron production dramatically boosts the anode-to-cathode current (Figure 4).

The hot anode in our arc discharge model is simulated as a 3D heat-conduction problem, with electron bombardment acting as the heat source on the surface and the transfer of energy through the anode tracked dynamically. A vapor-pressure equation is used to compute the rate of titanium particle vaporization into the void as a function of the anode surface temperature. Each computational particle carries a weighting that signifies how many real particles each one represents. To reduce computational cost, this weighting can

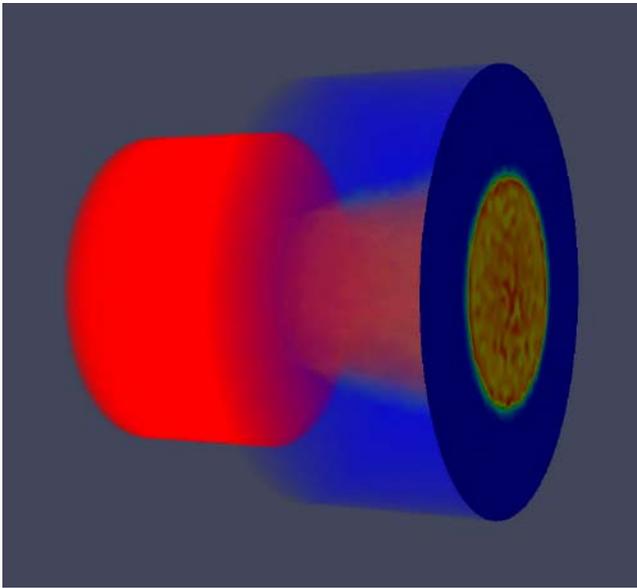
vary spatially and temporally, and according to the species type, but must be properly considered when computing the particle-induced electrostatic fields and when performing particle collisions and reactions. In addition to the variety of particle physics capabilities required for our HAVA model, a range of diagnostic features was also needed. Aleph features a rich array of such diagnostics for easy output of temporally and spatially averaged quantities of interest, such as density distributions of the various species and particle fluxes or currents through surfaces.

In summary, we have developed an advanced plasma simulation software package intended for enhanced physical modeling of neutron tube performance. We have built confidence in Aleph's simulation capabilities by successfully modeling known plasma phenomena, and we have used the code to perform novel research in the area of arc discharge modeling.

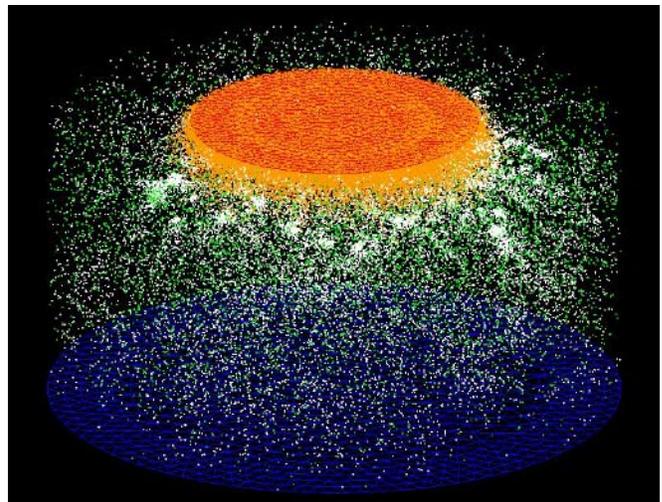
### Reference

Hopkins, M. M., Crozier, P. S., Hooper, R., et al. Progress Simulating Low Temperature Collisional Plasmas with a PIC-DSMC Method, 8th World Congress on Computational Mechanics (WCCM8); 5th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2008), June 30th – July 5, 2008, Venice Italy.

*This work has been supported by the National Nuclear Security Administration's (NNSA) Advanced Simulation and Computing (ASC) Program, Directed Stockpile Work (DSW), and Campaign 6.*



**Figure 3.** This Aleph-produced image shows an ion beam as it exits the aperture and transits the gap to the target. Also shown is the intensity of the beam incident on the target electrode.



**Figure 4.** Arc simulation image at  $t = 0.499$  ms showing virtual cathode (blue), anode (red), titanium neutrals (orange), titanium ions (white), and electrons (green). Titanium ions are produced in chain reaction bursts as evidenced by the highly concentrated pockets of white dots.

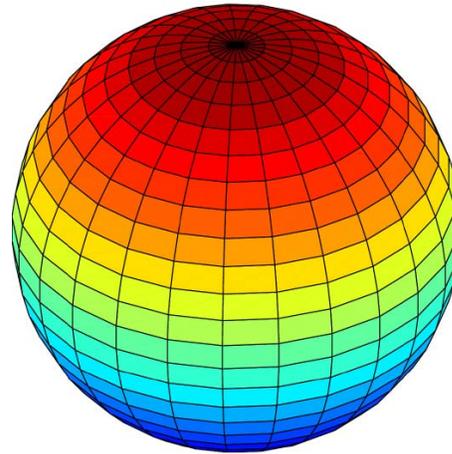
# COMMUNITY CLIMATE SYSTEM MODEL DEVELOPMENT FOR PETASCALE SUPERCOMPUTERS

Contact: Mark Taylor, Exploratory Simulation Technologies, [mataylo@sandia.gov](mailto:mataylo@sandia.gov)

*Team members at Sandia, The National Center for Atmospheric Research (NCAR), and Oak Ridge National Laboratory have ... been focusing on the integration of new, more scalable dynamical cores based on cubed-sphere grids into the Community Climate System Model (CCSM)... We now have a dynamical core in the CCSM with unsurpassed scalability and very competitive numerics.*

Climate change is already under way and will accelerate in this century, but there are still major uncertainties. Human societies will respond with both mitigation and adaptation strategies, which may include changes in consumption of energy and other resources, land use, agriculture, and migration. These stresses could also lead to a shift in alliances and to civil unrest and conflict. Adapting to new situations will require advanced understanding of climate change and its direct and indirect effects. Improved understanding of climate change is of paramount importance for mission planning, surveillance, facilities placement and design, and sustainable energy futures. Developing this understanding requires high-fidelity climate models that simulate many components of the Earth system, including atmospheric and oceanic circulation, land surface processes, sea ice, and chemical and biogeochemical cycles. Adding full chemical and biogeochemical cycles to today's climate models will require significantly more computational resources than are available today. Furthermore, an increase in resolution to allow for the accurate assessment of regional impacts of climate change also requires further increases in computational power. Achieving both of these goals will require petascale computer architectures.

The DOE petascale computers have hundreds of thousands of processors. Effectively using such machines remains a challenge because of several scalability bottlenecks present in all modern climate models, the largest of which is created by the numerical methods used in the dynamical core of the atmospheric model component. The dynamical core solves the partial differential equations governing the fluid dynamical aspects of the atmosphere. In addition, atmospheric models contain a suite of subgrid parametrizations for the many physical processes unresolved in an atmospheric model but which drive the dynamics such as convection, precipitation, and radiative forcing. Currently, most dynamical cores use latitude-longitude based grids (Figure 1). These grids create a logically Cartesian orthogonal mesh suitable for a wide array of numerical methods, including finite volumes and the spectral transform method. The grid points cluster at the



**Figure 1.** A latitude-longitude grid showing the clustering of grid points at the poles. This creates numerical difficulties in climate models that can be overcome, but at the cost of degrading parallel scalability.

pole, creating several computational difficulties collectively referred to as the pole problem. The biggest difficulty is the severe Courant-Friedrichs-Lewy restriction on the timestep caused by this clustering. There are many successful techniques to handle this pole problem; however, most of them substantially degrade parallel scalability by requiring too much inter-processor communication.

Team members at Sandia, the National Center for Atmospheric Research (NCAR), and Oak Ridge National Laboratory have thus been focusing on the integration of new, more scalable dynamical cores based on cubed-sphere grids (Figure 2) into the Community Climate System Model (CCSM). For our initial work, we used the spectral element dynamical core in NCAR's High-Order Method Modeling Environment (HOMME). The spectral element method used in HOMME is a type of  $h$ - $p$  finite element method and thus naturally supports cubed-sphere grids. This approach completely avoids the pole problem and has given us a version of the CCSM atmospheric component with unprecedented scalability. We can now run efficiently on 86,000 processors with a horizontal average grid spacing of 25 km. Even better scalability will be possible when computing with a global resolution of 10 km, DOE's long-term goal.

Parallel performance and scalability are only one of the necessary requirements of a petascale-ready dynamical core. The quality of the solutions generated is always the most important consideration. Our initial experiments with the HOMME dynamical core coupled to the full suite of CCSM atmospheric subgrid physics revealed three areas where the numerical methods needed improvements: The numerics in HOMME did not conserve mass or energy, which we addressed by developing a new *compatible* formulation of spectral elements, making HOMME the first dynamical core in the CCSM to conserve both mass and energy without the use of *ad hoc* fixers. The original limiter-based dissipation mechanisms in HOMME resulted in noticeable grid imprinting in the solution, which we addressed by replacing the limiters with an isotropic hyper-viscosity operator. Finally, we replaced the oscillatory tracer advection scheme with a conservative, nonoscillatory and sign-preserving advection operator. For more details, see [1, 2, 3].

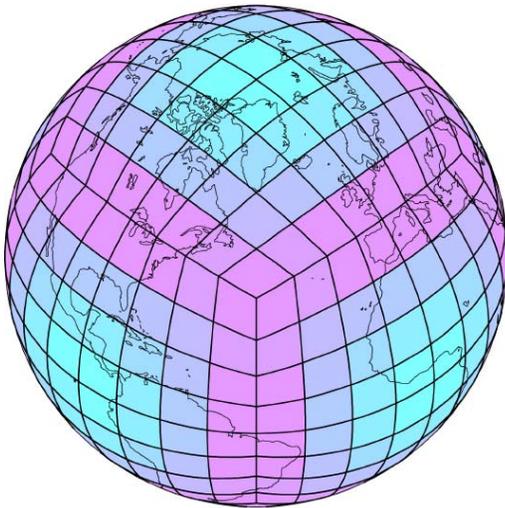
With the improvements listed above, we now have a dynamical core in the CCSM with unsurpassed scalability and very competitive numerics. We have completed extensive verification work using the *aqua planet* standardized atmospheric test. This test uses the full atmospheric physics from the CCSM with prescribed sea surface temperatures and without the CCSM land, ice, or ocean models. A snapshot is shown in Figure 3. Our current focus is the completion of more realistic AMIP (Atmospheric Model Inter-Comparison Program) simulations. AMIP simulations use the full

atmosphere and land models in the CCSM, but the ice and ocean modes are replaced by prescribed data sets that are derived from observations. Excellent AMIP simulations will establish that our new approach is suitable for use in climate science investigations.

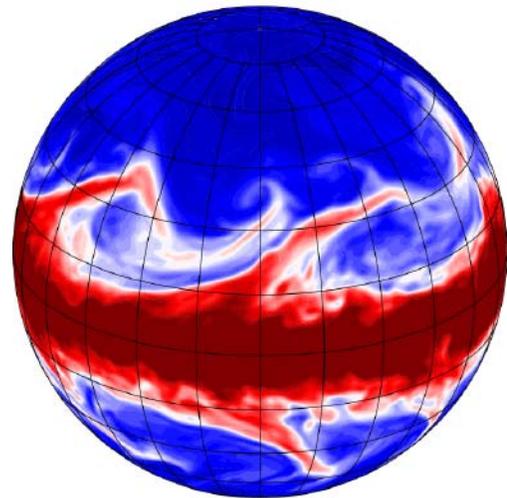
### References

1. Taylor, M. A., St.Cyr, A., and Fournier, A. (2009). A Non-Oscillatory Advection Operator for the Compatible Spectral Element Method (to be published in Computational Science ICCS 2009, *Lecture Notes in Computer Science*, Springer).
2. Taylor, M. A., Edwards, J., and St.Cyr, A. (2008). Petascale Atmospheric Models for the Community Climate System Model: New Developments and Evaluation of Scalable Dynamical Cores, *J. Phys. Conf. Ser.*, Vol. **125**.
3. Taylor, M. A., Edwards, J., Thomas, S., Nair, R. (2007). A Mass and Energy Conserving Spectral Element Atmospheric Dynamical Core on the Cubed-Sphere Grid, *J. Phys. Conf. Ser.*, Vol. **78**.

***This work supported by a grant from the Biological and Environmental Research Program of the Department of Energy, Office of Science: Grant # 06-13194, A Scalable and Extensible Earth System Model for Climate Change Science.***



**Figure 2.** A cubed-sphere grid used by the spectral element atmospheric model component of the CCSM. The color represents the area of each grid cell. The cells are much more uniform when compared to the cells in a latitude-longitude grid.



**Figure 3.** A snapshot of precipitable water over the surface of an Aqua planet, simulated using the atmospheric component of the CCSM with the spectral element dynamical core on a cubed-sphere grid. Aqua planet simulations are used to test new dynamical cores and other physical processes in climate models before coupling with other components such as land, ocean, and ice models. The color scheme is suggestive of clouds, but we note that even on petascale computers, climate models will lack sufficient resolution to directly model clouds. Instead, the effects of clouds on the resolved scales are modeled with subgrid parametrizations.

# INFORMATICS ENABLES DATA-DRIVEN DISCOVERY AND PREDICTION

Contact: Suzanne L. K. Rountree, Computer Science & Informatics, [slkroun@sandia.gov](mailto:slkroun@sandia.gov)

*Sandia is carving a unique niche and positioning itself as a leader in solving complex and large-scale informatics problems, especially those involving national security. ... Math and computer science researchers in informatics have teamed with mission-driven analysts and experts in uncertainty quantification and human factors to focus on decision-making in complex national security environments.*

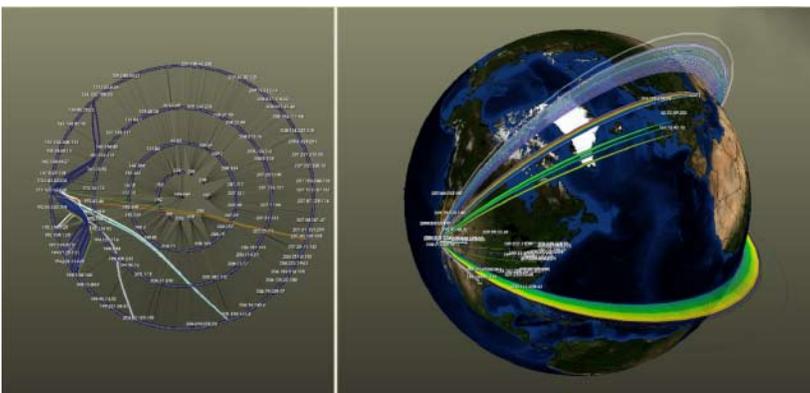
Informatics is the science that encompasses complex information and relationship-based analytic methods to support decision-making in uncertain and massive-data environments. Sandia is carving a unique niche and positioning itself as a leader in solving complex and large-scale informatics problems, especially those involving national security. To that end, Sandia is building a strong research program that utilizes many distinguishing laboratory strengths: discrete math, fast graph analytics, alternative data machines, linear algebra methods, information visualization, computer architectures, high-performance computing, and scalable algorithms.

Social networks, like terrorist networks, are a good example of complex relationships that challenge informatics capabilities. People can be described in terms of their human attributes, the characteristics of their activities, and the interrelationships between their activities that connect them to others. When people engage in suspicious activities that threaten national security — through networks of adversaries engaged in weapons proliferation, terrorism, cyber attacks, and other illicit activities — and when these adversarial networks in turn hide their activities among the complex interactions between legitimate and illegitimate secondary networks (e.g., for communication, computing, supply chain, and financial transactions), then transformational methodologies are needed to ferret out the threats buried within massive amounts of benign information [1].

Math and computer science researchers in informatics have teamed with mission-driven analysts and experts in uncertainty quantification and human factors to focus on decision-making in complex national security environments. One research project in particular, focuses on informatics for discovery and prediction. A sub-team on that project recently demonstrated a thin slice of end-to-end capabilities in a prototype targeting cyber analysis in computer networks. Using fabricated network traffic test data from MIT Lincoln Labs, the sub-team processed raw data and analyzed it to answer four postulated questions: (1) What computers were communicating with each other? (2) What network file transfers crossed country boundaries? (3) What information was contained in the transferred files? (4) What suspicious network activities occurred?

Figure 1 depicts several linked visualizations that answer question #2 [2]. The ring-view (left) shows successive rings of computer networks, sub-networks, and individual computers (white IP addresses) exchanging files across the Internet. The geospatial view (right) integrates an IP address lookup function to show the countries of origin and receipt where the computers reside. The example leverages scalable data queries with analytical graph searches and visualization for presentation to the analyst.

Figure 2 also shows how information visualization can quickly depict unusual or suspicious behaviors that need further investigation (question #4). All the “normal” and



**Figure 1.** Depiction of network file transfers between computer IP addresses (shown as arcs in the ring-view on the left), and network file transfers crossing country boundaries (shown as arcs in the geospatial view on the right).

expected traffic fades to the background visually (as dim blue arcs), while the suspicious traffic is highlighted in a rainbow of colors where each color represents a different form of network activity between two computers (like file transfers). This helps the analyst to focus on the small subset of larger network traffic to answer whether the suspicious activities are legitimate or illegitimate. The scenario applies conditional statistics to the results of a data query before visualizing the results of the graph search. Figure 3 shows the results from powerful graph searches used to cluster and aggregate related information [3]: finding and identifying communities (groups of colored squares) and finding connected sub-graphs that give multiple paths (black lines) between two specific nodes (upper left red and lower right yellow).

Sandia has begun to address the design of an overall distributed systems architecture, targeting large informatics applications, that leverages the capabilities of different computers to perform the tasks that each does best: data machines for database searches, multi-threaded machines for rapid graph queries, client-server architecture for delivery of visual outputs to user-analysts, and high-performance computing and distributed memory machines for math computations. In a recent success demonstrating part of a

distributed systems architecture, Sandia was able to link data accesses on a Netezza data machine with visual analytics running on our Red Storm high performance computer (Figure 4). Our early prototype and systems architecture design are positioning Sandia to analyze informatics problems of unprecedented scale and complexity.

**References**

1. Colbaugh, R., Gosler, J., Glass, K. (2008). Some Intelligence Analysis Problems and Their Graph Formulations, invited paper, *Journal of Intelligence Community Research and Development*.
2. Wylie, B. (2008). GSpace: A Linear Time Graph Layout, Visualization and Data Analysis (VDA2008), San Jose, CA.
3. Hendrickson, B., Berry, J. (2008). Graph Analysis with High Performance Computing, *Computers in Science and Engineering*, 10 (2):14-19.

***This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.***

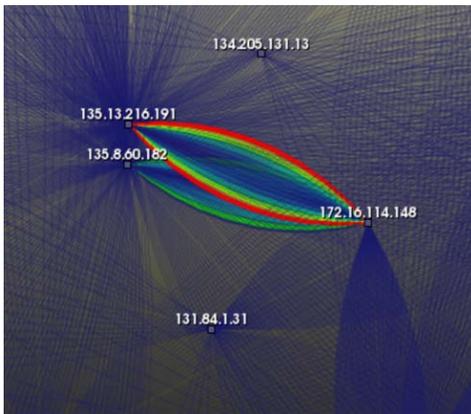


Figure 2. (Left) Example of information visualization with brightly colored arcs to highlight suspicious network activities. "Normal" network traffic fades visually when represented as dim blue arcs.

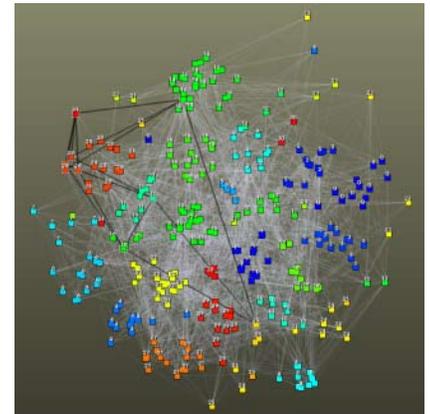


Figure 3. (Right) Powerful graph searches are used to cluster and aggregate related information (shown as commonly colored groupings). Connection sub-graphs depict multiple paths between two nodes in a graph (e.g., lines between upper left red node and lower right yellow node).

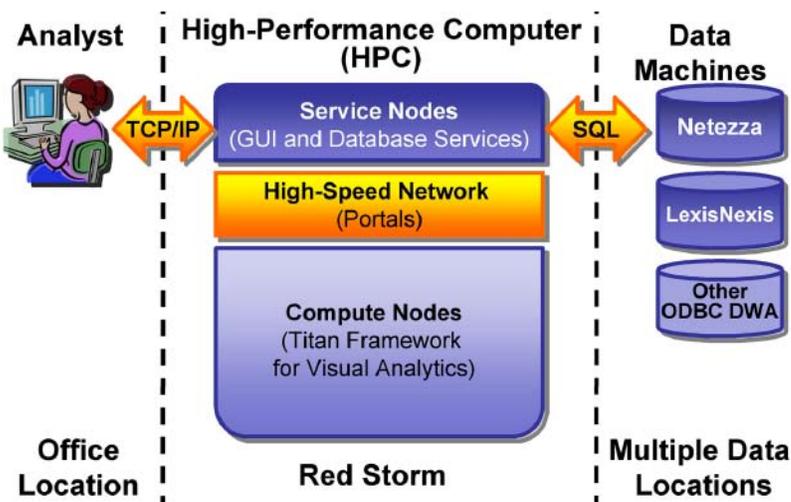


Figure 4. (Left) Distributed Systems Architecture: An analyst sitting at her workstation can run a high-performance computing (HPC) application on the Red Storm computer. The distributed systems architecture enables the HPC application to leverage remote data warehouse appliances (DWA) for efficient data access and management.

# VERIFICATION AND VALIDATION METHODOLOGY FOR COMPUTATIONAL COGNITIVE, BEHAVIORAL, AND SOCIAL MODELS

Contacts: Laura A. McNamara, Exploratory Simulation Tech., [lamcnam@sandia.gov](mailto:lamcnam@sandia.gov)  
 Timothy G. Trucano, Optimization and UQ, [tgtruca@sandia.gov](mailto:tgtruca@sandia.gov)  
 George A. Backus, Exploratory Simulation Tech., [gabacku@sandia.gov](mailto:gabacku@sandia.gov)  
 Scott A. Mitchell, Computer Science and Informatics, [samitch@sandia.gov](mailto:samitch@sandia.gov)

*Sandia National Laboratories is targeting a leadership role in developing Verification and Validation (V&V) approaches for computational social science applications.*

Across government and industry, computational modeling methodologies such as agent-based models, systems dynamics, and artificial neural networks are assuming a prominent role in simulations involving social, cognitive, and behavioral phenomena. As leaders make greater investments in computational simulations for policy decisions, methodologies for evaluating models will become increasingly important. Sandia National Laboratories is targeting a leadership role in developing Verification and Validation (V&V) approaches for computational social science applications. This requires understanding the unusual characteristics of V&V in computational social science, as well as its technical difficulties. We have suggested a methodological framework for addressing these challenges that builds on our previous experience in computational physics and engineering and that can be systematically incorporated into decision makers' assessments of computational social simulations.

In the United States, across government and industry, mathematical modeling and computational simulation have long been perceived as critical to decision making, in areas from nuclear weapons design and development to operations research. Until the 9/11 attacks, simulations of sociocultural phenomena were less common in policymaking circles. However, since 2001, federal agencies have invested heavily in computational modeling and simulation software to help analysts and decision makers better understand “soft” problems, like insurgency and terrorism. A 2006 article in *IEEE Spectrum* pointed out that the United States' national security enterprise is betting heavily that “computers equipped with the right software can give vital insights into the minds and motives of terrorists and the structure and critical links in their organizations” [1].

Largely lacking are comprehensive approaches to evaluating the goodness of the insights that computers provide decision makers. Sandia is particularly well placed to address this problem: Under the Advanced Simulation and

Computing Program (ASC), Sandia assumed a leading role in developing verification and validation (V&V) methodologies for assessing the internal and external correctness of computational simulations. In our work, we have been leveraging the principles of the ASC V&V Program in developing a robust framework for evaluating computational simulations of social and/or psychological phenomena.

In computational science and engineering (CS&E), verification and validation refer to a suite of methodologies that assess the internal (verification) and external (validation) correctness of a model-based simulation. Most of the specific methods developed for V&V at Sandia derive from computational physics or engineering problems, where mathematical or logical proof can be combined with planned experiments to develop rigorous assessments of a code's adequacy for the decisions its developers intend it to support. Developing verification and validation approaches appropriate for the simulation techniques used in computational social science—which range from game-based training models to systems dynamics models—presents a range of challenges, from methodological to organizational. These include the development of verification methods, the difficulty of designing and conducting high-fidelity social experiments for validation, and the problem of balancing resources for simulation evaluation against equally pressing demands for model development and application in resource-constrained environments.

To address these challenges, we have focused on developing a high-level framework for simulation evaluation that incorporates a range of domain-independent concepts for verification and validation activities. We begin by defining verification and validation *as the systematic accumulation of evidence that a model/simulation is adequate for application in the intended decision context*. This means that planning for verification and validation activities begins well before coding starts, when a decision context is identified and simulation requirements specified.

In this regard, several elements of the ASC V&V Program can be leveraged to support robust evaluation. Two examples are the Predictive Capability Maturity Model (PCMM) and the Phenomenon Identification and Ranking Table (PIRT), tools used to systematically plan and implement evaluation activities. A PCMM identifies elements of the simulation and maps requirements for their technical maturity to the application of the simulation. High-consequence decisions that rely heavily on simulation results demand correspondingly robust demonstrations of maturity, which implies significant verification and validation investments. The PIRT requires that teams systematically identify the phenomena that the simulation is attempting to replicate and documents the relative importance of each phenomenon and the degree to which it is adequately understood. In doing so, the PIRT helps modeling teams prioritize V&V investments in relation to the evaluation requirements implied by the PCMM [2].

To demonstrate the cross-domain applicability of ASC-based approaches, we worked with a cognitive science modeling project to develop an evaluation-planning rubric that incorporated many ASC approaches and concepts, including the PCMM and the PIRT. This resulted in the development of an extensive evaluation approach for a particular cognitive model of human memory formation, details of which are available in our recent report [3].

## References

1. Goldstein, H. (2006). Modeling Terrorists: New Simulators Could Help Intelligence Analysts Think Like the Enemy, *IEEE Spectrum* (September 2006), pp. 26-35.
2. Oberkampf, W. L., Pilch, M., and Trucano, T. G. (2007). *Predictive Capability Maturity Model for Computational Modeling and Simulation*, Sandia National Laboratories, Albuquerque, NM.
3. McNamara, L.A., Backus, G. A., Trucano, T. G., Mitchell, S. A., Slepoy, A. (2008 ). *R&D for Computational Cognitive and Social Models: Foundations for Model Evaluation through Verification and Validation* (Final LDRD Report), SAND 2008-6453, Sandia National Laboratories, Albuquerque, NM.

*This work has been supported by SNL's Laboratory Directed Research and Development (LDRD) Program.*

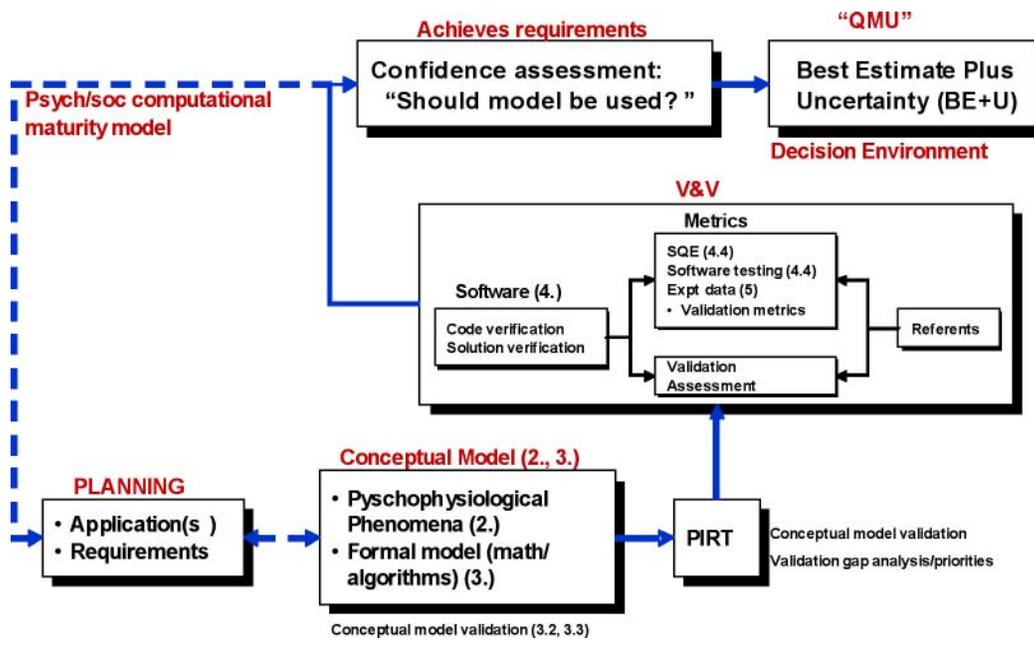


Figure 1. Adapting ASC V&V methods to psychological and social computational simulation.

## *AFTERWORD*

Organizationally, the Computation, Computers, Information and Mathematics Center (CCIM) originated in the mid-1980s. At that time, it was an enormous challenge to advance massively parallel computational science and engineering, let alone implement it usefully within the mission-driven research and development environment of Sandia. Perhaps the key strategic foundation for developing a world-leading HPC center was the deep belief that the organization needed to have three major intellectual themes at the core of its R&D – fundamental research in computational algorithms, fundamental research in advanced HPC architectures, and intersection of frontier mission-driven applications. The synergy and integration among these three elements was envisioned to be critical to advance massively parallel computing and create sustained impact for Sandia’s mission spaces.

When considering the summaries of CCIM work presented in this report, the importance of this three-element strategic principle remains, perhaps surprisingly, fresh and relevant over 20 years later. We have, of course, emphasized these elements in the way that we have organized this report in its major sections. But the work summarized also reinforces the critical nature of interaction and integration within our organization’s R&D environment. The technical work presented illustrates our reliance upon multidisciplinary approaches that incorporate algorithmic, hardware, and applications perspectives to leading-edge computational science and engineering for the 21st century. This collection demonstrates that vertical integration and our differentiating approach are essential elements in supporting and impacting a wide spectrum of Sandia’s national security missions.

It may go without saying that if one examined a similar compilation of CCIM technical work from 20 years ago, the portfolio of basic research in algorithms and hardware would be significantly different. HPC has been, and continues to be, a very dynamic field. Furthermore, Sandia’s mission spaces have diversified into many new national security thrusts over the past twenty years. This report illustrates the enormous agility of the organization’s response and impact on these very fluid environments.



NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

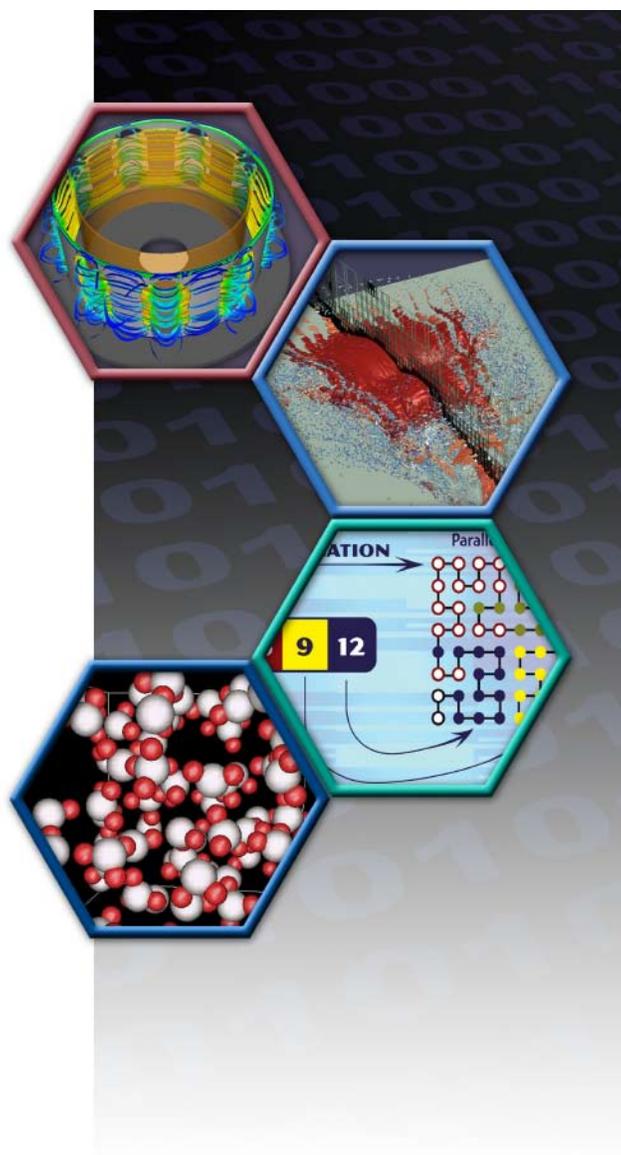
U.S. Department of Energy  
Office of Scientific and Technical Information  
P.O. Box 62  
Oak Ridge, TN 37831

Telephone: (865) 576-8401  
Facsimile: (865) 576-5728  
E-Mail: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)  
Online ordering: <http://www.osti.gov/bridge>

Available to the public from

U.S. Department of Commerce  
National Technical Information Service  
5285 Port Royal Rd.  
Springfield, VA 22161

Telephone: (800) 553-6847  
Facsimile: (703) 605-6900  
E-Mail: [orders@ntis.fedworld.gov](mailto:orders@ntis.fedworld.gov)  
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>





Sandia National Laboratories